EXPERT JUDGMENT IN THE AUDIT TEAM: AN EXAMINATION OF SOURCE CREDIBILITY

The Ohio State University

University Microfilms International

Copyright 1980 by Bamber, Edward Michael

All Rights Reserved
PLEASE NOTE:

In all cases this material has been filmed in the best possible way from the available copy. Problems encountered with this document have been identified here with a check mark.

1. Glossy photographs
2. Colored illustrations
3. Photographs with dark background
4. Illustrations are poor copy
5. Print shows through as there is text on both sides of page
6. Indistinct, broken or small print on several pages
7. Tightly bound copy with print lost in spine
8. Computer printout pages with indistinct print
9. Page(s) lacking when material received, and not available from school or author
10. Page(s) seem to be missing in numbering only as text follows
11. Poor carbon copy
12. Not original copy, several pages with blurred type
13. Appendix pages are poor copy
14. Original copy with light type
15. Curling and wrinkled pages
16. Other
EXPERT JUDGMENT IN THE AUDIT TEAM:
AN EXAMINATION OF SOURCE CREDIBILITY

DISTRIBUTION

Presented in Partial Fulfillment of the Requirements for the
Degree Doctor of Philosophy in the Graduate School of
The Ohio State University

By
E. Michael Bamber, B. Ec. (Honors), M. Ec.

***

The Ohio State University
1980

Reading Committee
Professor Thomas J. Burns
Visiting Professor Ernest L. Hicks
Associate Professor Richard J. Jagacinski
Assistant Professor Thomas E. Nygren

Approved by:
Thomas J. Burns
Adviser
Department of Accounting
ACKNOWLEDGEMENTS

I wish to acknowledge the invaluable contribution made to this study by the members of my dissertation committee: Thomas J. Burns, adviser and committee chairman, Ernest L. Hicks, Richard Jagacinski, and Thomas E. Nygren.

I am indebted to the audit managers of the accounting firms who gave time to participate as subjects in the experiment.

Special thanks are due to Lynette Gillick who patiently typed countless drafts of the test instruments.

I wish to acknowledge the Ph.D. students, the faculty, and the staff of the Department of Accounting at OSU who provided an atmosphere in which this study could be conducted with the least possible pain.

I wish to acknowledge the financial assistance in the form of a Presidential Fellowship provided by the Graduate School, OSU.

Finally, I want to thank my parents for their encouragement and support.
VITA

November 30, 1950 ............... Born - Preston, England
1956 .............................................. Immigrated to Melbourne, Australia
1973 .............................................. B. Ec., Honors
1973-1974 ...................................... Teaching Assistant, Department of Accounting, Monash University
1975 .............................................. Travelled
1976-1979 ...................................... Teaching Assistant, Department of Accounting, The Ohio State University

FIELDS OF STUDY

Major Field: Accounting
Minor Fields: Economics
Statistics in Psychology
# TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>ACKNOWLEDGEMENTS</td>
<td>ii</td>
</tr>
<tr>
<td>VITA</td>
<td>iii</td>
</tr>
<tr>
<td>LIST OF TABLES</td>
<td>vii</td>
</tr>
<tr>
<td>LIST OF FIGURES</td>
<td>ix</td>
</tr>
<tr>
<td>Chapter</td>
<td></td>
</tr>
<tr>
<td>I. INTRODUCTION</td>
<td>1</td>
</tr>
<tr>
<td>II. LITERATURE REVIEW</td>
<td>15</td>
</tr>
<tr>
<td>III. A CASCADED INERENCE MODEL OF THE MANAGER–SENIOR RELATIONSHIP</td>
<td>45</td>
</tr>
</tbody>
</table>

**Chapter I. INTRODUCTION**
- Purpose of the Study
- Method of Investigation
- The Experiment
- Contributions of the Study
- Organization of the Study

**Chapter II. LITERATURE REVIEW**
- The Audit Team
  - Description of the Sequential Audit Process
  - Description of the Audit Team
  - Judgment in the Context of the Audit Team
- Experimental Auditor Judgment Research
  - Auditor Judgment as Expert Judgment
  - Early Research on Auditor Judgment
  - Directions of Recent Research on Auditor Judgment
  - The Problem Recognized
- Source Credibility: Experimental Psychology Research
- Attitude and Opinion Change Research
- Experimental Cascaded Inference Research
- Source Credibility and Auditor Judgment

**Chapter III. A CASCADED INERENCE MODEL OF THE MANAGER–SENIOR RELATIONSHIP**
- A Bayesian Approach
- The Audit Team: A Bayesian Model
VI DESCRIPTION OF THE RESEARCH METHODOLOGY: A TWO-STAGE EXPERIMENT

A Two-stage Experiment: Justification
Subjects
The First Stage of the Experiment
  Experimental Task
  Independent Variables
  Dependent Variable
  Research Hypotheses
  Statistical Analysis

The Second Stage of the Experiment
  Experimental Task
  Independent Variables
  Dependent Variables
  Research Hypotheses
  Statistical Analysis
  Pretest of the Test Instruments
  Chapter Summary

V STATISTICAL ANALYSIS AND INTERPRETATION OF THE RESULTS

Stage One: Assumptions of the Analysis
Analysis of Subjects' Responses
  Hypothesis 1
  Hypothesis 2
  Hypothesis 3
  Firm Effect
  Subject Effect
Analysis of Difference Scores
  Hypothesis 4
  Firm Effect
  Subject Effect

Stage Two
Assumptions of the Analysis
Analysis of Subjects' Responses
  Hypothesis 5
  Hypothesis 6
  Hypothesis 7
  Hypothesis 8
  Hypothesis 9 and 10
Additional Analysis..........................................................
  Firm Effect.....................................................................
  Subject Effect..................................................................
  Multidimensional Scaling............................................
Subjects Comments and the Post Experimental Questionnaire..........................
  Subjects Comments..........................................................
  Post Experimental Questionnaire.................................
Chapter Summary..........................................................

VI  SUMMARY AND CONCLUSIONS 156

Summary.................................................................
Conclusions.................................................................
  Second Research Objective..........................................
  Third Research Objective............................................
  Fourth Research Objective..........................................
  The Subject Effect....................................................
Implications......................................................................
  Auditor Judgment Research........................................
  Auditing Practice.......................................................
  Experimental Psychology...........................................
Limitations......................................................................
Suggestions for Future Research....................................

APPENDIX

A  A BAYESIAN REPRESENTATION OF THE AUDIT TEAM MEMBERS' JUDGMENTS.......................................................... 182
B  EXTENSION OF THE BAYESIAN MODEL TO INCLUDE A MEASURE OF SOURCE CREDIBILITY................................................. 186
C  STEPS INVOLVED IN CALCULATING THE SUBJECT'S INFERED LIKELIHOOD RATIO AND THE MODEL'S DERIVED LIKELIHOOD RATIO................................................................. 189
D  THE TEST INSTRUMENT FOR STAGE ONE OF THE EXPERIMENT..... 196
E  THE TEST INSTRUMENT FOR STAGE TWO OF THE EXPERIMENT..... 209

BIBLIOGRAPHY................................................................. 226
# LIST OF TABLES

<table>
<thead>
<tr>
<th>Table</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.</td>
<td>SUMMARY OF EARLY RESEARCH ON AUDITOR JUDGMENT</td>
</tr>
<tr>
<td>2.</td>
<td>LIKELIHOOD RATIOS FOR AN INTERNAL CONTROL EXAMPLE</td>
</tr>
<tr>
<td>3.</td>
<td>ANALYSIS OF VARIANCE FOR A TWO-FACTOR REPEATED MEASURES DESIGN</td>
</tr>
<tr>
<td>4.</td>
<td>SUMMARY OF ACCOUNTS RECEIVABLE</td>
</tr>
<tr>
<td>5.</td>
<td>SUMMARY OF ANOVA RESULTS</td>
</tr>
<tr>
<td>6.</td>
<td>SUMMARY OF INDIVIDUAL RESPONSE BEHAVIOR</td>
</tr>
<tr>
<td>7.</td>
<td>FIRM REPRESENTATION WITHIN CLUSTERS</td>
</tr>
<tr>
<td>8.</td>
<td>SUMMARY OF ANOVA BY CLUSTER</td>
</tr>
<tr>
<td>9.</td>
<td>SUMMARY OF ANOVA RESULTS - DIFFERENCE SCORES</td>
</tr>
<tr>
<td>10.</td>
<td>FIRM REPRESENTATION WITHIN CLUSTERS - DIFFERENCE SCORES</td>
</tr>
<tr>
<td>11.</td>
<td>SUMMARY OF ANOVA BY CLUSTER - DIFFERENCE SCORES</td>
</tr>
<tr>
<td>12.</td>
<td>SUMMARY OF THE RESULTS OF THE SECOND-_STAGE EXPERIMENT</td>
</tr>
<tr>
<td>13.</td>
<td>MANOVA RESULTS FOR D1-D</td>
</tr>
<tr>
<td>14.</td>
<td>ANOVA RESULTS FOR E</td>
</tr>
<tr>
<td>15.</td>
<td>ANOVA RESULTS FOR D1</td>
</tr>
<tr>
<td>16.</td>
<td>ANOVA RESULTS FOR D2</td>
</tr>
<tr>
<td>17.</td>
<td>ANOVA RESULTS FOR D3</td>
</tr>
<tr>
<td>18.</td>
<td>ANOVA RESULTS FOR D4</td>
</tr>
<tr>
<td>19.</td>
<td>ANOVA RESULTS FOR G</td>
</tr>
<tr>
<td>20.</td>
<td>ANOVA RESULTS FOR H</td>
</tr>
<tr>
<td>Table</td>
<td>Description</td>
</tr>
<tr>
<td>-------</td>
<td>-----------------------------------------------------------------------------</td>
</tr>
<tr>
<td>21.</td>
<td>BREAKDOWN OF SUBJECTS' RESPONSES INTO CHANGE/NO CHANGE CATEGORIES</td>
</tr>
<tr>
<td>22.</td>
<td>FURTHER BREAKDOWN OF SUBJECTS' RESPONSES</td>
</tr>
<tr>
<td>23.</td>
<td>TEST FOR A FIRM EFFECT ON SUBJECTS' MEAN SAMPLE SIZE JUDGMENTS</td>
</tr>
<tr>
<td>24.</td>
<td>SUBJECTS' SAMPLE SIZE JUDGMENTS COMPARED TO SENIOR'S RECOMMENDED SAMPLE SIZES BY FIRM</td>
</tr>
<tr>
<td>25.</td>
<td>INDIVIDUAL SUBJECT CLASSIFICATION OVER ACCOUNTS RECEIVABLE STRATA</td>
</tr>
<tr>
<td>26.</td>
<td>SUMMARY OF ANOVAS ON SUBJECT GROUPINGS</td>
</tr>
<tr>
<td>27.</td>
<td>SUMMARY OF HYPOTHESIS TESTS</td>
</tr>
</tbody>
</table>
## LIST OF FIGURES

<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.</td>
<td>THE AUDIT AS A SEQUENTIAL PROCESS</td>
<td>17</td>
</tr>
<tr>
<td>2.</td>
<td>EXAMPLES OF AUDIT TEAMS</td>
<td>21</td>
</tr>
<tr>
<td>3.</td>
<td>THE AUDIT TEAM</td>
<td>26</td>
</tr>
<tr>
<td>4.</td>
<td>DEPENDENCIES AMONG AUDIT TEAM MEMBERS</td>
<td>50</td>
</tr>
<tr>
<td>5.</td>
<td>SOURCE CREDIBILITY MEANS BY CONFIDENCE LEVEL</td>
<td>106</td>
</tr>
<tr>
<td>6.</td>
<td>SOURCE CREDIBILITY MEANS BY FIRM, WITH DUNCAN'S MULTIPLE RANGE TEST FOR MEANS</td>
<td>109</td>
</tr>
<tr>
<td>7.</td>
<td>SOURCE CREDIBILITY MEANS BY CLUSTER</td>
<td>114</td>
</tr>
<tr>
<td>8.</td>
<td>MEAN DIFFERENCES AS A FUNCTION OF SOURCE CREDIBILITY FOR BOTH STATISTICAL CONFIDENCE LEVELS</td>
<td>117</td>
</tr>
<tr>
<td>9.</td>
<td>MEAN DIFFERENCES AS A FUNCTION OF SOURCE CREDIBILITY BY CLUSTER</td>
<td>121</td>
</tr>
<tr>
<td>10.</td>
<td>PERFORMANCE RATING MEANS FOR EACH RESPONSE CATEGORY BY GROUP</td>
<td>143</td>
</tr>
<tr>
<td>11.</td>
<td>ALSCAL PLOT OF STIMULI ON TWO DIMENSIONS</td>
<td>147</td>
</tr>
<tr>
<td>12.</td>
<td>MDPREF PLOT OF SUBJECTS ON TWO DIMENSIONS</td>
<td>148</td>
</tr>
<tr>
<td>13.</td>
<td>MEAN DIFFERENCES AS A FUNCTION OF SOURCE CREDIBILITY UNDER TWO DIFFERENT STATISTICAL CONFIDENCE LEVEL ASSUMPTIONS</td>
<td>194</td>
</tr>
<tr>
<td>14.</td>
<td>MEAN DIFFERENCES AS A FUNCTION OF SOURCE CREDIBILITY BY CLUSTER, BASED ON EXPERIMENTWISE CONFIDENCE LEVEL</td>
<td>195</td>
</tr>
</tbody>
</table>
CHAPTER I

INTRODUCTION

Auditing is a human evaluation process to establish the adherence to certain norms, resulting in an opinion or judgment. (Schandl, 1978, p. 4)

In this rare attempt to set out a coherent theory of auditing, Schandl recognizes that audit activity necessitates judgment. Furthermore, the audit process "consists of a chain of interdependent judgments" (Schandl, op. cit. p. 29). The pervasiveness of judgment in the audit process has long been recognized by the accounting profession. Yet a problem exists in how to obtain insight into the nature of auditor judgment. The 1972 American Accounting Association Committee on Basic Auditing Concepts said of the audit task that "the mental leap from simple evidential propositions to broad generalizations is supported only by the vaguest system of inference." (Committee on Basic Auditing Concepts, 1972, p. 55).

In an attempt to clarify this vague system of inference, judgment-related auditing research has increased considerably since 1972. At least three factors have contributed to the growth of this research, and also to its importance. The first factor is the common finding in previous studies of fallibility in and differences among auditors' judgments. In particular, Weber (1977) finds that auditors do not generally meet
the psychometric criteria indicative of expert judgment as specified by Einhorn (1974). This is a potentially serious failing considering that the professional auditor "can be held to standards of performance and judgment somewhat higher than those of an ordinary person." (Mautz and Sharaf, 1961, p. 131). The second factor is the increased litigation brought against auditors. For example, in June of 1968, Fortune estimated that as many suits were filed against auditors in the previous 12 months as in the previous 12 years (Peat, Marwick, Mitchell and Company, 1976, p. 88). The increased litigation has resulted in a higher standard of care in auditing (Causey, 1976).

At the same time that the quality of auditor judgment is being questioned, auditors are under pressure to extend the range of their judgments. The demand for extending the scope of the audit function is the third factor creating an interest in auditor judgment. The increased litigation brought against auditors in recent years has also resulted in a broader scope of duty (Causey, op. cit.). This trend continued with the Report, Conclusions, and Recommendations of the Commission on Auditors' Responsibilities (Commission on Auditors' Responsibilities, 1978). These three factors imply significant interest in how auditor judgments are made and how well they are made.

While the concern with auditor judgment has significantly increased auditor judgment research, this research has concentrated on the individual auditor. But an audit is usually conducted by an audit team. Recognition of this fact allows the division of labor and resulting dependencies that occur between audit team members to be considered. One such dependency is between auditors' judgments. An auditor will often
form a judgment on the basis of audit evidence compiled and possibly analyzed for him by a subordinate auditor. The inferential value (or information content) of the audit evidence will in part be a function of the reliability or credibility of the subordinate auditor.

The fallibilities in judgment and the significant judgment differences found in previous research suggest that the credibility of the information source is a factor that, at least implicitly, will enter the auditor's judgment process. In the audit review process, the audit team members' credibility is of central concern. A consequence of this audit team perspective is that if the auditor is sensitive to the credibility of his subordinates, an inaccurate judgment made at one level may be compensated for at a higher level within the hierarchical structured audit team. Such a possibility has been recognized in a different context by Cyert, March and Starbuck (1961). Yet Mock and Turner (1979b), in the only existing study to consider the existence of the audit team, found that the review of the auditor's work by a superior did not affect the outcome (sample size recommendations). This result occurred despite the existence of large differences between auditors' initial recommendations.

The potential consequences of deficiencies in auditor judgment, such as economic inefficiencies in conducting the audit, losses by investors, litigation and even outside regulation of the audit function, make continued research into auditor judgment desirable. This study addresses a problem which is largely ignored in existing auditor judgment research: the impact of audit team dependencies on auditors' judgments.
Purpose of the Study

The credibility of audit team members should be of concern to the auditor once it is recognized that, within the audit team, an auditor often forms a judgment on the basis of work performed by another team member. The dependency between judgments of this form is studied in psychology through the concept of source credibility.

The concept of source credibility in psychology refers to the accuracy, reliability or credibility of the sender (source) of a message to a receiver who does not directly observe the occurrence of the event reported in the message. Hence the receiver's concern with the credibility of the sender. Note that no distinction is commonly made in the psychology literature between source credibility and source reliability, nor will a distinction be made here. Under a strict psychometric definition, reliability would not include judgment bias. Such bias would, however, affect the credibility of the source. Furthermore, the concept of source credibility should be considered sufficiently broad to encompass not only the source's competence, but also various motivational factors which affect his veracity and biases.

The purpose of the study is to employ the psychological concept of source credibility to examine the judgmental dependencies existing in the audit team; in particular, to determine whether source credibility is a factor that enters into the auditor's judgment and, if so, what effect it has. In an auditing context, a subordinate auditor is the sender; he conducts an audit procedure and reports the results to a more senior auditor. Source credibility refers to the senior auditor's belief (subjective probability) that the procedure has been conducted and
reported accurately. A formal probabilistic definition of source credibility as applied to an auditing context is provided later. The advantage of adopting the source credibility concept from psychology is that it enables the auditor's dependence on his subordinate's credibility to be studied in isolation. The need to deal with the complex interactions that occur within the audit team during the conduct of the audit is avoided.

Within the overall purpose of the study, four specific research objectives are established: (1) to model the auditor's judgment process from an audit team perspective incorporating a measure of source credibility, (2) to test to what degree the auditor's judgment is sensitive to source credibility, (3) to test whether the auditor modifies the audit program to compensate for source credibility, and (4) to evaluate what factors about a subordinate influence the auditor's appraisal of the subordinate's credibility. The purpose of the first objective is to provide a normative benchmark to gauge the potential impact of source credibility on the auditor's judgment process. The purpose of the second objective is not only to discover whether the auditor is sensitive to source credibility, but also to provide information on the extent of the auditor's reaction to different levels of source credibility. That an auditor is sensitive to source credibility does not necessarily imply some action is taken, such as modifying the audit program. The third objective considers this possibility. Finally, the purpose of the fourth objective is to provide additional understanding of the operation of source credibility within the audit team, by attempting to determine the critical factors which affect the auditor's appraisal of the subordinate's credibility.
The focus of the study is on the audit manager and his relationship with the audit senior, who is at the next lower level in the audit team hierarchy. The reason for this focus is that the audit manager is more directly involved than the audit partner in conducting an audit: in making important audit decisions, in resolving conflicts between seniors in continually reviewing the work and judgments of seniors and, when judged necessary, in requiring the expansion of a particular area of the audit investigation. In searching for a point in the audit process where source credibility is particularly important, a logical level to look at is the audit manager.

Method of Investigation

The study applies both an analytical methodology and an experimental methodology to investigate the four research objectives. The analytics entail the use of cascaded inference theory, which is based on Bayes' Theorem, to model the auditor's judgment process. The result is a Bayesian model which incorporates a measure of source credibility. The advantage of a Bayesian approach is that the probabilistic representation of the auditor's judgment explicitly recognizes the inherent uncertainty existing in the audit. The model predicts that small reductions in the credibility of an auditor will lead to large reductions in the informativeness of the audit evidence.

To determine whether auditors, specifically audit managers, are sensitive to the credibility of their subordinates, and to consider the third and fourth objectives, a laboratory experiment is chosen. A laboratory experiment permits manipulation of the variables of interest, randomization of treatments and control of extraneous influences. By manipulating source credibility in the experiment, it is possible to
determine whether participating auditors are particularly sensitive to certain levels of source credibility. In addition, the experiment can include factors which are hypothesized to influence the auditor's evaluation of a subordinate's credibility and formal tests of hypotheses. Due to the complexity of the audit process, the control which can be maintained in the laboratory experiment is a particularly important advantage.

The above considerations give the laboratory experiment a high degree of internal validity. Often at issue in laboratory experiments is the external validity. Several factors contribute to the external validity of the study. A realistic case was used to present the test instrument to the subjects. The case includes materials normally present in the conducting of an audit. The content of the experimental materials was discussed with practicing auditors and pre-tested before the experiment was conducted. Finally, the subjects in the experiment were practicing auditors.

The Experiment

Thirty five audit managers participated in the experiment. The experiment was presented to the subjects in two stages. The first stage was administered to the subjects, usually individually, in their own offices. The subjects' task was to read the case, and on the basis of the information contained in it, to make probabilistic judgments in internal control evaluation. Included in the information provided were alternative descriptions of the credibility of the senior who conducted the relevant audit procedures. Eight judgments were required in all, each under a different assumption concerning the senior's credibility.
and the statistical confidence level at which the senior conducted the audit procedures. The first-stage experiment was designed to meet the second research objective. Furthermore, it was designed so that the subjects' responses can be compared to the model's predictions.

The experimental materials comprising the second stage of the experiment were left with the subjects, after the administration of the first stage, to be returned to the researcher by mail. Thirty-one subjects returned the completed experimental materials. The same case was used in both stages. In the second stage, the focus was not on internal control evaluation but on a subsequent step in the audit process, confirmation of accounts receivable. The subjects' task was to make sample size judgments, based on the information in the case study, under different descriptions of the senior. The subjects' responses provide evidence on whether the subjects, the audit managers, modify the audit program in response to a change in the audit senior's credibility. The modification was in the form of a change in the sample sizes for the confirmation procedure.

The form of the audit senior's description differed between the two stages. In the first stage, an explicit probabilistic appraisal of the senior's credibility was provided; in contrast, in the second stage only factors pertaining to his credibility were provided. These factors are the senior's experience level, his performance rating and the sampling approach he used. By manipulating these factors, those, if any, considered to be important by the subject in determining the extent the senior is to be relied on were identified. The use of surrogates for source credibility had another advantage. A subject may not respond to the explicit manipulation of source credibility in the first stage because a manager's
judgment in appraising a senior is more implicit than explicit. The second stage goes some way to cover this possibility.

The two stages of the experiment permit direct consideration of the second, third, and fourth research objectives and statistical testing of the related hypotheses. Ten hypotheses are posited; the test results may be summarized as follows:

1. Subjects were found to be sensitive to source credibility. Subjects differed as to when they discounted (recognize the reduction in informativeness of) the results of the audit procedures because of a reduction in source credibility.

2. When subject responses (internal control judgments) were compared with model predictions, the majority of subjects discounted the results for the decrease in source credibility more than the model, while the discounting by the remaining subjects was less than that of the model.

3. The confidence level at which compliance tests were conducted, but not the interaction between confidence level and source credibility, accounted for a statistically significant amount of variance in subject internal control judgments.

4. Subject sample size judgments differed across descriptions of the senior, but not to the extent of the weight the subjects indicated they placed on the senior's recommendation in making their judgments. The sample size judgments did not reflect the full reaction of the subjects to the different descriptions of the senior.

5. The accounting firm the subjects work for was not a statistically significant factor in explaining differences between subjects' internal control judgments, but was a statistically significant factor
in explaining differences between subject sample size judgments.

6. Both the senior's rating in the formalized performance evaluation and the senior's experience level were statistically significant in explaining subject sample size judgments. Of the two, performance rating was more important.

7. The sampling approach, statistical or judgmental, employed by the senior was not a statistically significant independent variable in explaining subject sample size judgments. Subjects did differ, however, in their preference for statistical versus judgmental sampling.

**Contributions of the Study**

The contributions of the study are grouped under three headings: auditor judgment, audit practice and experimental psychology. How auditors go about making the complex judgments required in conducting the audit has recently been the focus of widespread attention. The study of source credibility within the audit team is considered an important extension of existing research into auditor judgment. Source credibility is an example of context effects which have been shown to impact upon judgment (Swieringa, Gibbins, Larsson and Sweeney, 1976; Gibbins, 1977); auditor judgments are made in the context of the audit team. This study represents a first step in the analysis of auditor judgment from the perspective of the audit team. A related contribution is the use of a probabilistic approach to analyze source credibility. This approach confronts a key characteristic of auditing, the auditors' operation in an uncertain environment. A probabilistic approach has been used previously in deriving normative models of parts
of the audit process. However, this research not only models part of
the audit process but empirically tests the interesting propositions
derived from the modelling process. Since no sufficiently comprehen-
sive theory of auditing exists to provide a standard for evaluating
empirical auditing research (Schandl, op. cit., Weber, op. cit.), this
is noteworthy.

Adopting the team perspective enables some of the previous find-
ings on auditor judgment to be put in a new more relevant perspective.
While a relationship between internal control evaluation and the extent
of the audit plan proposed is hypothesized, the evidence is mixed
(Weber op. cit., Mock and Turner op. cit.). Since the study finds audi-
tors adjust the audit plan in response to changes in the quality of
compliance test information, a relationship is demonstrated.

The study also provides evidence on the impact of the lack of con-
sensus and other questions raised about auditors' judgments. Studies
on auditor judgment often conclude by stating that the impact of the
auditors' judgment differences found is a function of many considera-
tions (Joyce, 1976, Mock and Turner op. cit.). The audit team perspec-
tive is one such consideration. The sensitivity to source credibility
exhibited by the participating managers indicates that they are sensi-
tive to differences in the ability of subordinates and in the quality
of information provided by them. Furthermore, if the managers adjust
sample sizes in response to different descriptions of the subordinate,
questionable judgments by auditors may be compensated for at a higher
level within the audit team. The sequential audit process provides
many opportunities for such compensation.
The study provides several implications for audit practice. Managers in the study were found to be sensitive to source credibility. This result is reassuring for accounting firms. The impact of study findings on accounting firms can be considered in relation to audit risk. One component of overall audit risk is the quality of the audit work performed. Manager sensitivity to source credibility suggests that they recognize this component. In respect to statistical testing, audit risk is assessed through $\alpha$ and $\beta$ error rates. While $\beta$ is not always directly considered in audit sampling, the results of a $\beta$ error are generally more significant. Manager behavior can be argued to reflect concern for $\beta$.

The study also highlights the role of auditor training and communications as means of reducing audit risk. Improved training of auditors should have the effect of increasing their source credibility. The model developed suggests that, even at initially high levels of source credibility, subsequent increases in source credibility have a significant affect on information content. One aspect of auditor training should be to instruct auditors in how to evaluate their subordinates. Finally, a consequence of the division of labor within the audit team is the importance of communication. So that an auditor is familiar with those he works with, communication is important across as well as within audit teams.

While the concept of source credibility is obtained from the psychology literature, the research also contributes to experimental psychology. The study provides an application of cascaded inference theory, which has undergone considerable theoretical development but
little practical application. In addition, the study uses expert judges rather than student surrogates in a realistic task setting. In psychology there is a call for greater realism in experimental designs (Petrinovitch, 1979). The second stage experiment, because of the same subject and task considerations, contributes to attitude change and opinion change research. While some results of the psychology studies are replicated, others are not.

Organization of the Study

The next chapter presents the literature review. Literature which supports the view of the audit process from the perspective of the audit team is presented first. This perspective is used to review the accounting literature on auditor judgment and the psychology literature on source credibility. Chapter III models the audit team and the audit manager's judgment process within the audit team. A measure of source credibility is included in this model. In an example, the model is applied to internal control evaluation. On the basis of the model and the literature review in chapter II, the research hypotheses are specified. Chapter IV on the research methodology employed describes the experimental environment and tasks used in the two stages of the experiment. It defines the independent and dependent variables and their respective levels. The statistical tests used to analyze the experimental results are also described. Chapter V presents the results of the two stages of the experiment. Chapter VI, the final chapter, summarizes the study in terms of the results of the hypothesis tests and their relation to the four research objectives. Finally,
limitations of the study as well as areas for future research are discussed.
CHAPTER II

LITERATURE REVIEW

This chapter reviews the auditing literature relevant to the study of expert judgment in the audit team, and also reviews the psychological literature pertaining to source credibility. The purpose of the chapter is: (1) to determine a relevant model of the audit investigative process for the research, (2) to justify the problem addressed, and (3) to justify the research approach selected. The chapter is organized into three sections. The first section describes the audit process and the functioning of the audit team. The second section reviews the experimental auditor judgment research. The third section reviews the experimental psychological research dealing with source credibility.

The Audit Team

Description of the Sequential Audit Process

The function of an audit may generally be stated as carrying out an examination of the client's financial statements sufficient to enable the auditor to express an opinion on them [American Institute of Certified Public Accountants (AICPA) 1977, p. 5]. In turn, the audit process is typically described as a sequence of processes, beginning with a study of the environment and ending with the issue.
of the audit report, which presents the auditor's opinion on the client's financial statements. Figure I, adapted from Arens and Loebbecke (1976), shows the main elements in the sequence.

The majority of model related audit studies have focused on a particular element of the audit process. In their review of audit research, Lin, Mock, Newton and Vasarhelyi (LMNV) (1978) identify three elements which have been researched: the evaluation of internal controls, statistical sampling as used in compliance and substantive testing, and substantive tests in the form of analytical review and tests of detail. Some aspect of decision theory underlies all these research efforts. The authors conclude that further research is required and that a problem with existing auditing research is that "researchers have pursued rather narrow research objectives and too often have failed to consider where their results fit into the overall field of auditing" (LMNV p. 1).

Three studies (Kinney, 1975, Bailey and Jensen, 1978, and LMNV op. cit.) explicitly recognize the sequential nature of the audit process. Kinney uses decision theory to integrate the role of internal control system design on compliance tests and tests of detail in audit planning. He concludes by questioning the economic feasibility of extending the model beyond the single account analyzed. Bailey and Jensen demonstrate a completely Bayesian decision theoretic model that links compliance and substantive tests. A Bayesian linkage is said to offer three advantages. First, consistency of beliefs throughout the part of the audit process modelled is assured. Second, variables often treated as exogeneous are shown to be endogeneous in
FIGURE 1
THE AUDIT AS A SEQUENTIAL PROCESS
the model relating compliance and substantive tests. Third, other relevant variables, such as management override, can be incorporated in extensions of the model. The authors also question the feasibility of the model. As the number of decision stages increases and as multiple testing is recognized, "the calculation of the expected costs required for the selection of the test levels at all stages may become intractable" (Bailey and Jensen, p. 30). LMNW extend a model initially proposed by Warren (1975) to integrate previous research efforts into an overall model of the audit process. The model is:

\[ R = 1 - (1-C)^x(1-A)^y(1-D)^z \]

Where

- \( R \) = Overall audit reliance
- \( C \) = Internal control reliance
- \( A \) = Analytical review reliance
- \( D \) = Substantive test reliance
- \( x, y, z \) are weighting parameters.

The model is concerned with testing the (main) hypothesis that there is no material error in the presentation of the financial statements.

About the model they say:

"The assignment of the weights \( x, y \) and \( z \), the role of human information processing heuristics in the estimation of \( C \) and possibly \( A \) are some of the key questions still in need of further research. It is important, however, that the decision theory approach used in the estimation of \( D \) be integrated in the human information processing model for the formulation of the test of the main hypothesis." (LMNW, p. 33)
The authors emphasize the relationship between audit procedures and the relationship between human information processing and decision theory approaches in auditing research. The credibility of the auditors who have conducted the various audit procedures may be a factor entering into the estimation of the weighting parameters. The authors make no reference to this possibility. IMNV's study, like the other studies referenced, looks at the related procedures in the audit process from the perspective of a single auditor. This is not a criticism of the studies as they are concerned with analyzing the interdependencies existing between audit procedures. Nevertheless, the fact that an audit is usually conducted by several auditors is a cause itself of interdependencies between auditors' judgments.

Description of the Audit Team

The first standard of field work in Codification of Statements on Auditing Standards (SAS) recognizes that an audit may be conducted by more than one auditor:

"The work is to be adequately planned and assistants, if any, are to be properly supervised." (AICPA op. cit. p. 33)

The number of auditors making up the audit team which conducts an audit is a function of the size and nature of the client, and the type of accounting firm. Examples of audit teams are presented in Figure 2. The examples in Figure 2, together with the following descriptions of functional levels, are taken from Anderson (1977). As Anderson is Canadian, the Certified Public Accountant (CPA) title is replaced by the Chartered Accountant (CA).
1. Reviewers — responsible for participating in the planning, for conducting the review, and for approving the issue of the audit report (their titles may be partners, client partners, audit partners, principals, managers, or supervisors and usually, in medium size and large practices, at least two review levels are involved);

2. Audit seniors — responsible for participating in the planning, for conducting and/or supervising the field work, for reviewing the work of their assistants, and for reporting the results to the reviewers (their titles might be staff seniors, senior staff accountants, etc. and they might or might not have already obtained their CA certificates);

3. Audit assistants — responsible for conducting assigned sections of the field work under the supervision of the audit senior, each assistant also being responsible for supervising the work of any assistants under him (the titles of audit assistants might be staff accountants, intermediates, juniors, etc.; usually they would be students-in-accounts each working towards a CA though some of them might be audit technicians not working towards professional designations);

4. Specialists — responsible for planning, conducting, reviewing or merely consulting on specialized areas such as taxes, computer auditing, testing, regulatory requirements (they may be full-time specialists or audit partners or managers with part-time specialist responsibilities). (Anderson, p. 402)

The functional titles may differ across accounting firms, but the descriptions of the levels are quite general. Note that in all but small accounting firms there are two distinct levels beyond the audit senior: the audit manager and the partner. The partner has final responsibility for the audit; he signs the audit report. The manager usually is in charge of the actual conduct of the audit and does
<table>
<thead>
<tr>
<th>CLIENT</th>
<th>SMALL FIRM</th>
<th>REGIONAL FIRM</th>
<th>NATIONAL FIRM</th>
</tr>
</thead>
<tbody>
<tr>
<td>RETAIL STORE</td>
<td>1 partner</td>
<td>1 partner</td>
<td>1 manager</td>
</tr>
<tr>
<td>(one location)</td>
<td>1 manager</td>
<td>1 manager</td>
<td>1 audit senior (non-CA)</td>
</tr>
<tr>
<td>100 hour audit</td>
<td>1 audit senior (non-CA)</td>
<td>1 partner</td>
<td>1 audit senior (non-CA)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>1 manager</td>
<td>1 audit senior (non-CA)</td>
</tr>
<tr>
<td>SMALL MANUFACTURER</td>
<td>1 partner</td>
<td>1 partner</td>
<td>1 partner</td>
</tr>
<tr>
<td>(one location)</td>
<td>1 audit assistant (technician)</td>
<td>1 partner</td>
<td>1 manager</td>
</tr>
<tr>
<td>300 Hour audit</td>
<td></td>
<td>1 manager</td>
<td>1 audit senior (CA)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>1 audit senior (CA)</td>
<td>1 audit senior (CA)</td>
</tr>
<tr>
<td>LARGER MANUFACTURER</td>
<td>1 partner</td>
<td>1 partner</td>
<td>1 partner</td>
</tr>
<tr>
<td>(one factory plus ten retail outlets in two provinces)</td>
<td>1 audit senior (CA)</td>
<td>1 manager (serving as senior)</td>
<td>1 manager</td>
</tr>
<tr>
<td>1,000 hour audit</td>
<td>2 audit assistants (CA's)</td>
<td>3 audit assistants</td>
<td>1 audit senior (CA)</td>
</tr>
<tr>
<td></td>
<td>(occasional consultation with outside computer specialist)</td>
<td>one a CA)</td>
<td>3 audit assistants</td>
</tr>
<tr>
<td></td>
<td></td>
<td>1 tax specialist</td>
<td>1 tax specialist</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>1 computer audit specialist (CA)</td>
</tr>
</tbody>
</table>

FIGURE 2. EXAMPLES OF AUDIT TEAMS
22.

most of the reviewing. Regarding the proficiency of each auditor SAS states:

The nature and extent of supervision and review must necessarily reflect wide variances in practice. The auditor charged with final responsibility for the engagement must exercise a seasoned judgment in the varying degrees of his supervision and review of the work done and judgment exercised by his subordinates, who in turn must meet the responsibility attaching to the varying gradations and functions of their work. (AICPA op. cit., p. 233)

The audit process can be described as a sequence of procedures over time. Once the existence of the audit team is recognized, the audit process can also be considered as a sequence of interactions between the members of the audit team who perform the audit. As described by Anderson, the audit team members are distinguished by functional levels; there is a division of labor within the audit team. There is a movement of responsibility and instructions down the hierarchical levels of the audit team, and an upward flow of information that has been collected, organized and analyzed. Whether for the final audit opinion or an intermediate judgment, the auditor uses subordinate team members to make his task more manageable; the audit team and consequent division of labor enables auditors to conduct today's large scale audits.

Note that the existence of this organizational form is not surprising. In analyzing the relationship between types of economic markets and alternative organizational structures, Williamson (1975) demonstrates that simple hierarchy is the more efficient manner of group organization. It allows for specialization of judgment and decision making and economy in communication costs. It allows for adopting to uncertainty in a coordinated way and, through audit and
review, permits narrowing of the information gap between team members.

Research on the audit team per se is limited. Watson (1975) considers whether a management services team faces greater task environment uncertainty than an audit team and whether the greater the uncertainty the more formalized the team structure. He finds that the management services environment is less certain than the auditing environment. The evidence in respect to the degree of the formalized structure of the teams is mixed. However, the audit team is found to have a hierarchical structure in terms of the work relations between team members. This finding is consistent with Anderson's descriptions. Two recent studies (Jiambalvo, 1979, and Blocker, 1979) recognize the existence of the audit team. Because they are limited to means of evaluating and improving performance of audit team members, particularly audit seniors, they are not relevant to this study and are not discussed further.

Judgment in the Context of the Audit Team

The division of labor places an additional burden on the integration of information within the auditor's judgment process. The auditor needs to determine the capability of a subordinate to whom he assigns an audit procedure and the extent to which he can rely, in making his own judgments, on the results of the work conducted by the senior. In other words, the auditor must evaluate the credibility of the subordinate as an information source (source credibility). If the auditor has the choice over work assignment, the evaluation should be first carried out before a subordinate is assigned a task. Then, when the task is completed and the results reported, the auditor should review
The subordinate's work to determine to what extent the initial evaluation was appropriate, to what extent the subordinate has followed instructions and, in consequence, to what extent he can rely on the results.

One aspect of this evaluation is the subordinate's competence. Schandl (op. cit.) includes "the principle of competence of the judge" as one of the five principles of audit judgment he considers. The five principles of judgment are:

1. validity of the judgment
2. competence of the judge
3. independence of the judge
4. completeness of the judgment
5. communicability of the judgment.

He describes competence in judgment as:

the knowledge, experience, and ability to select the proper norms, apply them correctly to the well chosen evidence, and come up with a good or valid judgment. (Schandl, p. 189)

The evaluation of a person's competence is said to be a matter of opinion as there is "no absolute yardstick for measuring it." (Ibid.)

The principle has two applications:

The first one refers to the internal quality of the person making the judgment. The second refers to the belief or trust in the competence of the judge by anybody who accepts the judgment and thereby renders it socially effective. (Ibid.)

The second implication refers to external users of the audit report and also to members of the audit team. In fact, Schandl says that the principle takes on added importance in "group audits." So the second implication becomes what is called the source credibility issue in this study. It can be hypothesized that the more competent the subordinate the greater is his credibility and, consequently, the greater is the
weight which can be given to his work. The concept of source credibility, however, is more inclusive than Schandl's concept of competence. Other factors can be expected to affect credibility, for example, specific client knowledge, as distinct from general competence, subordinate motivation and integrity.

Given the division of labor within the audit team, there are additional considerations which further complicate an auditor's judgment which is based on information provided to him. One is the inability of humans to provide others with all the information that is available to them in making a judgment (Schandl, p. 191). Evidence of this inability has already been found in an auditing context (Mock and Turner, op. cit.). Another is that evidence may be evaluated at a level at which the subordinates cannot discriminate between cues that may become important later in the audit or that may relate to a part of the audit which is not their direct concern. On the other hand, the audit process is structured to minimize these problems. An example of this structuring, and an important part of the interaction within the audit team is the review process. The audit assistant's work is reviewed by the audit senior; the senior's immediate supervisor, normally the manager, reviews the senior's work and also reviews the work of the assistant, but less thoroughly; finally, the partner reviews all working papers, those prepared by the manager more thoroughly than the others (Arens and Loebbecke, op. cit.). All these factors enhance the role of competence or source credibility within the audit process.

The division of labor within the audit team means that an auditor has to rely on the judgments of his subordinates. Consequently, the
subordinate's competence and the auditor's evaluation of the reliability of the information provided to him by the subordinate is relevant to his judgment process. The relationships and judgmental dependencies between members of the audit team are not captured by the description of the audit process as a sequence of procedures over time (Figure 1). They can be captured when the audit team and its hierarchical structure are considered. This alternative view of the audit process is summarized in Figure 3. Use of this alternative view or sequence is more appropriate for considering the research objectives.

---

**Experimental Auditor Judgment Research**

**Auditor Judgment as Expert Judgment**

Auditor judgment is generally recognized as an example of expert or professional judgment. According to SAS: 'The independent auditor must undergo training adequate to meet the requirements of a
professional man" (AICPA op. cit., p. 23). Theoretical studies also recognize the expert judgment required of the auditor (Mautz and Sharif, op. cit., Schandl, op. cit.). Schandl's five principles of judgment listed earlier give some idea of what is entailed in expert judgment. A more quantifiable set of necessary criteria for expert judgment, at least in a psychometric sense, is provided by Einhorn (op. cit.). The three criteria are:

1. Experts should tend to cluster variables in the same way when identifying and organizing cues.

2. Expert judgment should be highly reliable (intragrade reliability), show both convergent and discriminant validity, and be relatively free of judgmental bias when measuring cues.

3. Experts should weight and combine information in similar ways (interjudge reliability or consensus). (Einhorn, p. 562)

Studies of auditor judgment generally assume the expert nature of auditor judgment and use parts, if not all, of Einhorn's criteria in their analysis. A review of auditor judgment research follows.

Early Research on Auditor Judgment

Judgment related studies in auditing consider the audit process from a human information processing (HIP) perspective. Reviews of all but the most recent literature are found in Joyce (1976), Libby and Lewis (1977) and Lin, Mock, Newton and Vasarhelyi (op. cit.). In general, the auditing studies support conclusions obtained in psychological research into judgment; namely, that man's ability to process
information when making judgments is quite limited, even for expert judges (Slovic, 1972). Table 1 presents a summary of this research.

Joyce summarized the evidence as follows:

There is a consistent, albeit small, body of research in auditing that suggests that nontrivial differences exist among auditors in the way they use information in making decisions about the amount of audit work to perform. This consistency is more impressive when one considers it was achieved despite use of considerably different methodologies, subjects, and tasks. (Joyce, p. 53)

He indicated that the worth of reducing the differences found depends on their impact and the costs of reducing them. Both these issues need further research. He also indicated that additional research was needed (1) using more representative designs and (2) incorporating the dynamics of the audit review process. In respect to the latter point, Joyce suggested that the review process "probably reduces judgment variance" but gave no support for this view.

Directions of Recent Research on Auditor Judgment

Recent research on auditor judgment has progressed in three directions: (1) the replication and extension of early research; (2) the study of the nature of information processing characteristics of auditors; and (3) the consideration of ways to improve auditors' judgment performance.

A number of studies have entailed the replication and extension of earlier research. Hamilton and Wright (1977) extended Ashton's study (1974) to specifically consider the impact of auditor experience on internal control evaluation. Seventeen auditors were asked to rate, on a six point scale, 32 combinations of "yes" and "no" answers to five
<table>
<thead>
<tr>
<th>Reference</th>
<th>Research Approach</th>
<th>Audit Problem</th>
<th>Results</th>
</tr>
</thead>
<tbody>
<tr>
<td>Aly &amp; Duboff (1971)</td>
<td>BLM&lt;sup&gt;a&lt;/sup&gt;</td>
<td>Judgment vs statistic sampling</td>
<td>Differences&lt;sup&gt;b&lt;/sup&gt;</td>
</tr>
<tr>
<td>Coreless (1972)</td>
<td>Bayesian</td>
<td>Error Distribution in payroll</td>
<td>Differences, Inconsistency, Non-Bayesian</td>
</tr>
<tr>
<td>Kinney &amp; Ritts (1973)</td>
<td>BLM</td>
<td>Inventory sampling</td>
<td>Differences</td>
</tr>
<tr>
<td>Ashton (1974)</td>
<td>BLM</td>
<td>Internal control strength</td>
<td>Consensus (.7), Consistency (.81)</td>
</tr>
<tr>
<td>Boatsman &amp; Robertson, (1976)</td>
<td>BLM</td>
<td>Materiality</td>
<td>High predictability</td>
</tr>
<tr>
<td>Felix (1976)</td>
<td>Bayesian</td>
<td>Error distribution in payroll</td>
<td>Positive effect of training</td>
</tr>
<tr>
<td>Joyce (1976)</td>
<td>BLM</td>
<td>Extent of substantive testing</td>
<td>Differences (.37), Consistency (.86)</td>
</tr>
<tr>
<td>Moriority &amp; Barron (1976)</td>
<td>BLM</td>
<td>Materiality</td>
<td>Differences</td>
</tr>
<tr>
<td>Hofstedt &amp; Hughes (1977)</td>
<td>BLM</td>
<td>Materiality</td>
<td>Differences</td>
</tr>
</tbody>
</table>

<sup>a</sup>BLM stands for the Brunswic Lens Model type approach. For a description of this and the Bayesian approach see Report of the Committee on Human Information Processing, AAA, 1977.

<sup>b</sup>Differences should be interpreted as the opposite of consensus.
items on an internal control questionnaire. From these responses, a multiple linear regression model was constructed for each subject. The overall results of their study were comparable to those of the Ashton study with average consensus of .66 and average proportion of the variance explained by the five factors being 82 percent. The authors hypothesized that experience would be positively correlated with measures of consensus, predictability and self-insight. A significant positive correlation was found only with respect to consensus. Subjects were divided into two groups: those with three years or less of auditing experience and those with more than three years of experience. The group with more experience had greater consensus (correlation coefficients of .777 versus .621).

Ashton and Kramer (1978) replicated the Ashton study with a group of thirty auditing students. They found more unexplained variance in students' judgments than in judgments by practicing auditors. On average, the separation of duties factors accounted for only 40 percent of the variance in the students' judgments, compared to 51 percent for the auditors. The student consensus (.66), however, was similar to that of the auditors (.70).

Ashton and Brown (1979) extended Ashton's study by including two additional questions on the internal control questionnaire, and the eight questions were presented to the 31 auditor subjects in two different randomly assigned orders. The purpose was to test for an order of presentation effect. Each auditor rated 160 cases and an ANOVA model for each auditor was constructed. The researchers found that the three separation of duties factors accounted for 51 percent of the judgment.
variance. Only one of the eight factors explained significantly more judgment variance in one order of presentation than in the other. The average consensus found (.67) was consistent with that in the previous studies.

Weber (op. cit.) looked at accuracy and consensus aspects of auditors' evaluation of an internal control system's reliability. Of note in this study was the detailed case material given to the auditor subjects. A lack of consensus was found in carrying out the assessment of overall system reliability. Psychological and demographic variables examined had only a perversive influence on auditor decision making, but both increasing audit experience and decreasing risk-taking propensity were positively correlated with more extensive audit programs. Finally: "This research joins other research in questioning the nature of the relationship between the evaluation of internal control and the substantive testing plan proposed" (Weber, p. 188). Weber suggested future research efforts have a theoretical base and concentrate on the manipulation of task variables rather than the investigation of psychological and demographic variables.

Mock and Turner (1979a, 1979b) studied the effect of changes in internal controls and differences in guidance on audit program decisions. As Weber did, they attempted to use a representative design by providing the auditor subjects with extensive case material. Subjects responded to improvements in internal controls by reducing the planned sample sizes in the audit program. Furthermore, improvements from "weak" to "strong" controls resulted in smaller sample size selection than improvements from "weak" to "fair" controls. While revisions in
the right directions were made, there was significant variability in program decisions, especially in poor control environments. Demographic variables such as years of experience, specialized training and client-mix experience did not account for a statistically significant amount of judgment variance.

The second direction research has taken is into the nature of information processing characteristics of auditors. This research is based on the work of Tversky and Kahneman (1974). Their basic finding was that subjects do not follow the principles of probability theory in judging uncertain events, but rather use heuristics which sometimes yield accurate estimates and sometimes do not.

Uecker and Kinney (1977) studied the use of heuristics by auditors. The extent to which 120 auditors made systematic errors in the judgment evaluation of sample outcomes, through the auditors' use of representativeness and protectiveness heuristics, was considered. Each subject examined five cases. Seventy-four percent of the subjects made at least one error; 54 percent made at least one error of representativeness and 37 percent at least one error of protectiveness. There was no significant correlation between subjects' experience and number of correct responses. The authors consequently warned against auditors' mixing random selection with judgmental evaluation.

Gibbins (1977) conducted four experiments involving the representative heuristic with 71 auditor subjects. The experimental results provided mixed evidence on the existence of representativeness - induced biases in auditors' judgments. The subjects were generally sensitive to the effects of both prior probabilities and sample sizes. Evidence
of the illusion of validity with respect to the consistency among highly correlated variables was found; that is, subjects' responses were influenced by the consistency while the redundancy due to intercorrelation was ignored. However, the significance of these results is questionable given the nature of the task surrogation employed. The author concluded that "results have been and will continue to be highly sensitive to the specific circumstances (context effects) and measures used" (Gibbins, p. 25).

Mock and Turner (op. cit.) also considered heuristics, particularly, the anchoring heuristic and the halo effect. A realistic experimental design was utilized. Subjects' planned sample sizes were analyzed with regard to their dependence on last year's audit results (anchoring) and their dependence on the current audit's findings on other correlated internal controls (halo effect). Both phenomena were found to be present, although neither their significance nor effect was determined. If significant, the authors hypothesized that appropriate changes in the design of audit procedures may be desirable.

The third direction of relevant research involves the consideration of ways to improve individual judgment performance. In his study, Weber (op. cit.) incorporated a simulation decision aid which provided an estimate of the dollar error in the inventory account balances produced by the client's accounting system. The purpose of the decision aid was to assist in substantiating inventory. While the decision aid increased judgment accuracy, it had no effect on the audit plan proposed by the subject.
An alternative to a decision aid is group judgment. Reckers and Schultz (1979) used a choice shift framework setting to study the effects of several cues on individual versus group-assisted contingency loss evaluations by auditing students. To suppress risk in these types of judgments, the authors suggest certain forms of consultation (e.g., advisory consultation to be done by telephone and face-to-face consultation carry with it decisive authority) and that the use of groups leads to more consistency in audit judgments. These are strong suggestions given the use of student surrogates.

Mock and Turner (op. cit.), in the only study to explicitly recognize the existence of the audit team, established audit manager - audit senior teams wherein a manager reviewed the senior's work. The purpose of this part of their study was to see the effect of guidance, in the form of manager review, on seniors' sample size recommendations. No significant differences were found in the before and after sample size recommendations. Given the large differences among seniors in the initial recommendations, this was a surprising result. The authors cite two possible factors contributing to this result. The first was the lack of comprehensiveness in seniors' rationale memos supporting their sample size recommendations. This lack of documentation might have made it difficult for the managers to make a meaningful review. The second factor concerns task complexity, about which the authors state:

Perhaps the most pervasive, yet general, finding that arose both from the experimental and literature review phases of this study concerns the significant complexity involved in internal control evaluation. This conclusion is valid even from the limited perspective of an external auditor reviewing controls
purely as an input into audit program design. Internal control reviews with more general objectives would seem to exhibit even greater complexity. (Mock and Turner, 1979b, pp. 223-224)

The Problem Recognized

In summary, the literature shows that there are differences (1) between the judgments of individual auditors (lack of consensus) and (2) between auditors' judgments and normative models. These differences remain when elaborate attempts at task surrogation, called for by Joyce, are made to convey greater task realism to subjects. The evidence on whether psychological and demographic variables, including experience and training, can explain these differences is mixed. Research into how auditors make judgments has focused on the use of heuristics in judgment. The primary finding is that the results are sensitive to the context effects surrounding the judgment.

One such context effect in auditor judgment is the audit team. The first section of this chapter described how, in an audit conducted by an audit team, auditors' judgments depend on work carried out and judgments made by fellow team members. The literature review indicates that the context effect provided by the audit team has not been considered as yet; its consideration in the light of auditor judgment differences is the purpose of this study. The study of Mock and Turner (op. cit.) became available after this study was begun. Their finding that the audit review process had no apparent effect further questions the relationship between audit team members and adds to the timeliness of this study.
The existence of fallible judgment and judgment differences between auditors suggests that credibility of subordinate team members should be of concern to the auditor. To review the research conducted on source credibility, the psychological literature has to be consulted.

Source Credibility: Experimental Psychological Research

Attitude and Opinion Change Research

In experimental psychology, source credibility is considered in attitude and opinion change literature. This literature goes back many years; a review of the early research is provided by Havland, Janis and Kelly (1966). They organize the results according to the nature of (1) the communicator, (2) the communication, and (3) the audience. The results concerning the communicator are relevant. Among these results is the finding that communications from low credibility sources compared to those from high credibility sources are appraised as more biased and unfair by subjects. Information tests revealed equally good learning of what was said regardless of source credibility so that the effects on opinions were not the result of differences in the amount of attention or comprehension.

More recent studies are found in a variety of psychology journals. Source credibility is expressed as a quantitative independent variable, by means of some form of probability estimate of correct communication, or a qualitative independent variable. As a qualitative variable, source credibility has been expressed in terms of expertise (a professional or specialist's opinion versus that of the average man in the street), the source's familiarity with the person/object under study,
and general statements of public opinions. While all the recent studies find source credibility a significant factor affecting opinion change, there are conflicting results about how it interacts with other independent variables (Rhine and Severance, 1970).

Rhine and Severance (op. cit.) claim that a cause of the apparent conflicting results is that most studies employ only two of the three factors of interest: discrepancy between communication and subject's opinion, source credibility, and ego-involvement which is the importance of the issue to the subject. When the authors experimentally manipulated all three factors, they found source credibility significantly affected attitude change, although to a lesser extent when ego-involvement was high. The greater the discrepancy between the communication and the subject's initial opinion, the greater the attitude change. The interaction between source credibility and discrepancy was only significant at the .10 level. Other studies have found no such interaction (Tong-He Choo, 1964).

To give an idea of the typical experiment employed in this literature, one of the scenarios used by Rhine and Severance is described. For the high ego-involvement case the issue of an increase in university tuition was used. Student subjects' attitudes to this issue were measured before and after the receipt of the view of either a high or low credible source. The high credible source was a "Yale Professor of Educational Economics" and the low credible source was a "Private First Class, United States Army." A pilot test was conducted to ensure accurate credibility classifications.
Information integration theory has been used to quantify the weights attached to the variables inducing attitude change. When Rosenbaum and Levin (1968) gave subjects a communication consisting of two contradictory descriptive statements supplied by two different sources identified by occupation, they found that the weight a subject assigned to an informational item of a given value was a function of source credibility. In a follow up paper, Rosenbaum and Levin (1969), the authors used positive and negative descriptions instead of contradictory ones and found that source credibility had a greater impact on positive information than on negative information. Refraining from making a specific explanation of this result, they concluded "all we can say at this time is that the effects of source credibility depend in some way on the values of the information being presented" (Rosenbaum and Levin, p. 37).

Among other studies employing an information integration theory methodology, Birnbaum, Wong and Wong (1976) found that source credibility magnified the impact of the information provided; and that, when two sources were involved, the multiplicative effect of a source's credibility was inversely related to the credibility of the other source. Beach, Mitchell, Deaton and Prothers (1978) found a multiplicative effect for both source credibility and information relevance. Two experiments were conducted in which subjects either assessed the probability of success or the acceptability of hypothetical job candidates. The response mode did not affect the results.

In summary, the unambiguous finding of attitude and opinion change research is that the more credible the information source, the greater
is the attitude change towards the information provided. Other factors, including ego-involvement, communication discrepancy and information relevance also impact upon attitude change.

Experimental Cascaded Inference Research

Another area of experimental psychology that has considered source credibility is cascaded inference. Cascaded inference is founded upon Bayesian statistics; much of the research has been at a theoretical level rather than an empirical level. Cascaded inference theory is considered further in the next chapter when it is used to model auditor judgments within the audit team; the empirical results are considered here.

These studies follow empirical Bayesian studies in utilizing a variance of the "poker chip and bookbag" experiment. As well as manipulating the conditional probabilities concerning the hypotheses of interest, a probabilistic measure of source credibility is included and manipulated in the experiment. Snapper and Fryback (1971) were the first to investigate subjects' probabilistic inferences based on partially reliable reports. Two spinners with attachable discs were used to generate outcomes. The discs had different proportions of two colors. The experimenter selected a disc by the flip of a coin, spun the spinner to determine the outcome (color), spun the second spinner to decide whether he would report the initial color truthfully or falsely, then reported the appropriate outcome. The subjects knew the possible color proportions (conditional probabilities), but could not see which discs were selected. Their task was to provide likelihood ratio estimates about which disc was selected. The procedure was repeated for each
subject a number of times. The subjects responded to reduced report reliability by overestimating the diagnostic impact of a report. In other words, compared to the normative model subjects insufficiently discounted for imperfect source credibility.

Similar results were found by Schum, DuCharme and DePitts (SDD) (1973), Youssef and Peterson (YP) (1973) and Gettys, Kelly and Peterson (GKP) (1973). In the SDD study subjects performed a probabilistic inference task based on their own unreliable observations of flashing visual stimuli. They found subjects' odds estimates excessive compared to the odds provided by the normative model. The authors suggest that subjects did not adjust sufficiently for their own unreliability. Furthermore,

The apparently nonobvious feature of inferences based on unreliable observations is that the impact in observations of events with large inferential impact should be degraded more drastically, for a fixed reduction in observational reliability, than the impact in observations of events with lower inferential impact. (SDD, p. 318).

YP compared subjects' cumulative posterior log odds revision under cascaded and noncascaded conditions both to each other and to the normative model's values. Subjects' odds were slightly excessive at lower likelihood values and became increasingly conservative for higher values. This finding can be reconciled with that of SDD. The posterior odds estimates for subjects under cascaded conditions were consistenly greater than those under noncascaded conditions with equivalent inferential impact ratios. The insufficient discounting for source credibility was offset by excessive Bayesian revision (conservatism).

The primary finding of research on human inferences is that opinion revisions are conservative when compared to those of Bayes' theorem.
The above referenced studies suggest this conservatism does not apply to source credibility; rather, an opposite tendency exists. To explain this phenomena, GKP refer to a "best guess" hypothesis. Under this hypothesis, the subject first infers the posterior odds ignoring credibility then reduces the odds to reflect the source's lack of credibility. While their evidence supported this hypothesis, evidence provided by Funaro (1975) did not. However, all these studies do find that source credibility impacting upon subjects' judgments, although to a lesser extent than a theoretical performance standard suggests.

Finally, Cyert, March and Starbuck (1961) considered source credibility or more specifically bias in an organizational setting. They established three-person experimental organizations where the third person made a decision for the group based upon information provided by the first two persons. Informational bias was introduced by biasing the individual group member's payoff function. The third group member anticipated the bias in the information he received due to his sources' biased payoffs; for every bias there was a bias discount. They concluded that members of an organization consider the whole environment (estimates, information, communication with corresponding biases, payoffs and decision context) as part of their decision situation.

Source Credibility and Auditor Judgment

In discussing the functioning of the audit team earlier in this chapter, the dependencies between judgments of audit team members were recognized. These dependencies were described as context effects of auditor judgments within the audit team. When making a judgment based on work conducted by a subordinate, the auditor either explicitly or
implicitly evaluates the subordinate in relying on the subordinate's work. The judgment differences found among auditors in existing research imply that this evaluation should be a critical input into the auditor's judgment process. The doubt cast on auditors' ability to explain the rationale for their judgments makes this evaluation all the more important.

The auditor's evaluation of the subordinate fits the source credibility scenario investigated in experimental psychology. The auditor is the receiver of the message from the subordinate auditor, the sender. The message contains the results of the audit procedure, conducted by the subordinate, which the auditor requires to make an audit judgment. The credibility of the sender affects the informativeness of the message. Note that credibility is the term commonly used in the psychology literature. Appropriately defined, other terms such as reliability and competence can be and have been used. This study will follow the psychology literature and use the term source credibility to describe the result of the auditor's evaluation of the subordinate.

An advantage of the source credibility approach, in the study of how the audit team context affects the auditor's judgment, is that it enables the effect of the subordinate's credibility on the auditor's judgment to be studied in isolation. Following the psychology experiments, the receiver, the auditor, is the only subject required, while a hypothetical information source, the subordinate auditor, is manipulated. This means that actual audit teams are not required. This is a significant practical consideration. In addition, it permits greater internal validity. Mock and Turner (1979b) did use actual two-member audit teams
to study the impact of the audit review process on auditor judgment. They found no effect. Experimental design problems, not acknowledged by Mock and Turner, could have contributed to this finding. There was no control over the credibility of the subordinate auditors used as subjects, and a selection bias may have been present. Possibly of more importance was the lack of motivation for the managers to adjust the seniors' recommendations. In fact, the managers may have been motivated not to make any adjustment because they did not want to publicly contradict the seniors' judgments. Such counfounding is minimized, if not avoided, by using a hypothetical senior as the source whose credibility is at issue.

The source credibility literature finds that the receiver will adjust the weight he gives to the message as a function of the sender's credibility. This study's four research objectives are designed to examine the operation of source credibility in the audit context. To gauge the potential impact of the source credibility issue in auditing, the first object is modelling the auditor's judgment process so as to include a measure of source credibility. Whether the judgment dependencies existing in the audit team, and the resulting issue of source credibility, are recognized by auditors is captured by the second objective of the study; namely, to test to what degree the auditor's judgment is sensitive to source credibility. The study's third objective, to test whether an auditor adjusts an audit program to compensate for source credibility, considers the common finding in previous research that, while the independent variables manipulated may affect the auditor's judgment, the audit plan recommended is not necessarily affected.
The mixed evidence on the ability of auditor characteristics to explain judgment differences raises the issue of how an auditor would evaluate the credibility of a subordinate auditor. The fourth objective, to evaluate what factors about a subordinate influence the auditor's appraisal of the subordinate's credibility, examines the issue.
CHAPTER III

A CASCADED INFERENCE MODEL OF THE MANAGER–SENIOR RELATIONSHIP

This chapter presents the theoretical development of the study. The first research objective is addressed. This objective is to include a measure of source credibility in a model of the auditor's judgment process. The purpose of the model is to express the process formally, considering the audit team context and the existing judgment dependencies. The impact of the judgment dependencies is considered through the source credibility concept.

As shown in the literature review, the psychology research on source credibility has essentially involved straightforward laboratory experiments without a representative context. The model assists in conceptualizing the relationship between the judgment dependencies existing in the audit team and the role of source credibility. By incorporating a measure of source credibility, the model enables the potential impact of source credibility on an auditor's judgment to be considered.

Finally, the model contributes to the analysis of subjects' responses, obtained from the experimental part of the study, by providing a benchmark in the form of normative responses for comparison.

The first section of the chapter justifies the adoption of a Bayesian approach to the modelling process. A hierarchical Bayesian model is then presented which represents the judgment occurring at each
level within the audit team. The audit manager's judgment is focused upon and explained in terms of the model. While the manager's dependence on the senior is captured by the model, a specific representation of source credibility, which can be manipulated, is not provided. This representation is achieved by utilizing cascaded inference theory. A measure of source credibility is included in the model and its impact on auditor judgment is considered through an internal control example.

A Bayesian Approach

The probabilistic nature of auditing is evidenced by the audit report: it presents the auditor's opinion on the financial statements under examination rather than a statement of fact about them. Biddle and Joyce (1979) recognize the inherent probabilistic nature of auditing and the frequency with which probabilistic assessments of either an explicit or implicit nature are made. The auditor bases nearly all his judgments and consequent decisions on incomplete information. Given that the probabilistic nature of professional judgment has been recognized and studied with respect to meteorology, medical diagnosis and business (Beach, 1975), the probabilistic nature of auditor judgment should be recognized and studied. Furthermore, in psychology there is a questioning of whether behavioral issues are addressed at an adequate level of complexity (Petrinovich, 1979). The recognition of the complexity of behavior, "the vicarious and hierarchical texture of both the stimulus events and the responses available to the organism...forces us to adopt a probabilistic approach if we are to develop explanatory constructs that can be applied to organisms in interaction with their
environment or with other organisms." (Petrinovich, p. 384).

The Bayesian approach follows naturally from the recognition of the probabilistic nature of auditing. The approach uses Bayes' Theorem to express how probabilistic opinions should be revised in the light of relevant new information. Bayes' Theorem is founded on statistical decision theory and specifies internally consistent relationships among probabilistic opinions. Essentially Bayes' Theorem states that an event (sample datum D) is observed at one level and conclusions are then drawn (P(H|D)) about an event at the next higher level (H: hypothesis about the population being sampled). The appeal of this model in auditing is that it provides a means for dealing with the difficulty the auditor faces in relating his prior beliefs to the evidence gained through the audit investigative process (Sorenson, 1969, Corless, 1972). Biddle and Joyce (op. cit.) use the Bayesian approach to provide the following description of the audit judgment process.

1. Determine the overall level of assurance desired from audit procedures.

2. Make prior assessments of the probability (amount) of error in the population.

3. Evaluate the probability of the audit test discovering this error.

4. Gather sample evidence until the desired assurance level is attained.

5. Make a posterior estimate of the probability (amount) of error in the population.

6. Combine priors and test results into an overall assessment.

The ability of Bayes' Theorem to handle successive revisions of opinion enables the sequential judgments that are part of the audit
process to be traced through time. An example of this use of the Bayesian approach, referenced earlier, is the Bayesian model of Bailey and Jensen incorporating both compliance and substantive testing. Variables relevant to the audit process can be included by the extension of the simple Bayesian model. In particular, situational variables such as the experience of subordinates and, as will be developed in this chapter, source credibility can be recognized through the auditor's priors and various probability estimates. The Bayesian approach enables the judgment dependence existing in the audit process to be traced through the hierarchical levels of the audit team.

Another reason for adopting a Bayesian approach is that Bayesian analysis is being applied more and more to model the audit process. Most of this application has concerned either compliance testing or substantive testing. As referenced earlier, larger segments of the audit process are now being modeled, for example, Bailey and Jensen's model. Bayesian analysis is also being increasingly used in judgment related research although, as indicated in the literature review, few researchers have used this approach in an auditing context. Nevertheless, studies in other accounting settings (e.g., Barefield, 1972; Dickhaut, 1973; and Swieringa, Gibbons, Larsson and Sweeney, 1976) have demonstrated the success with which Bayes' Theorem can be used as a standard against which actual judgments can be compared. It is especially useful for examining the impact of information set variables and variables characterizing the decision maker and decision rule (Libby and Lewis, 1977).

There is, however, evidence showing that, for whatever reason, subjects do not process information according to Bayes' Theorem (Slovic,
Fischoff and Lichtenstein, 1977). This result has led some researchers (e.g., Beach, Mitchell, Deaton and Protheris, 1978) to question the prominence given to Bayes' Theorem. But the Bayesian approach has uses other than as a performance ranking device. It provides both a starting point for considering how judgments are made (Einhorn, in press), and a standard to aid in comparing findings of different studies. Both of these uses are particularly relevant for this research, which involves one of the first attempts to introduce source credibility and the team perspective into the study of auditor judgment. In addition, as indicated earlier, the psychology research on source credibility has essentially involved straightforward laboratory experiments without a representative context. Since in this research a specific context, the audit conducted by an audit team, is of concern, judgment and, in particular, source credibility, should be modelled within this context.

The Audit Team: A Bayesian Model

Most real world processes, including audits, are more complex than the single stage revision process described by Bayes' Theorem. Often multiple levels of uncertainty are involved, where events are observed at some lower level and inferences must pass or cascade through multiple levels in the hierarchy to the hypothesis of interest (Gettys and Willke, 1969). This view of the world applies to the judgment sequence present in the audit process. Figure 4 gives a diagrammatic representation of this judgment process.
<table>
<thead>
<tr>
<th>TEAM MEMBER</th>
<th>BELIEF-EVIDENCE INTERDEPENDENCY</th>
</tr>
</thead>
<tbody>
<tr>
<td>ASSISTANT:</td>
<td>Data Collection (D)</td>
</tr>
<tr>
<td>SENIOR:</td>
<td>( \text{PR}_1 \rightarrow D = P_01(H_0</td>
</tr>
<tr>
<td>MANAGER:</td>
<td>( \text{PR}_2 \rightarrow P_0 \rightarrow D = P_02(H_0</td>
</tr>
<tr>
<td>PARTNER:</td>
<td>( \text{PR}_3 \rightarrow P_02 \rightarrow P_01 \rightarrow D = P_03(H_0</td>
</tr>
</tbody>
</table>

**FIGURE 4. DEPENDENCIES AMONG AUDIT TEAM MEMBERS**

A description of the model goes as follows: The audit assistant carries out much of the actual data gathering function. He works under explicit instructions from the audit senior. The audit senior gathers the evidence from the audit assistant (D) and, combining this evidence with the evidence he himself has gathered and his prior beliefs on the matter at issue (\( \text{PR}_1(H_0) \)), forms an opinion (\( P_01(H_0 | D) \)) in respect to the hypothesis (\( H_0 \)) being considered. A description of the investigative process and a statement of the results are contained in the working papers (D). A similar description can be given for the work performed by the manager at the next level of the audit team hierarchy. Finally, the partner carries out the final review, considering the views of the team members under him, the quality of the work done, and the results of the tests made, all of which are documented in the working papers. The audit opinion is then issued. Note that the working papers reflect the dynamic nature of the audit process. As the audit progresses more evidence is added to them. Similarly, as the working papers move between levels of the audit team hierarchy they are changed, either by an
auditor reviewing the work included and thus adding to their verifiability, or by an auditor directing that additional work be conducted in a specific area.

The Bayesian formulation of the belief-evidence interdependency shown in figure 4 is now given; the detailed derivation of the models is explained in Appendix A. Again, the models do not describe how the auditor actually goes about making an audit judgment; the models are merely representations. They are useful in that they provide a consistent probabilistic representation of the combining of the auditors prior beliefs with the results of the audit procedures conducted. They also highlight the dependencies between auditors' judgments. Furthermore, what is especially important for the research objective conducted in this chapter, the models provide the framework for introducing the source credibility factor as an input into the auditor's judgment. The following notation is employed:

\[ D \] : the relevant audit evidence as summarized in the working papers.

\[ P_{R_n} \] : prior probability of the nth auditor.

\[ P_{O_n} \] : posterior probability of the nth auditor.

\[ P_n \] : likelihood or conditional probability of the nth auditor

\[ n = 1,2,3 \] : for audit senior, manager and partner, respectively.

\[ H_i \] : the ith state of nature, where \( H_0 \) is the hypothesized state.

Senior:

\[ P_{O_1}(H_0|D) = \frac{P_1(O_1|H_0)P_{R_1}(H_0)}{\sum_{i=1}^{3} P_i(D|H_i)P_{R_i}(H_i)} \]
Manager:

\[
PO_2(H_0|PO_1,D) = \frac{P_2(PO_1|H_0,D)P_2(D|H_0)PR_2(H_0)}{\sum_{i} P_2(PO_1|H_i,D)P_2(D|H_i)PR_2(H_i)}
\]

Partner:

\[
PO_3(H_0|PO_2,PO_1,D) = \frac{P_3(PO_2|H_0,D,PO_1)P_3(PO_1|H_0,D)P_3(D|H_0)PR_3(H_0)}{\sum_{i} P_3(PO_2|H_i,D,PO_1)P_3(PO_1|H_i,D)P_3(D|H_i)PR_3(H_i)}
\]

The above models provide a decomposition of the variables which enter the auditor's judgment process at any one level of the audit team. As is to be expected, the higher the hierarchical level the more complex the judgment process. In reality, the judgment process is more complicated than that specified because, rather than a single path judgment tree, there will normally be several auditors at the senior level and often more than one manager in the audit team for a particular client. A more complicated hierarchical Bayesian model could incorporate an increase in the number of auditors. Such a model, however, is not required for the purposes of this research. The focus here is upon the relationship between audit team members at different levels, not at the same level, within the hierarchical structure of the audit team. The structure of the audit team provides the setting for subordinate evaluation and, where necessary, adjustment to the audit work completed to date. The judgments involved are reflected in the model through the estimation of the conditional probabilities. The judgment task of the manager is considered next.
The Audit Manager

The focus of the research is on the audit manager, and his relationship with the audit senior who is at the immediately lower level in the audit team hierarchy. There are three reasons for this focus. First, while the hierarchy described in the previous section encompasses all the levels which comprise the audit team, consideration of all of them is beyond the scope of this research. As a first effort in analyzing judgment dependencies within the audit team, concentrating in detail on a part of the audit team is considered to be preferable to a more comprehensive but more general study.

A second reason for the particular focus is the role of the manager. The manager is more directly involved than the partner in the conduct of the audit: in making important audit decisions, in resolving conflicts between seniors, in continually reviewing the work of the seniors and, when judged necessary, requiring the expansion of a particular area of the audit investigation. In searching for a place where judgment differences are going to be recognized and adjusted for in the audit process, a logical place to look to is the audit manager level.

A third reason for the focus on the audit manager is the primary focus of previous research. As pointed out in the literature review chapter, the majority of previous experimental research on auditor judgment used audit assistants and audit seniors as subjects. The majority of existing research findings relate to these lower levels; considering the implications of these findings, it is appropriate to consider a higher level in the audit team structure: the audit manager.

The basic judgment model of the manager is symbolized as:
\[ P_{O_2}(H_0|P_{O_1}, D) = \frac{P_2(P_{O_1}|H_0,D)P_2(D|H_0)PR_2(H_0)}{\Sigma_i P_2(P_{O_1}|H_i,D)P_2(D|H_i)PR_2(H_i)} \]

which can also be formulated as follows by dividing the numerator and denominator by \( P(D) \):

\[ P_{O_2}(H_0|P_{O_1}, D) = \frac{P_2(P_{O_1}|H_0,D)P_2(H_0|D)}{\Sigma_i P_2(P_{O_1}|H_i,D)P_2(H_i|D)} \]

Referring to either formulation, the manager's judgment is influenced by two factors, the audit test results (D) and the recommendation of the senior (\( P_{O_1} \)). In making a judgment on an aspect of the audit process, the manager relies on the work of his subordinates. The audit procedures conducted and their results are summarized in the working papers. The manager has this evidence and the recommendation of the senior with which to revise his initial opinion. The recommendation or opinion of the senior either is expressed explicitly, by memo or personally, or is provided implicitly when the senior signs the working papers to indicate that the audit procedures were conducted as planned and that the results were satisfactory.

Presumably the manager's evaluation of the evidence includes the evaluation of the senior who prepared the evidence. Factors such as the senior's ability, experience and observed bias may enter this evaluation. The manager will be aware of the senior's record through the accounting firm's formalized performance evaluation system, and there is the informal communication system which transmits information about the staff. Such factors take on added significance given the findings presented in the literature review of limited documentation ( Mock and Turner, op. cit.) and the difficulty of communicating judgmental rationale (Schandl,
These factors can be thought of as reflected in the model by way of the subjective probabilities; on this basis, source credibility is deemed to be represented implicitly in the model. For example, consider $P_2(PO_1|H_0,D)$, which is the manager's subjective likelihood for the senior providing his opinion ($PO_1$) given the results from the procedures conducted ($D$) and the hypothesized state ($H_0$). While the senior's opinion may not be supported by the evidence as unambiguously as the manager would like, he might give more weight (a higher subjective probability) to the senior's opinion that an internal control has functioned as intended, for instance, if the particular senior is experienced compared to unexperienced. However, testing whether the manager is sensitive to source credibility requires an explicit representation of the concept. This representation is considered next.

A Probabilistic Definition of Source Credibility

The model developed is based upon a conception of source credibility proposed in the psychology literature by Schum and DuCharme (1971). They define source credibility in terms of a conditional probability which is incorporated into the Bayesian model. Schum (1978) explains the underlying contention with the following example from jurisprudence. Let $D$ be the event of the testimony by a witness that the streetlight at the scene of the accident was on and $D^*$ the event of the streetlight being on. As a juror in a negligence case events $D$ and $D^*$ are not the same and it is possible that testimony $D$ is consistent with both $D^*$ and its complement $D^*$. The juror needs to question the credibility of the
information source of D when he personally did not observe the event's occurrence or non-occurrence.

Schum's example can be extended to an auditing context. The audit manager has the senior make the observation for him. Of concern to the manager is whether the testimony or report of the senior corresponds to the event which took place. Obviously, in an auditing context, source credibility is of greater concern than the above example suggests. The senior will not normally passively make an observation and report it, but will conduct an audit procedure. In conducting a test, for example, judgments are required on sample size, the type of sampling and, often in checking items selected, irregularities arise which require a judgment to decide whether they need recording. Furthermore, as the description of the audit team indicated, the actual checking or observing may not be carried out by the senior but by one of his subordinates.

The audit procedure and sample results are not independent of the source. Consequently, D* has to be conceptualized as the sample results or audit evidence which would result if no judgment inaccuracies existed, that is, if the audit senior were completely competent and reliable. In a sense, D* represents the evidence that would be contained in the working papers of the ideal audit. Even in the ideal audit, cost considerations would keep the auditor from examining the entire population; thus, the risk inherent in evidence gathering and sampling procedure might cause the ideal audit to incorrectly characterize the true state. As will become apparent, even if the ideal audit is unobservable and possibly not unique, the development of a useful source credibility construct is not prevented. The following definitions extend the previous
notation and summarize the above. (The notation with respect to $D$ and $D^*$ is reversed to that of Schum and DuCharme).

$H_i$ : as in the first model, the $i$th state of nature.

$D_k$ : also used previously, but now subscripted, the audit working papers. $D_k$ may represent the working papers completed to date which the audit manager is reviewing or audit evidence contained in the working papers which the manager is using to make a specific judgment. It is assumed that the evidence contained in the working papers provides a message about the existing state. $D_k$ denotes the message identifying state $k$ as the existing state.

$D^*_k$ : the abstraction of the audit working papers to the condition of containing no judgment inaccuracies as described above. $D^*_k$ denotes the message provided by the ideal audit identifying state $k$ as the existing state.

$P(D^*_k | H_i)$ : is the probability (P) of the ideal audit identifying state $k$ when state $i$ is the true state. For $k = i$, it measures potential informativeness of the audit procedure (even impact). As suggested above, it can be interpreted as sampling risk.

$P(D_k | D^*_k)$ : is the probability (r) of the audit identifying state $l$ when state $k$ would have been identified by the ideal audit. Since the only difference between $D_k$ and $D^*_k$ when $l = k$ is, by definition, due to error caused by
less than perfect credibility, the conditional probability is a probabilistic measure of source credibility.

The following development of the source credibility measure is based on the analysis of Schum and DuCharme (op. cit.). To evaluate the evidence contained in the working papers, a measure of information content is required. The likelihood ratio is frequently used as a measure of information diagnosticity, where diagnosticity is the power of the data to discriminate between hypotheses. Accepting the likelihood ratio as an appropriate measure of the diagnosticity of the working papers, two likelihood ratios can be formulated.

\[ L_{o j} = \frac{P(D^*_k | H_0)}{P(D^*_k | H_j)} \]

\[ A_{o j} = \frac{P(D_k | H_0)}{P(D_k | H_j)} \]

Where \( L_{o j} \) is the likelihood ratio for the evidence from the ideal audit discriminating between hypothesized states. \( A_{o j} \) is the likelihood ratio for the actual audit's evidence discriminating between hypothesized states. It will be called a measure of audit diagnosticity, the information content of the audit work. \( A_{o j} \) can be expanded (see Appendix B) to obtain:

\[ A_{o j} = \frac{\sum_k P(D^*_k | D_k^*) P(D^*_k | H_0)}{\sum_k P(D^*_k | D_k^*) P(D^*_k | H_j)} \]
This likelihood ratio is composed of two elements. \( P(D^*_k|H_0) \) is the probability of the ideal audit indicating state \( k \) when state \( o \) is the true state. It measures the potential impact of data, the audit evidence, abstracted from judgment differences. When \( P(D^*_k|H_1) = 1 \), where \( k = i \), then the audit evidence is perfectly diagnostic (\( L_{o|j} \) approaches infinity) and will always identify the true state. The second component of \( A_{o|j} \) is \( P(D^*_k|D^*_k) \). It specifies the impact of source credibility on the audit's diagnosticity. When \( P(D^*_k|D^*_k) = 1 \), where \( k = k \), the working papers are said to exhibit perfect source credibility in the same sense that \( A_{o|j} = L_{o|j} \), and the audit evidence contained in the working papers is the same as that which would be obtained by the ideal audit. With a less than perfectly credible source, diagnosticity declines and \( L > A > 1 \) (alternatively \( L < A < 1 \)). A likelihood ratio of one indicates the evidence has no diagnostic value.

The definition of source credibility as \( P(D^*_k|D^*_k) \) is sufficiently broad to encompass not only the source's competence as an auditor, but also various motivational factors which affect his veracity and biases. Investigation of such factors is beyond the scope of this research, but is believed to offer an interesting opportunity for extension of the research. As previously indicated, the focus of this research is on the audit manager and his sensitivity to the senior's credibility, not on the nature and causes of the senior's possible lack of credibility.

**Source Credibility and the Audit Manager**

The Bayesian model of the audit manager discussed earlier can be expanded to include the measure of source credibility just presented.
In the model below $\sum_{k} P_{2}(D_{k}^{*} | D_{k}^{*})P_{2}(D_{k}^{*} | H_{1})$ is substituted for $P(D_{k} | H_{1})$. The explanation of this substitution is provided in appendix B.

$$P_{2}(H_{o} | P_{o1}, D_{k}) = \frac{\sum_{k} P_{2}(D_{k} | D_{k}^{*})P_{2}(D_{k}^{*} | H_{o})P_{2}(P_{o1} | H_{o}, D_{k})P_{2}(H_{o})}{\sum_{ik} P_{2}(D_{k} | D_{k}^{*})P_{2}(D_{k}^{*} | H_{i})P_{2}(P_{o1} | H_{i}, D_{k})P_{2}(H_{i})}$$

The elements combined in the manager's judgment consist of his priors, the senior's opinion, the potential inferential value of the audit procedures conducted, and the credibility of the person (the senior) who conducted the audit work. All probabilities are shown as subjective probabilities as they all require estimation by the manager. The conditional probability $P_{2}(P_{o1} | H_{o}, D_{k})$ does not include $D_{k}^{*}$ as an element because the audit senior is believed to form his opinion on the results of the audit work he has conducted.

Since the focus of the research is on the manager's sensitivity to the credibility of his subordinates, the above model can be manipulated to give greater emphasis to the source credibility factor. First, the likelihood form of Bayes' Theorem can be used:

$$P_{2}(H_{o} | P_{o1}, D_{k}) = \frac{\sum_{i} P_{2}(D_{i} | D_{i}^{*})P_{2}(D_{i}^{*} | H_{o})P_{2}(P_{o1} | H_{o}, D_{i})P_{2}(H_{o})}{\sum_{ij} P_{2}(D_{j} | D_{j}^{*})P_{2}(D_{j}^{*} | H_{j})P_{2}(P_{o1} | H_{j}, D_{j})P_{2}(H_{j})}$$

A second way of simplifying the expression is to collapse the senior's opinion into the working papers. In the first model the senior's opinion was distinguished from the audit evidence to emphasize the hierarchical information processing sequence occurring within the audit team. Recognition of the hierarchical relationships existing within the audit team enable source credibility to be introduced. Now what is of
interest is how source credibility impacts upon the inferential value of the audit evidence relied on by the manager. Treating the senior's opinion as part of the audit evidence enables the model to emphasize the impact of source credibility. The final form of the model is:

\[
\frac{PO_2(H_o|D_j)}{PO_2(H_j|D_j)} = \sum_k P_2(D_j|D^*_k)P_2(D^*_k|H_o)PR_2(H_o)
\]

or

\[
\Omega_1 = A_{oj} \Omega_o
\]

Treating the senior's opinion as part of the audit evidence included in the working papers avoids the necessity of clarifying the distinction between the conditional probabilities \(P_2(D_j|D^*_k)\) and \(P_2(PO_1|H_1,D_j)\). The conditional probability \(P_2(D_j|D^*_k)\) was defined as a measure of source credibility. When discussing \(P_2(PO_1|H_1,D_j)\) it was recognized that the probability may be a function of personal characteristics of the senior. Experience was mentioned as one such factor; possibly source credibility could be included. Alternatively, the manager may regard the senior's opinion and the related conditional probability estimate as an impersonal piece of information independent of source credibility. By including \(PO_1\) as part of \(D_j\), the personnel evaluation aspects of the manager's judgment are all included in the source credibility factor.

**An Example: Internal Control Evaluation**

With the specification of the judgment model incorporating source credibility, the impact of source credibility can be considered normatively. In order to apply this model to the manager's judgment process
an audit setting is required. The setting to be employed is one of interest, internal control evaluation. This example provides an introduction to the experiment, described in the next chapter, which deals with internal control evaluation. To make this example consistent with the experiment, it is assumed that only two states of internal control can exist: strong and weak ($H_i: i = 1, 2$). If desired, a model including three internal control states could be developed; the primary results, to be presented later, would not differ. Furthermore, internal control evaluation can result either in substantial reliance of in no reliance being given to the system. Each of these levels of reliance may be indicated by the ideal audit ($D_k^*: k = 1, 2$) and the actual audit ($D_l*: l = 1, 2$).

The following matrix outlines the state-event paradigm for internal control under ideal audit conditions, where $P(D_k^*|H_i) = p$ for $k = i$.

<table>
<thead>
<tr>
<th></th>
<th>$H_1$</th>
<th>$H_2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$D_1^*$</td>
<td>$P$</td>
<td>$1-P$</td>
</tr>
<tr>
<td>$D_2^*$</td>
<td>$1-P$</td>
<td>$P$</td>
</tr>
<tr>
<td>$E$</td>
<td>$1.0$</td>
<td>$1.0$</td>
</tr>
</tbody>
</table>

Assuming that $P(D_k^*|H_i)$ is symmetric means that the ability of the audit procedures to identify the true state is not a function of the state which actually exists. In terms of statistical sampling terminology, this implies that the type I error rate equals the type II error rate. Similarly, a symmetric assumption is also made for source credibility $P(D_l^*|D_k^*)$. The source credibility matrix is:
Here the symmetry assumption implies that the credibility of the audit senior who has conducted the audit procedures entailed in the evaluation of internal control is not a function of the state which actually exists. That is, the accuracy of the results of the audit procedures conducted does not depend on whether the internal control system is strong or weak. While both symmetry assumptions are made primarily for convenience, they are considered reasonable assumptions in an auditing context; significantly different results would not be obtained when the symmetry assumptions did not hold for values of \( p \) and \( r \) likely to exist in auditing. Schum and DuCharme (op. cit.) do consider other possible assumptions.

Under the assumptions of symmetry, the likelihood ratios for measuring the diagnosticity of the audit evidence are the same for \( D_1 \) and \( D_2 \). The likelihood ratio for the receipt of message \( D_1 \) from the audit senior is considered below. Table 2 contains the numerical likelihood ratios for various values of \( p \) and \( r \).

\[
A_{12} = \frac{2 \sum_{k=1}^{2} P(D_1 | D^*_k) P_2(D^*_k|H_1)}{2 \sum_{k=1}^{2} P(D_2 | D^*_k) P_2(D^*_k|H_2)}
\]

\[
= \frac{rp + (1-r)(1-p)}{r(1-p) + (1-r)p}
\]
The respective impacts of \( p \) and \( r \) are seen in Table 2. As source credibility increases, the impact of the message approaches its potential diagnostic value; then, as \( p \) approaches one, it becomes perfectly diagnostic. Of interest is the rate of decline in diagnosticity as source credibility falls. From a normative standpoint, the following table indicates that, source credibility interacts with potential evidence diagnosticity and that a relatively small reduction in source credibility causes a relatively large reduction in diagnosticity. This observation is important because in an auditing context the relevant probabilities are quite high. Sampling procedures are normally conducted with sample sizes providing 90 percent, 95 percent and 99 percent levels of confidence. Further, an incompetent auditor who had little credibility probably would not keep his job very long; auditors must meet high levels of educational and professional standards.

**TABLE 2**

<table>
<thead>
<tr>
<th>( r )</th>
<th>( A(P=.1.0) )</th>
<th>( A(P=.99) )</th>
<th>( A(P=.95) )</th>
<th>( A(P=.90) )</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.00</td>
<td>( \infty )</td>
<td>99.00 (1.996)</td>
<td>19.00 (1.279)</td>
<td>9.00 (0.954)</td>
</tr>
<tr>
<td>0.90</td>
<td>9.00 (0.95)</td>
<td>8.26 (0.916)</td>
<td>6.14 (0.788)</td>
<td>4.56 (0.659)</td>
</tr>
<tr>
<td>0.80</td>
<td>4.00 (0.602)</td>
<td>3.85 (0.585)</td>
<td>3.35 (0.525)</td>
<td>2.85 (0.455)</td>
</tr>
<tr>
<td>0.70</td>
<td>2.33 (0.367)</td>
<td>2.29 (0.360)</td>
<td>2.13 (0.328)</td>
<td>1.94 (0.288)</td>
</tr>
<tr>
<td>0.60</td>
<td>1.50 (0.176)</td>
<td>1.49 (0.173)</td>
<td>1.44 (0.158)</td>
<td>1.38 (0.140)</td>
</tr>
<tr>
<td>0.50</td>
<td>1.00 (0.000)</td>
<td>1.00 (0.000)</td>
<td>1.00 (0.000)</td>
<td>1.00 (0.000)</td>
</tr>
</tbody>
</table>

*Likelihood ratios are derived for hypothetical symmetrical data: \( P(D_1|D^*_1) = P(D_2|D^*_2) = r, P(D^*_1|H_1) = P(D^*_2|H_2) = p \), and \( i = l = k = 1, 2 \). The log likelihood ratios are given in brackets.*
Chapter Summary

The purpose of the chapter was to consider the first research objective: to formulate a model of the auditor's judgment which includes a measure of source credibility as an input. The initial model presented highlighted the dependencies and relationships which were found to exist in the analysis of the audit team from the previous chapter. Within the audit team, the focus was on the relationship between the audit manager and audit senior from the perspective of the audit manager.

A Bayesian model incorporating a measure of source credibility, developed by Schum and DuCharme (op. cit.) in the psychology literature, was used to model the audit manager's judgment. A numerical example containing values of \( p \) and \( r \) considered typical of an auditing situation was analyzed. The results suggested that source credibility can have a significant impact even when the relevant probabilities are high.

The critical question now is whether the audit manager is sensitive to source credibility; that is, does he perceive the impact of source credibility and, if he does, how do his responses compare to those of the model? These questions were among the issues investigated in the experiment. A description of the experiment and the specific hypotheses tested is contained in the next chapter.
CHAPTER IV

DESCRIPTION OF THE RESEARCH METHODOLOGY: A TWO-STAGE EXPERIMENT

The purpose of the experiment is to consider the second, third and fourth research objectives. The second research objective is to test to what degree the auditor's judgment is sensitive to source credibility. This objective is considered in the first stage of the experiment. The third and fourth research objectives are, respectively, to test whether the auditor will modify the audit program to compensate for variations in source credibility, and to find out what characteristics of a subordinate influence the auditor's appraisal of the subordinate's credibility. Both these objectives are considered in the second stage of the experiment.

The chapter is organized into five sections. The first section justifies both the use of an experimental methodology and the use of a two-stage experiment. The second section describes the subjects used in the experiment. The third and fourth sections describe the two stages of the experiment. Each stage is described in terms of experimental task, independent variables, dependent variables, hypotheses, and statistical analysis. Finally, the fifth section describes the pretest.

A Two-stage Experiment: Justification

Three methods are commonly used to collect data on research hypotheses. The first of these, the field study, can be discarded
immediately. The field study methodology is used to observe behavior. In the present study, this method would entail analyses of the audits of actual clients. The members of the audit team would be observed to determine the credibility of the auditors involved, and how the auditors' superiors dealt with source credibility. There would be significant practical problems with this approach, including the difficulty of measuring the auditors' perceived credibility and the lack of rationale documentation. Also, a major problem would be the inability to attribute auditors' behavior to the credibility of their subordinates, rather than to a multitude of other situational specific factors. These problems and the infeasible time demands of this method mean that an experimental method has to be utilized.

The choice is between a field experiment and a laboratory experiment. The field experiment, like the field study, takes place in a realistic setting but, unlike the field study, is designed to enable control and manipulation of predetermined variables. A laboratory is "any setting that allows investigators to control closely the conditions under which observations can be made." (Filley, House and Kerr, 1976, p. 43). In a laboratory experiment, an artificial setting is often substituted for a realistic setting to obtain the greatest possible control and internal validity. The choice of experimental method essentially consists of a tradeoff, constrained by economic considerations, between external validity and internal validity.

The field experiment is also rejected. To satisfy this method, the subjects should not be aware that they are participating in an experiment; rather they should believe that they are conducting a real audit.
Creating such a scenario is difficult because of the complexity of the audit process, the physical investigations of the client undertaken, and the differences in audit technologies between accounting firms. Therefore, a laboratory experiment is chosen.

A laboratory experiment is a research study in which the variance of all or nearly all of the possible influential independent variables not pertinent to the immediate problem of an investigation is kept at a minimum. This is done by isolating the research in a physical situation apart from the routine or ordinary living and by manipulating one or more independent variables under rigorously specified, operationalized, and controlled conditions. (Kerlinger p. 398)

The strength of the laboratory experiment is in the internal validity provided by controlling extraneous environmental influences, and in the freedom allowed the researcher in the design of the experimental materials. The gain in internal validity, however, is at the expense of external validity. The question over external validity arises because the experimental situation controls many elements which would otherwise give reality to the task. Also, the independent variables manipulated are often surrogates for real world variables.

The external validity of the experiment is enhanced, as will become evident when the stages of the experiment are described, by several factors. First, audit managers are used as subjects. Because of their role within the audit team, described earlier, they should be particularly concerned with source credibility. Second, an audit case study is used to provide the setting for the manipulation of the independent variables. The case study contains information normally encountered in an audit; it is designed to facilitate the auditor-subjects relating to the experiment. Third, the experiment is conducted
in a realistic setting: the auditor-subject's office. Finally, a two-stage experiment is used, in part, to deal with the issue of the surrogation of real world variables.

One reason for utilizing a two-stage experiment is the difficulty in representing source credibility. In the first stage, an explicit probabilistic measure of source credibility is provided. This representation is used so that the experiment is consistent with the Bayesian model presented in the previous chapter. The comparison of the subjects' responses with the model's predictions aids in the analysis of the experiment's results. Furthermore, a probabilistic mode recognizes the uncertainty inherent in the audit process. The potential problem (not substantiated in the results) is that auditors may not be accustomed to dealing with so explicit a representation of source credibility. Consequently, in the second stage an implicit measure of source credibility is provided by manipulating factors which have been found in psychological studies to affect job performance.

The inclusion of these factors in the second stage also enables the fourth research objective to be considered. This is achieved by analyzing the results to determine which of the factors manipulated has the greatest affect on the subjects' behavior. Another reason for having the second stage is that it provides another view of auditors' behavior in response to differences in source credibility. In the first stage, the auditor is asked to make a probabilistic judgment about the state of the client's internal control system. In the second stage, the auditor is asked to make sample size judgments. Sample size judgments are somewhat more concrete, and also directly involve cost considerations.
Thus, the second stage also investigates the third research objective.

Subjects

The research topic, briefly stated: determining whether source credibility is a factor in an auditor's judgment, makes it essential that the subjects be auditors rather than student surrogates. Further, given the nature of the functional levels in the audit team, preferably the subjects should be audit managers.

Six "Big Eight" public accounting firms with offices in Columbus, Ohio were contacted and asked to provide subjects; 35 audit managers participated. These six firms will be referred to as F1-F6. The number of subjects from each firm was eight, seven, six, six, five, and three, respectively. The offices were of similar size, having an average of 100 professionals and less than ten audit managers. Thus, bias from a non-random selection of subjects is not viewed as a concern. The subjects being from one geographic location and from offices of similar size restricts the generalizability of the results, although the researcher does not believe a serious bias is induced, given the research topic.

The test instrument making up the first stage of the experiment was administered to the subjects in their offices with the researcher present. In one instance, two subjects, and in another instance, three subjects, answered the test instrument at the same time although independently. The subjects took between fifteen and thirty-five minutes and averaged twenty-five minutes to complete their task. Afterwards, they were asked if they had had problems with the task, or wished to make any comments on it. The advantage here was that any difficulties
the subject had with the task could be cleared up immediately (verbal communication with the subjects is also reported on in the results chapter). This was an important consideration. Subjects in the first stage were required to deal with uncertainty in the form of percentages and probabilities. Since some auditors may have felt uncomfortable dealing with uncertainty in such explicit terms, the researcher administered the test instrument of the first stage personally. As it turned out, no problems were encountered with the administration and, on the occasions that the researcher was asked a question, minor clarifications were involved. The subjects appeared to the researcher to be interested in their task and highly motivated.

At the end of the session with each subject, the test instrument for the second stage of the experiment was left with the subject. The subjects were asked to complete it at their convenience and to return it to the researcher, in the stamped addressed envelope provided. Thirty-one test instruments were returned. The researcher not personally administering the second-stage test instrument does threaten internal validity. Subjects from the same accounting firm may not have worked independently, and additional information may have been differentially consulted by subjects. Neither possibility, however, was considered likely by the researcher. In addition, the degree of consensus is not a major research issue.

The First Stage of the Experiment

The purpose of the first stage of the experiment is to provide empirical evidence on the second research objective. The research question is: are auditors sensitive to the credibility of their
subordinate team members; or, in terms of the Bayesian model developed, can support be found for the intuitive suggestion that the inferential value of information provided by a source depends not only upon the potential inferential impact (diagnosticity) of the event reported but also on the credibility of the source? As discussed earlier, there is reason to believe that source credibility should be of concern in an audit setting.

Experimental Task

The use of a representative task to enhance external validity requires the modification of the typical task used in cascaded inference experiments involving source credibility. A look at these experiments, however, does indicate the data requirements for this type of experiment. Since the design of all the empirical studies (Snapper and Fryback, op. cit.; Schum, DuCharme and DePitts, op. cit.; Youssef and Peterson, op. cit.; Funaro, op. cit.) is similar, only one study is outlined here. The results of the first study, the Snapper and Fryback study, were utilized by the others. Their purpose was to investigate subjects' inferences based on partially credible reports. The binomial hypotheses and credibility levels were presented in the experiment by means of discs and spinners. Each pair of binomial hypotheses was represented by two discs having different proportions of red and blue. In conditions with \( r \) (credibility) less than one, a grey and white disc represented the credibility of the source with \( r \) equal to the proportion of the disc that was white. The numerical values of the proportion of each color were printed on each disc, and the discs were displayed to the subjects throughout the experiment.
The experimenter had identical discs. He flipped a coin to select one disc from the appropriate pair of red-blue discs and attached a spinner to its center. Another spinner was attached to the center of the appropriate credibility disc and both were spun. The experimenter, but not the subjects, observed the outcome of both spins and, if the credibility spinner landed on grey, he falsely reported the color on which the other spinner stopped; otherwise a truthful report was made. The students' task, having their own discs in front of them and knowing the reporting rule, was to write down likelihood ratio estimates about which red-blue disc had been selected. The procedure was repeated within each condition until both possible reports had occurred. The condition consisted of the parameter values employed. The values of \( r \) were 1.0, 0.9, and 0.7 while the four parameters for each pair of hypotheses, \( P(D^*|H_a) \), \( P(D^*|H_b) \) and \( P(D^*|H_a) \), \( P(D^*|H_b) \) were: (a) .33, .67 and .33; (b) .80, .20 and .60, .40; (c) .90, .10 and .45, .55; (d) .25, .75 and .75, .25. As was mentioned in the literature review, the subjects responded to reduced source credibility by overestimating the diagnostic impact of the unreliable report.

In this experiment a mechanical device, such as a spinner, is not used to obtain outcomes during the experiment. Rather, outcomes are predetermined and the parameters manipulated in written alternative situations to which the subjects respond. The disadvantage of this alternative procedure is that some of the spontaneity of the experiment is lost. The advantage is that the parameters can be put into a representative context for the subject. Following Tversky and Kahneman (1974) and as outlined in the literature review, this approach to
probabilistic research has been followed in accounting by Swieringa et al., Gibbins, and Biddle and Joyce.

The experimental situations are not completely independent in that each represents an alternative outcome of a single phase of an audit. The audit is described in a comprehensive case study representing a realistic audit situation. A copy of the case study is provided in Appendix D. It contains materials normally prepared during an audit, including the relevant results of last year's audit procedures and this year's compliance tests. The subjects (audit managers) participating in the experiment are asked to assume manager responsibility for an audit that is currently underway. Their task, in the first stage of the experiment, is to first provide, on the basis of the initial evidence contained in the case study, a likelihood (prior probability) that the client's internal control system for sales and receivables has functioned sufficiently well to justify placing substantial reliance on it in determining the extent of confirmation of accounts receivable. Last year no reliance was placed on the internal control system. This year the client has made several improvements in the system. The compliance test results reflect this improvement and the senior recommends substantial reliance. After reading the results of compliance tests conducted by the senior and his recommendation, subjects provided a second likelihood judgment (posterior probability). Eight alternative situations in which two independent variables are manipulated are included in the case study.
Independent Variables

In the experiment conducted by Snapper and Fryback, the information provided by the source was subject to two types of uncertainty. First, there was the uncertainty associated with the occurrence of the event itself. This uncertainty was defined as $P(D_k^*|H_0)$ and was called event impact in the model presented in the previous chapter. The second type of uncertainty concerned the ability of the source to report the event reliably. This uncertainty was defined as $P(D_{\alpha}|D_k^*)$ and was called source credibility.

In the first stage of the experiment, source credibility is manipulated through the description of the senior who conducts the audit procedures. The subject is informed that the consensus among the audit managers who have worked with the particular senior, including the subject himself, is that the senior has a certain level of credibility. Credibility is explained to the subject in terms of the senior's technical ability, in particular his reliability in conducting audit work and presenting the results. An anchored reliability scale is also provided. The case of 100 percent credibility is described as follows:

**Evaluation of Tacket:** The consensus among the audit managers, including yourself, who have worked with Tacket is that he is completely reliable in conducting audit work and in presenting the results. In terms of the rating scale given below, he is considered to be 100% reliable—which means that, for example, in the past the results he has presented have been found in review to be always acceptable.

<table>
<thead>
<tr>
<th>50%</th>
<th>60%</th>
<th>70%</th>
<th>80%</th>
<th>90%</th>
<th>100%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Completely Unreliable</td>
<td>Highly Unreliable</td>
<td>Fairly Unreliable</td>
<td>Reliable</td>
<td>Fairly Reliable</td>
<td>Highly Reliable</td>
</tr>
</tbody>
</table>
Four levels of reliability are used in the experiment. The probabilistic reliability values and the corresponding descriptive anchors are: (1) 100 percent completely reliable, (2) 90 percent - highly reliable, (3) 80 percent - fairly reliable, and (4) 70 percent - fairly unreliable. Note that in a two-state case, which is the situation represented here, a value of 50 percent is equivalent to complete unreliability as there is an equal probability of getting a true or false report. While descriptive anchors are inescapably arbitrary, the ones used here were extensively pre-tested. Lower levels of reliability are not used because it is believed unlikely that they would be found in audit seniors. Even the 70 percent level is considered to be somewhat unrealistic; it was included to determine the subjects' reaction to it and to provide a spread of the variable. The normative model suggested that small reductions in credibility would have a significant impact.

The second independent variable is event impact; in the audit context, this is the informativeness of the audit procedures undertaken. These audit procedures are the compliance tests conducted by the senior. Since statistical sampling techniques are commonly used in making compliance tests a probabilistic measure of event impact is readily obtained; it is the confidence level at which the compliance tests are conducted. Confidence levels of 90 percent and 95 percent, common for auditing, are used in the experiment. Lower confidence levels are believed not to be commonly used in auditing and so are not used in the experiment. Anderson (op. cit.) describes the 95 percent confidence level as providing normal to high reliance on internal controls, and the 90-percent confidence level as providing normal to low reliance.
Since the primary interest is in the effect of source credibility, the second independent variable is included primarily to see how source credibility interacts with it.

The experimental design can be compared to the design employed by Snapper and Fryback. Their two alternative hypotheses were which of two spinners was selected by the flip of a coin; the corresponding hypotheses the auditor-subject considers are whether the client's system has or has not functioned sufficiently well to justify substantial reliance on it. Snapper and Fryback controlled for prior probabilities by selecting the spinner by flipping a coin; here, the subjects are asked for their priors based on the information provided in the case study. The senior's recommendation concerning reliance on the system and the compliance tests on which he based his recommendation take the place of the reported color the spinner landed on as the information the subject uses in revising his opinion. Snapper and Fryback controlled report (overall) diagnosticity by manipulating event impact, $P(D^*|H_o)$, and source credibility, $P(D^*|D_k)$, by the color proportions on the appropriate spinners. Here, as described previously, event impact is manipulated by the statistical confidence level at which the compliance tests are conducted. Source credibility is manipulated through the description of the senior and his reliability rating.

**Dependent Variable**

The subjects are asked to revise their initial judgment, about the probability that the client's system has functioned sufficiently well to justify substantial reliance on it, in determining the extent of confirmation of accounts receivable. The evaluation of internal
control for this purpose is a common audit judgment. Snapper and Fryback asked their subjects to provide likelihood ratio estimates. Instead of likelihood ratios, subjective probabilities, to be recorded on a one hundred point scale from 0.00 to 1.00, are asked for in the experiment. Auditors are believed to be more comfortable with this form of response than with odds estimates. There are no clear preferences in the psychology literature between likelihood ratio and probability judgments (Slovic, Fischhoff and Lichtenstein, 1977). A one hundred point scale has been used in accounting research before (Reckers and Schultz, op. cit., and Kennedy, 1975, for example) and is commonly used in source credibility and opinion change literature.

**Research Hypotheses**

To summarize the first stage of the experiment, subjects are asked to review their initial judgment, about the probability that the client's system has functioned sufficiently well to justify substantial reliance on it, on the basis of the compliance tests conducted by the senior and his recommendation. Uncertainty is introduced into the experiment in the form of the confidence level at which the compliance tests are conducted and, more importantly, the level of the senior's credibility. Given two confidence levels and four levels of credibility, each subject provides eight responses (2 x 4) in a complete repeated measures design.

The subjects' responses are used to test the following four hypotheses relating to the first research objective. The first three hypotheses follow directly from the behavior of the normative model presented in the last chapter, but only the first two can be said to be
intuitively obvious. The fourth hypothesis concerns the direct comparison of the subjects' behavior with the benchmark provided by the model. The hypotheses are stated in their null form.

H1: The subjective probability that the client's system has functioned sufficiently well to justify substantial reliance on it is unaffected by the level of source credibility.

This hypothesis is of primary interest; it should be rejected. The thesis of the research is that source credibility affects the inferential value of information and is of concern to the auditor. The psychological studies cited in the literature review found source credibility to be a statistically significant independent variable: the higher the source credibility the higher the information's inferential value. Therefore, the subjective probability should be larger the higher the source credibility level.

H2: The subjective probability that the client's system has functioned sufficiently well to justify substantial reliance on it is unaffected by the confidence level of the compliance tests.

This hypothesis should be rejected. The higher the statistical confidence level at which compliance tests are conducted, the higher the event impact and inferential value of the tests results. Therefore, the subjective probability should be larger at the higher confidence level.

H3: The impact of source credibility is unaffected by the confidence level of the compliance tests.

This hypothesis should be rejected according to the predictions of the normative model. The model suggests an interaction occurs between the two independent variables; source credibility has more impact when the confidence level is higher.
H4: The derived likelihood ratios are approximated by the likelihood ratios calculated from the model.

This hypothesis should be rejected if the results of previous research on cascaded inference are replicated. This research has generally found that subjects are sensitive to source credibility, but that they do not adjust for source credibility variations as much as the normative model suggests they should. As previously stated, no attempt is made to present the results of the normative model as the correct responses for auditors; rather the normative model merely provides a benchmark for evaluating subjects' responses. The fourth hypothesis is important for another reason. Most of the psychology research into source credibility, including the research using the source credibility model developed by Schum and DuCharme (op. cit.), has employed students as subjects and has given them artificial tasks. In contrast, this research employs as subjects people who are expert judges in their own profession and has given them a familiar and realistic task.

Statistical Analysis

Given two statistical confidence levels of compliance testing and four levels of source credibility, each of the 35 subjects provided eight responses in a completely repeated measures design (2 x 4 x 35). There are three reasons why the use of a repeated measures or within-subject design is particularly appropriate in the study. First, all treatment combinations are observed so that all interaction effects are testable, while subjects are efficiently utilized. In research that is largely exploratory, as auditor judgment research is, being able to
test all interactions allows greater confidence to be attached to the results. In addition, the efficient utilization of subjects is important when expert judges are used as subjects; both their numbers and their time are scarce resources.

The second reason for the use of repeated measures is the control provided over individual subject differences. According to Winer (1971);

> the primary purpose of repeated measures on the same elements is the control that this kind of design provides over individual differences between experimental units. In the area of the behavioral sciences, differences between such units often are quite large relative to differences in treatment effects which the experimenter is trying to evaluate. (Winer, p. 517)

As explained in the literature review, large differences in judgments have been found among auditors.

The third reason for using repeated measures relates specifically to the nature of the experimental task. The auditor-subjects are asked to make successive internal control judgments based on different evaluations of the audit senior. Such a situation is not unfamiliar to the audit manager. The audit manager usually is a member of more than one audit team, and on each of the audit teams there is likely to be more than one audit senior. The result is that the audit manager is used to dealing with several seniors at one time and, it is hypothesized, is used to making judgments about their relative credibility and its effects.

The danger with a repeated measures design is the possibility of inducing significant results (called demand effects) through the subjects knowing what variables are being manipulated. This danger was
minimized by emphasizing to the subjects, before they commenced their
task, that the study was exploratory and that there were no right or
wrong answers. The researcher believes that the nature of the subjects
made it unlikely that demand effects would be a major concern. The
reason for this belief is that audit managers are experienced, success­
ful professionals who would have no motive to provide misleading
responses and would be unlikely to be intimidated by the experimental
task. If student surrogates had been used as subjects, the possibility
of intimidation would have been of greater concern. The results sug­
egest that significant demand effects were not present in subjects'
responses. The argument supporting this claim is given when the
results are presented.

A two-factor repeated measures analysis of variance (ANOVA) pro­
vides a convenient method of testing the first three hypotheses. The
tests are conducted by F-ratio tests on the ANOVA model's main effects
and interaction effects. The ANOVA model is the standard approach of
experimental studies in both source credibility research and auditor
judgment research, particularly when repeated measures are employed.
The appropriate ANOVA design, together with F-tests, is given in Table
3. The two main effect F-ratios and one interaction F-ratio provide
the tests required for the first three research hypotheses.

Of interest is not only whether source credibility accounts for a
statistically significant amount of variability in subjects' responses,
but also how subjects respond to changes in source credibility (levels
of the main effect). In particular, is there a certain level of source
credibility which causes a major change in subjects' responses?
### TABLE 3
ANALYSIS OF VARIANCE FOR A TWO-FACTOR REPEATED MEASURES DESIGN

<table>
<thead>
<tr>
<th>SV</th>
<th>df</th>
<th>SS</th>
<th>EMS</th>
<th>F</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total</td>
<td>abn - 1</td>
<td>( \sum \sum \sum Y_{ijk} - C )</td>
<td>( \sigma^2 + b\sigma^2_{AB} + b\theta^2 )</td>
<td>( MS_A )</td>
</tr>
<tr>
<td>A</td>
<td>( a - 1 )</td>
<td>( \sum_i (\sum_j \sum_k Y_{ijk})^2 - C )</td>
<td>( \sigma^2 + ab\sigma^2_{AB} + a\theta^2 )</td>
<td>( MS_A )</td>
</tr>
<tr>
<td>B</td>
<td>( b - 1 )</td>
<td>( \sum_i (\sum_j \sum_k Y_{ijk})^2 - C )</td>
<td>( \sigma^2 + a\sigma^2_{BS} + a\theta^2 )</td>
<td>( MS_B )</td>
</tr>
<tr>
<td>S</td>
<td>( n - 1 )</td>
<td>( \sum_i (\sum_j \sum_k Y_{ijk})^2 - C )</td>
<td>( \sigma^2 + ab\sigma^2 )</td>
<td>( MS_A )</td>
</tr>
<tr>
<td>AB</td>
<td>((a - 1)(b - 1))</td>
<td>( \sum_i (\sum_j \sum_k Y_{ijk})^2 - C - SS_A - SS_B )</td>
<td>( \sigma^2 + \sigma^2_{ABS} + n\theta^2_{AB} )</td>
<td>( MS_{AB} )</td>
</tr>
<tr>
<td>AS</td>
<td>((a - 1)(n - 1))</td>
<td>( \sum_j (\sum_i \sum_k Y_{ijk})^2 - C - SS_A - SS_S )</td>
<td>( \sigma^2 + b\sigma^2_{AS} )</td>
<td>( MS_A )</td>
</tr>
<tr>
<td>BS</td>
<td>((b - 1)(n - 1))</td>
<td>( \sum_i (\sum_j \sum_k Y_{ijk})^2 - C - SS_B - SS_S )</td>
<td>( \sigma^2 + a\sigma^2_{BS} )</td>
<td>( MS_A )</td>
</tr>
<tr>
<td>ABS</td>
<td>((a - 1)(b - 1)(n - 1))</td>
<td>( \sum \sum \sum Y_{ijk}^2 - C - SS_A - \cdots - SS_{ABS} )</td>
<td>( \sigma^2 + \sigma^2_{ABS} )</td>
<td>( MS_A )</td>
</tr>
</tbody>
</table>

Duncan's multiple range test is used to test for significant differences on the main effect (source credibility) means. This a posteriori test uses "a protection level of $\alpha$ for the collection of tests rather than an $\alpha$ level for the individual tests" (Winer, op. cit., p. 196).

A practical reason for using Duncan's test is that it is available in the computer statistical program used in the data analysis (SAS Users Guide, 1979). Given that the independent variable, source credibility, has only four levels in the experiment, Duncan's test gives results similar to, although slightly less conservative, the Newman Keuls test (Winer, op. cit., p. 197). The advantage of the test in this study is that it provides more power than Newman Keuls and other similar tests. This characteristic is an advantage since the test is employed here as a descriptive technique, not to provide a strict test of a null hypothesis.

ANOVA procedures are initially run at two levels of interest: overall and firm. Firm ANOVAs are run to consider how subjects employed by a specific firm behave, and to look for a firm effect. Individual subject ANOVA's are of limited value because of the lack of error term. To overcome this problem cluster analysis is performed to identify groups of subjects with similar responses. The purpose of the clustering is descriptive, not inferential; that is, the purpose is to assist in describing subjects' behavior, not to infer reasons for it. Cluster analysis is an exploratory classification technique. The cluster procedure in SAS is used (SAS Users' Guide, op. cit.). These groups then provide a level of ANOVA analysis.
The testing of hypothesis four requires a more complex manipulation of the data than does the testing of the first three hypotheses. First, a means of comparing subjects' responses (S) with the model's predictions (M) is required; decision research in psychology has employed several such measures (Beach, 1975). The measure selected for the study is:

\[
\text{Difference Measure (DM)} = \log \Omega_1(S) - \log \Omega_1(M) \\
= \log A(S) - \log A(M)
\]

There are several reasons for this choice. Because prior probabilities are not controlled for in the experiment, model predictions have to be calculated separately for each subject; this, in turn, means that the analysis has to utilize a form of difference or ratio measure. A difference measure is selected because it is believed to be more easily interpreted. The difference measure is simply the logarithmic transformation of a ratio measure. Such a transformation is common practice in decision research in psychology. It assists in the data analysis because the distribution of likelihood ratios is usually skewed. It also has the advantage of making the Bayesian model additive, which means that the difference between log posterior odds equals the difference between log likelihood ratios as the priors cancel. Finally, the focus is on the difference in likelihood ratios rather than the difference in posterior probabilities. This focus is because the likelihood ratio, as explained in the last chapter, is a common measure of information diagnosticity. Of interest is the effect of source credibility on the inferential value or diagnosticity of the information as perceived by the subjects. Also, the calculation of the
likelihood ratios for the model requires less calculation than the calculation of the posterior probabilities.

The calculation of the subjects' inferred likelihood ratios and the model's derived likelihood ratios is demonstrated in Appendix C. Since each subject provides a prior probability and a posterior probability for a two-state situation, the calculation of the subjects' inferred likelihood ratio is straightforward. The difficulty is in calculating the model's likelihood ratio. The problem with the model is that its information requirements are high in that it requires as inputs not just the likelihood ratio, but the individual conditional probabilities which comprise the likelihood ratio. In a statistical auditing setting this means that an estimate of the error rate under the alternative hypothesis (type II error) is required as well as the error rate under the null hypothesis (type I error). However, no direct consideration is usually given to the type II error in compliance testing. The second conditional probability required by the model can not be meaningfully included in the case information; it has to be estimated.

The second conditional probability is solved in the case of perfect source credibility by equating the model's likelihood ratio to the subjects' likelihood ratio. The three assumptions underlying this procedure are discussed and justified in Appendix C. Once this probability is estimated, it is used in the other three cases (levels) of source credibility to obtain the model's likelihood ratio. This means that the comparison of the two likelihood ratios can only be made at three levels of source credibility. The differences obtained are
analyzed with the same statistical procedures the responses are analyzed with, including overall, firm, and cluster ANOVAs.

The Second Stage of the Experiment

The purpose of the second stage of the experiment is to provide empirical evidence on the third and fourth research objectives. Specifically, will the subjects adjust sample sizes in response to variations in the senior's source credibility, and how does a given set of factors influence the subjects' perception of the audit senior's credibility? The second stage does not become redundant if source credibility is found to be insignificant in the first stage. There is always the possibility that the first stage does not capture significant variations in source credibility from the subjects' viewpoint, or that they are not used to considering source credibility explicitly. Rather, auditors may implicitly make allowances for source credibility by considering various attributes of the credibility of the source.

Experimental Task

A copy of the second-stage test instrument is included in Appendix E. The audit case study used in the first stage is extended and used in the second stage. The first stage focuses on the functioning of the clients' internal control system. In the second stage, a next step in the audit process is considered: the confirmation of the client's accounts receivable (a process in which the client's debtors are asked to communicate directly with the auditor). The information originally provided in the case-study, including the compliance test results, remains. In addition, the results of last year's confirmation are
included and, in place of the senior's recommendation concerning reliance on the internal control system, the senior's recommended sample sizes for this year's confirmations are provided.

The subjects' task is to consider the information given and to make sample size judgments. Note that the probabilistic context of the task is dropped. Also, the senior's credibility is not explicitly referenced in the experimental situations; rather, factors which are hypothesized to contribute to, or interact with, the perceived credibility of the senior are manipulated. Eight situations are contained in the test instrument and for each situation the subjects make six judgments. Four are sample size judgments; the accounts receivables are broken down into four strata and the subjects have the option of sampling each stratum. The other two responses are self-reports on the weight the subjects give to the compliance test results and the senior's sample size recommendations in arriving at their own sample size judgments.

**Independent Variables**

The four factors included in the experiment are referred to in either prior auditor judgment research or in the psychology literature, in attitude/opinion change research or personnel performance rating research. The four factors are: (1) experience of the senior, (2) past performance rating of the senior, (3) sampling approach employed by senior, and (4) importance of the issue to the subject. The first three factors relate to the source; experience and performance rating concern the perceived credibility of the source, and the judgment approach concerns how the source arrived at his sample size
recommendations. These factors are manipulated across experimental situations. The fourth factor is not so manipulated. It is considered by means of the materiality of strata in the accounts receivables.

Experience has been the subject of a great deal of auditor judgment research (Hamilton and Wright, op. cit., Uecker and Kinney op. cit., Weber, op. cit., Mock and Turner op. cit.). As was noted in the literature review, the evidence is mixed on whether more experienced auditors are better judges than less experienced auditors. The concern here is not whether they are better judges, but how an audit manager perceives the effect of experience on a subordinate's work and judgment. In the performance rating literature, there is evidence to show that ratings of performance were higher in the condition of higher perceived experience (Leventhal, Perry and Abrami, 1977). Two levels of experience, little and extensive, are used in the experiment. Their descriptions are:

EXPERIENCE - Little - Tacket has been a senior for only about a year. Erwin is his first wholesaling client.

Extensive - Tacket has been a senior for four years. He has had extensive commercial experience with both small and large clients, including wholesalers.

Personal performance evaluations are commonly used in accounting firms. Audit managers normally fill out an evaluation report on the senior at the end of an audit engagement. These reports then form the basis of the senior's overall evaluation, which is used in determining pay and promotion. It is likely that a manager will take a senior's previous evaluation into account when reviewing his work. Performance
rating research consistently finds, not unexpectedly, that performance level and ability have the strongest effects on rating (Landy and Farr, 1979). Consequently, the senior's evaluation may influence the manager's reliance on the senior, that is, the senior's perceived source credibility. Two levels of performance evaluation, low and high, are used in the experiment. They are described as:

**EVALUATION - Low** - Tacket has received only mediocre ratings in the firm's formalized performance evaluations. Although he is regarded as competent in administration and personnel matters, his technical skill, his care in carrying out audit work, and the soundness of his audit judgments are considered barely acceptable. Your own appraisal of the quality of his work supports these ratings.

**High** - Tacket has received excellent ratings in the firm's formalized performance evaluations. He is regarded as competent in supervising field work, adept at dealing with audit and client personnel, qualified in technical matters, careful in his audit work, and sound in his audit judgments. Your own appraisal of the quality of his work supports these ratings.

The third independent variable, sampling approach, is included in the experiment to test whether the senior's rationale for his sample size recommendations affects the manager's behavior. Rather than contributing directly to source credibility, this variable may interact with source credibility. This possibility is consistent with the finding in the opinion change literature that the impact of source credibility is a function of the nature of information presented. For example, source credibility had a significant affect on positive information but not on negative information on the "likableness" of a person
Further motivation for including this variable is the finding by Mock and Turner (op. cit.) that, even when sample size recommendations differed significantly among seniors, the managers supervising the respective seniors accepted the senior's recommendations. The question raised by this finding is: does it matter how the senior justifies his judgment? In the test instrument, the senior is portrayed as using either a judgmental or statistical sampling approach. The judgmental approach describes the senior employing an anchoring heuristic. The use of judgment heuristics by auditors is receiving increased attention (Biddle and Joyce op. cit.). Subjects' responses may provide some insight on how audit managers view the use of this heuristic compared to the use of statistical sampling. Note that the statistical approach is really a mixed approach here in that a judgmental sample is first made of the minor strata, for example, accounts past due, and then a statistical sample size is calculated for the main stratum of accounts receivable. The descriptions used are:

- **SAMPLING APPROACH - Judgmental** - In a memo, Tacket explains that he arrived at the sample sizes he has recommended by adjusting last year's sample sizes to recognize the improvement he judged had taken place in the client's internal controls.

- **Statistical** - In a memo, Tacket explains that he derived the sample sizes he has recommended from sample size tables, having regard for the risk levels he judged acceptable.

The first two independent variables represent attributes of the audit senior that may act as surrogates for source credibility. While additional attributes could have been included, the two chosen are the
two most commonly references as associated with performance in the performance rating literature (Landy and Farr, op. cit.). The third independent variable is hypothesized to interact with source credibility. The fourth factor is not an independent variable in the sense of being manipulated across experimental situations; it is described now.

The importance of the issue to the judge (subject) has also been found to interact with source credibility. The importance of the issue has been studied in opinion change research under the title of ego-involvement (Rhine and Severance op. cit.). The evidence suggests that the impact of source credibility is not as strong when ego-involvement is high. This factor is considered by means of the four strata according to which accounts receivable are classified. A summary of the strata is contained in Table 4. The importance of each stratum in respect to confirmation is a function of its materiality in terms of nature and amount. It is hypothesized that the least important stratum is accounts over 90 days past due because of the small dollar amount involved. For confirmation purposes, the most important stratum is accounts over $2,500 because of the large value of each account.

<table>
<thead>
<tr>
<th>Stratum</th>
<th>Number</th>
<th>Amount</th>
</tr>
</thead>
<tbody>
<tr>
<td>Accounts over 90 days past due</td>
<td>14</td>
<td>$10,100</td>
</tr>
<tr>
<td>Accounts past due between 45 and 90 days</td>
<td>55</td>
<td>52,319</td>
</tr>
<tr>
<td>Accounts over $2,500</td>
<td>40</td>
<td>111,320</td>
</tr>
<tr>
<td>Other accounts</td>
<td>1,899</td>
<td>1,567,213</td>
</tr>
<tr>
<td>TOTAL</td>
<td>2,008</td>
<td>$1,740,952</td>
</tr>
</tbody>
</table>

TABLE 4
SUMMARY OF ACCOUNTS RECEIVABLE
Dependent Variables

For each experimental situation the subject is asked to provide six responses; four sample-size responses and two self-report responses. A sample-size judgment is required for each of the accounts receivable strata. An advantage of asking for a sample-size judgment is that it is a more objective dependent variable than the probability judgment required in the first stage of the experiment. Furthermore, auditors are continually having to determine specific sample sizes in an audit. Given that there is an unambiguous cost that depends on the level of confirmation testing, using sample sizes as a dependent variable tests whether the effect of the independent variable is strong enough to cause the subject to incur additional audit costs.

In addition to providing sample size estimates, subjects are asked to indicate: (1) the weight they placed on the compliance tests conducted by the senior in making their sample size judgments, and (2) the weight they placed on the sample size recommendations of the senior in making their sample size judgments. These two additional dependent variables are included in order to analyze cases where the subjects' perception of the senior, and the reliance placed on him, changes even though the subjects do not change their sample size estimates.

It is further hypothesized that there is a difference in the fineness of measurement between the various dependent variables. The weight given to the senior's sample size recommendations depends on the extent to which subjects feel they can rely on the senior to make the appropriate judgment. In contrast, the weight given to the
compliance tests conducted by the senior depends on the extent to which the subjects feel they can rely on the senior to conduct audit procedures. Possibly a subject's perception of the senior changes, yet he still feels the senior is sufficiently capable to conduct the audit procedures in question. In turn, even if subjects doubt the ability of the senior to conduct the audit procedures, subjects may not respond by altering their sample size judgments because of the sampling procedure's robustness to the added uncertainty, and the cost added by increasing the sample size. Therefore, in the extreme position, possibly the subject is sensitive to the various scenarios describing the senior, but the weight given to the senior's sample size recommendations is the only dependent variable to show statistical significance.

A question raised about self-insight measures is: do subjects really know how much weight they attach to a certain variable (Slovic and Lichtenstein, 1971). This issue is not a major concern here, as interest is not so much in the actual weighting, but in the change in weighting over experimental situations. Also, self-insight among auditors has generally been found to be high (see Table 1).

Research Hypotheses

In the second stage of the experiment, the subjects are asked to make sample size judgments based on work (the compliance tests) conducted by a certain audit senior. The motivation for the subjects to respond differently across experimental situations is essentially the same as in the first stage of the experiment; different descriptions of the senior are provided. In this second stage, the senior's credibility is not explicitly manipulated; rather, factors which have been
shown in psychological research to either affect or interact with source credibility are manipulated. However, while these factors may affect the subjects' perception of the senior, the sample size dependent variable may be too coarse a measure to catch the effect. Perhaps an auditor has a rule of thumb for determining sample sizes, for example 10 percent of the population, and that only extreme circumstances would make him depart from this rule. Two additional dependent variables are included in an attempt to deal with this possibility.

In all, the subjects provide six responses to each of eight experimental situations in which three independent variables, each having two levels, are manipulated. The purpose of the hypotheses tested in this stage of the experiment is to consider the third and fourth research objectives of the study. The hypotheses are stated in their null form.

**H5:** The senior's experience level has no affect on the subjects' responses.

This hypothesis should be rejected. If the subjects associate experience with performance, greater reliance should be placed on the compliance test results and the senior's recommended sample sizes when the senior has extensive experience, and less when he has little experience.

Greater reliance in terms of subjects' self-report responses is unambiguously expressed as higher values on the response scale. Greater reliance in terms of subjects' sample size judgments is not so unambiguously expressed as higher values on the response scale. The senior's recommended sample sizes are less than last year's sample sizes because of the improvements in the client's internal
control and the positive compliance test results. Greater reliance is suggested when the subjects' sample sizes approach those recommended by the senior. Exceptions to this rule, however, can be envisaged. They will not be pursued unless the results of the experiment suggest the presence of other behavior patterns.

H6: The senior's performance rating has no affect on the subjects' responses.

This hypothesis should be rejected. Greater reliance, as determined above, should be exhibited by the subjects when the senior has a high performance rating; less reliance when he has a low rating.

H7: The sampling approach adopted by the senior has no affect on the subjects' responses.

The outcome of the test of this hypothesis is unclear. A firm effect is possible because accounting firms differ in the emphasis they give to statistical sampling versus judgmental sampling. In respect to providing an insight into the operation of source credibility, the next hypothesis is of more interest.

H8: There is no interaction between the senior's sampling approach and the senior's experience level or performance rating on the subjects' responses.

This hypothesis should be rejected. According to the attitude change literature cited, the impact of source credibility is influenced by the nature of the information presented. As judgment or intuition plays a more important role in judgmental sampling than in statistical sampling, it is hypothesized that an interaction does exist. Specifically, more reliance is exhibited in the judgmental approach when the senior's experience level and performance rating are high.

H9: The independent variables do not affect the subject's sample size estimates for the relatively unimportant stratum of receivables.
H10: The independent variables do not affect the subject's sample size estimates for the relatively important stratum of receivables.

Hypotheses nine and ten relate to the fourth factor, the importance of the issue; they should be accepted. Attitude change literature suggests that source credibility has the greatest impact when the issue is unimportant (low ego-involvement). Thus, for the relatively unimportant stratum, the senior's sample size recommendations should be accepted irrespective of how the subjects perceive the senior's credibility. Conversely, for the relatively important stratum, subjects should follow their own sampling policy, ignoring the senior's recommendations. However, the subjects' policy may or may not be the same as that of the seniors; it should not change as the description of the senior changes.

Statistical Analysis

The statistical methodology in the second stage of the experiment is essentially the same as that used in the first stage. ANOVA is the main statistical testing procedure employed. The justification for the use of this procedure was outlined in the description of the first-stage experiment. Multivariate analysis of variance (MANOVA) is also used to take advantage of the existence of more than one dependent variable. Possibly sample size judgments are correlated with each other across the strata. Separate univariate tests for variables that are correlated are not independent of each other. MANOVA takes these intercorrelations into account. The other main procedure employed is a non-parametric procedure: multidimensional scaling (MDS). It is used to provide additional insight into subjects' self-report responses.
Hypotheses five through eight relate to the three independent variables described previously. As each variable has two levels, a test of each hypothesis is provided by a repeated measures 2^3 ANOVA. This ANOVA model is similar to the model used in the first stage outlined in Table 3, except that now there is one additional main effect and one additional level of interaction.

The ANOVA procedure is run for each of the dependent variables. Hypotheses nine and ten are tested with the 2^3 ANOVA's using the sample size judgment in the appropriate accounts receivable stratum as the dependent variable.

The subjects' sample size judgments are also compared with last year's sample sizes and the senior's sample size recommendations to see whether any specific groupings of subjects are apparent. In particular, a firm effect is tested for. Separate ANOVAs are then run for each group. MANOVA is run with the dependent variable vector first consisting of the sample sizes for the four accounts receivable strata and, second, consisting of the respective weights subjects' reported they attached to the compliance test results and the senior's sample size recommendations.

The subjects respond to the two self-report dependent variables by means of a seven-point scale. With such a scale, a familiar debate is whether the resulting data is ordinal or interval and, in turn, whether parametric or non-parametric statistics are the most appropriate. In accounting research the issues is often ignored (Grove and Savich, 1979). Parametric tests are advocated by various statisticians (see Anderson, 1961, and Kerlinger, 1973), and are frequently employed
in auditor judgment research (for example, Ashton op. cit., Ashton and Brown, op. cit., and Hamilton and Wright op. cit.). Parametric tests, specifically ANOVA, are used here but, in addition, non-parametric multidimensional scaling (MDS) is also utilized. MDS is a set of mathematical techniques that provide a method for uncovering the "hidden structure" or dimensions of a data base. Two of the set of techniques are used: ALSCAL and MDPREF.

Alternating Least Squares Scaling (ALSCAL) (Young and Lewychyj, 1979) is used to conduct a classical nonmetric MDS analysis. This analysis finds the location of the eight stimuli (descriptions of the senior) on the extracted dimensions for the subjects as a group. The data input for this analysis is a symmetric matrix of profile proximities between stimuli. The original subject-by-stimulus matrix (subjects' responses) is converted into the input-data matrix by a Euclidian distance transformation (Kruskel and Wish, 1979).

In addition to finding subjects' group preferences for the stimuli, Multidimensional Preference Analysis (MDPREF) finds each subject's preferences between the extracted dimensions. This technique enables the differences between subjects to be analysed. Subjects' responses in the form of a subject-by-stimulus matrix satisfy the data input requirements. The assumption underlying the use of MDPREF is that subjects' self-report responses, the weights given to certain items in forming their judgments, correspond to their preferences. That is, a higher weight implies a greater preference.
Pretest of the Test Instruments

The starting point for developing the case study around which both test instruments are built was the case study described in the auditing text by Arens and Loebbecke (op. cit.), and the case study utilized by Mock and Turner (op. cit.) in their research. Using these case studies as guides, the case study and the complete test instruments themselves were developed; these underwent many iterations before they were considered ready for pretesting.

Pretesting was conducted with two auditors who were also part-time accounting instructors at Ohio State University and two Ph.D. students at Ohio State University. Both Ph.D. students had previous auditing experience. Audit managers from the same population from which the subjects for the experiment were obtained were not used because of their value as scarce resources. The researcher believed that the auditing experience of the pretest subjects was sufficient to discover any potential problems with the test instruments.

The pretest uncovered no potential problems with either test instrument except for some minor questions with respect to wording, particularly as to the clarity of instructions. As a result of the pretest, some minor changes in wording were made. The subjects indicated that each test instrument could be completed conscientiously within half an hour.

Chapter Summary

The chapter began with a justification of the use of a two-stage experiment to examine the second, third and fourth research objectives. The appropriateness of using practicing audit managers as subjects was
also explained. Then the experimental design of each stage of the experi-
ment was described. Included in this description was the enumeration of
the specific hypotheses relating to the research objectives, and how the
hypotheses are to be tested. Statistical testing is to be carried out
primarily with ANOVA procedures.

After the successful pretest of the test instruments, the experi-
ment was conducted with the subjects, audit managers, in their respec-
tive offices in Columbus, Ohio. Thirty-five subjects participated in
the first stage of the experiment, and thirty-one of these subjects
returned their completed copies of the second-stage test instrument.
Statistical analysis of the data obtained from the experiment, together
with the interpretation of the results, is presented in the following
chapter.
CHAPTER V

STATISTICAL ANALYSIS AND INTERPRETATION OF THE RESULTS

The purpose of this chapter is to present the results of the experiment described in the previous chapter. The chapter consists of three presentations: (1) the results of the first-stage experiment, (2) the results of the second-stage experiment, and (3) the comments provided by subjects at the completion of the first-stage test instrument and in the post experimental questionnaire.

In the first two parts of the chapter, the tests of the assumptions underlying the ANOVA are discussed. For each stage, sections of the chapter analyze the subjects responses at overall, firm, and individual subject levels. Within the various sections, the results of the tests of the research hypotheses are presented. All hypotheses are statistically tested with the significance level of rejection provided. Graphical analyses, involving plots of the treatment means, are also provided to assist in the interpretation of the statistical tests. Additional analysis, which was described in the previous chapter, is presented in an attempt to provide further insight into subjects' behavior in response to the independent variables.

Stage One: Assumptions of the Analysis

The research hypotheses are tested by the F tests associated with the ANOVA model. Myers (1972, pp. 70-76 and pp. 175-186) and Winer 102.
(1971, pp. 273-283) discuss the assumptions underlying the ANOVA F tests and the implications of violations of the assumptions. These assumptions can be summarized as follows:

1. the error terms associated with each observation are independent and distributed as $N(0, \sigma^2_e)$,
2. the subject variables are independent and distributed independently of the error terms and interaction effects as $N(0, \sigma^2_i)$, and
3. the variances and covariances of the treatment populations are equal for all treatments, that is, the variance-covariance matrix exhibits compound symmetry.

Randomization of experimental procedures enables expected correlations among measurements to cancel as sample size increases, so that the independence assumptions are satisfied. Selection of subjects was discussed in the previous chapter. Both independent variables in the experiment were randomized across subjects, with the exception that all subjects responded to the situation of perfect source credibility first. This compromise to randomization was introduced because it was believed that it would be easier for subjects to respond to this situation first, and because this situation, as explained in the previous chapter, is an input into the model's predicted values. The compromise could have resulted in subjects adopting an anchoring strategy, anchoring on the perfect source credibility case. In a repeated measures design, this compromise to randomization does not involve the assignment of treatments to subjects, only the order in which the subjects received treatments. Furthermore, inspection of the responses suggested that no
order effect was present, although no statistical test could be conducted.

Normality depends upon the measure chosen. For example, the likelihood ratio, which is the measure used in the difference score analysis, tends to exhibit a skewed distribution. For this reason a log transformation was performed. The response frequencies for each of the eight situations and the difference scores for the six situations of less than perfect credibility were plotted to observe the distributional characteristics of the responses. No severe skewness was found. The difference scores were bell shaped, while the responses were generally unimodal and skewed slightly towards the lower end of the response scale.

The consequences of heterogeneity of covariance, or lack of compound symmetry, are greater than those for simple heterogeneity of variance in a between subject design. As the following analysis will show, see Table 5 for example, the mean squares for treatment-by-subject interactions are relatively large, which suggests that a non-additive model is appropriate. In the non-additive model, heterogeneity of covariance may cause a significant bias in the F ratios. Inspection of the variance-covariance matrix for responses, and Hartley's test (Winer, op. cit., p. 207) rejecting homogeneity of variance (tabled $F_{\text{max}} = 3.8$ for $p = .05$, while calculated $F_{\text{max}} = 11.0$), led to the adoption of Myer's recommendation of employing the conservative degrees of freedom F test (Myers op. cit., pp. 175-179). Subsequently, an effect classified as significant will also be significant at the .05 level under the conservative test, unless otherwise noted. This conservative test is
applied even when Hartley's test does not reject homogeneity of variance (for example, difference scores: \( F_{max}^{\text{tabled}} = 3.6 \) for \( p = .05 \), while calculated \( F_{max}^{\text{calculated}} = 2.74 \)). With the use of the conservative degrees of freedom test, the researcher believes the ANOVA F tests conducted are justified and reasonable.

**Analysis of Subjects' Responses**

The first three hypotheses are tested by the overall ANOVA run on subjects' responses. Table 5 and Figure 5 summarize the results of this ANOVA. Each hypothesis is now considered in turn.

**Hypothesis 1**

The first hypothesis concerns the effect of the source credibility independent variable:

\( H_1: \) The subjective probability that the client's system has functioned sufficiently well to justify substantial reliance on it is unaffected by the level of source credibility.

This hypothesis is rejected given the p-value of .0001. Duncan's test indicates a significant difference (\( \alpha = .05 \)) between each mean. The higher the level of the source's credibility, the higher the subjects' mean response, or subjective probability that the client's internal control system functioned sufficiently well to allow substantial reliance on it. This result suggests that the compliance test results (which supported substantial reliance) were discounted as the senior's source credibility fell.

**Hypothesis 2**

The second hypothesis concerns the effect of the confidence level independent variable.
### Table 5

#### Summary of ANOVA Results

<table>
<thead>
<tr>
<th>Source</th>
<th>DF</th>
<th>MS</th>
<th>F Ratio</th>
<th>P &gt; F</th>
<th>Duncan</th>
</tr>
</thead>
<tbody>
<tr>
<td>A (Confidence Level)</td>
<td>1</td>
<td>.0844</td>
<td>18.63</td>
<td>.0001</td>
<td></td>
</tr>
<tr>
<td>B (Source Credibility)</td>
<td>3</td>
<td>1.1173</td>
<td>60.56</td>
<td>.0001</td>
<td>B1&gt;B2&gt;B3&gt;B4</td>
</tr>
<tr>
<td>S (Subjects)</td>
<td>34</td>
<td>.988</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>A x B</td>
<td>3</td>
<td>.0001</td>
<td>.13</td>
<td>.9374</td>
<td></td>
</tr>
<tr>
<td>A x S</td>
<td>34</td>
<td>.0045</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>B x S</td>
<td>102</td>
<td>.0185</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>A x B x S</td>
<td>102</td>
<td>.0006</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Mean Score**

- A1 = 95%
- A2 = 90%

**Figure 5. Source Credibility Means by Confidence Level**
H2: The subjective probability that the client's system has functioned sufficiently well to justify substantial reliance on it is unaffected by the confidence level of the compliance tests.

This hypothesis is also rejected given a p-value of .0001. In general, subjects' subjective probabilities were higher when the compliance tests were conducted at the 95 percent level compared to the 90 percent level. This means that the vertical distance between curves A1 and A2 in Figure 5 is statistically significant.

Hypothesis 3

The third hypothesis concerns the existence of an interaction effect between the two independent variables:

H3: The impact of source credibility is unaffected by the confidence level of the compliance tests.

This hypothesis cannot be rejected given a p-value of .9374. In terms of Figure 5, rejection requires non-parallelism between the plots, which does not exist. Subjects did not discount source credibility more at the higher confidence level.

While subjects were clearly sensitive to the independent variables, including all levels of source credibility manipulated, the failure to reject the third hypothesis suggests subjects' behavior does not correspond to that of the model. However, these results have to be considered with caution as the large mean squares for subjects and treatment-by-subject interactions indicate differences existed between subjects. These differences are now investigated. The most obvious explanation is differences existing between firms.
Firm Effect

A priori reasoning might suggest that the most likely difference to appear between firms would relate to the confidence level variable, as the emphasis placed on statistical versus judgmental sampling differs between firms. Firm differences on source credibility could arise from such things as different evaluation procedures and different emphasis given to individual audit procedures, the tradeoff between compliance and substantive testing being one example.

The firm factor was tested by a one-way ANOVA on subjects' mean responses. No significant firm differences were found ($F[5,29] = 1.18, P = .34$). However, this is a rather inefficient test of differences between firms in respect to the source credibility and confidence level. A mixed model, $F \times (A \times B \times S)$, was tested with two subjects from F1, one subject from F2, along with F6 dropped from the analysis to provide equal firm size. Neither the firm factor nor any of the related interactions is significant, and the results of the other effects are the same as those reported in Table 5.

Separate ANOVAs were also run for each firm. The interesting result is that the confidence level is not significant for any firm (mean $p = .14$, min. $p = 0.06$ for F1; max. $p = 0.30$ for F6). The smaller degrees of freedom appear to be influencing these results. Except for firm six, source credibility is significant for each firm. What does differ across firms is the rate of discounting for source credibility, as shown by Duncan's tests in Figure 6. Again, however, care is required in interpreting these results. The mean square for subjects is large for each firm, suggesting that the meaningful differences
**Figure 6. Source Credibility Means by Firm, with Duncan's Multiple Range Test for Means**

**F1:** B1, B2 > B3 > B4*

**F2:** B1, B2 > B3, B4

**F3:** B1, B2, B3 > B4

**F4:** B1 > B3; B2, B3 > B4

**F5:** B1 > B3; B2 > B4

**F1-F5:** Firms one through five

The protected α level for these and all future applications of Duncan's test is .05.

*Read as the mean response at the first and second levels of source credibility was greater than the third level mean, but the difference between the first and second level means was not statistically significant.*
exist at the individual subject level.

Subject Effect

The subjects' behavior is summarized in Table 6. A general reaction to source credibility is apparent. While the overall ANOVA indicates that a difference between 90 percent and 95 percent confidence level exists, many subjects saw no difference within the context of the case study. This may explain to some extent why hypothesis 3 could not be rejected. The lack of an appropriate error term prevented individual subjects' behavior being modelled. Further insight into subjects' behavior is achieved by grouping subjects other than by firm. A cluster analysis was performed. The subjects were the objects classified into clusters and the subjects' eight responses were the variables on which the clusters were formed. The results are presented in Table 7.

A difficulty with cluster analysis is deciding upon the appropriate number of clusters. The approach adopted was to first inspect the cluster map and then to run several cluster analyses. The researcher selected the analysis that subjectively appeared to minimize differences between subjects within clusters and maximize differences between subjects across clusters. Most subjects fell into clusters one and three. Cluster four, consisting of two subjects who severely discounted reductions in source credibility, was the second cluster formed. Cluster three was the third cluster formed, while cluster two was the additional cluster in moving from three to four clusters. Inspection of subjects' responses in each cluster supported this particular grouping. The lack of a strong firm representation within the clusters (see Table 7) provides additional support for the finding of no
### TABLE 6

**SUMMARY OF INDIVIDUAL RESPONSE BEHAVIOR**

<table>
<thead>
<tr>
<th></th>
<th>F1</th>
<th>F2</th>
<th>F3</th>
<th>F4</th>
<th>F5</th>
<th>F6</th>
<th>Σ</th>
</tr>
</thead>
<tbody>
<tr>
<td>YES</td>
<td>4</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>2</td>
<td>2</td>
<td>20</td>
</tr>
<tr>
<td>NO</td>
<td>4</td>
<td>4</td>
<td>2</td>
<td>1</td>
<td>3</td>
<td>1</td>
<td>15</td>
</tr>
<tr>
<td>YES</td>
<td>8</td>
<td>7</td>
<td>5</td>
<td>6</td>
<td>5</td>
<td>2</td>
<td>33</td>
</tr>
<tr>
<td>NO</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1</td>
<td>2</td>
</tr>
</tbody>
</table>

A : confidence level  
B : source credibility  
YES : responses differed on at least one level of factor  
NO : responses the same on all levels of factor

### TABLE 7

**FIRM REPRESENTATION WITHIN CLUSTERS**

<table>
<thead>
<tr>
<th></th>
<th>F1</th>
<th>F2</th>
<th>F3</th>
<th>F4</th>
<th>F5</th>
<th>F6</th>
<th>Σ</th>
</tr>
</thead>
<tbody>
<tr>
<td>C1</td>
<td>5</td>
<td>3</td>
<td>3</td>
<td>2</td>
<td>1</td>
<td>2</td>
<td>16</td>
</tr>
<tr>
<td>C2</td>
<td>1</td>
<td>1</td>
<td></td>
<td>2</td>
<td></td>
<td></td>
<td>4</td>
</tr>
<tr>
<td>C3</td>
<td>2</td>
<td>3</td>
<td>3</td>
<td></td>
<td>4</td>
<td>1</td>
<td>13</td>
</tr>
<tr>
<td>C4</td>
<td></td>
<td></td>
<td></td>
<td>2</td>
<td></td>
<td></td>
<td>2</td>
</tr>
</tbody>
</table>

C1-C4: Clusters one through four
significant firm differences.

Once the clusters were decided upon, the cluster factor was tested by a one-way ANOVA on subjects' mean responses. In contrast to the firm factor, the cluster factor provided significant ($F[3,31] = 38.77$, $p < .0001$). This test should be viewed as a minimum criteria for the clustering analysis to be worthwhile. As subjects were grouped to maximize the differences between clusters, it was expected that the cluster factor would be significant. To determine differences between clusters, an ANOVA ($A \times B \times S$) was run for each cluster. The results are presented in Table 8 and Figure 7. Again the large subject mean squares suggested the clusters did not capture all subject differences. Two observations stand out. First is the difference in overall mean responses between clusters. Duncan's multiple range test gave $C_1, C_2 > C_3 > C_4$. A t test at B1 indicated a significant difference between $C_1$ and $C_3$ ($t[27] = 38.29$, $p < .01$), which are the largest clusters. The averaging across source credibility levels apparently caused Duncan's test to miss the significant differences between $C_1$ and $C_2$. The difference between $C_1$ and $C_2$ was also significant at B1 ($t[18] = 9.64$, $p < .01$).

The second observation concerns the reaction to reduced source credibility by subjects in the various clusters. Groups $C_1$ and $C_3$, which contain 29 of the 35 subjects, approximated linear mean response behavior, with $C_3$ possessing the greater discount rate. The four subjects comprising $C_2$ indicated no significant difference between B1 and B2 (Duncan's test), but exhibited substantial discounting behavior between B3 and B4 levels of source credibility. This group also
### TABLE 8
SUMMARY OF ANOVA BY CLUSTER

<table>
<thead>
<tr>
<th>CLUSTER</th>
<th>SOURCE</th>
<th>DF</th>
<th>MS</th>
<th>F RATIO</th>
<th>P &gt; F</th>
<th>DUNCAN</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>C1</td>
<td>A</td>
<td>1,15</td>
<td>.0378</td>
<td>6.05</td>
<td>.0265</td>
<td></td>
</tr>
<tr>
<td></td>
<td>B</td>
<td>3,45</td>
<td>.0947</td>
<td>29.39</td>
<td>.0001</td>
<td>B1&gt;B2&gt;B3&gt;B4</td>
</tr>
<tr>
<td></td>
<td>S</td>
<td>15,45</td>
<td>.0215</td>
<td>278.64</td>
<td>.0001*</td>
<td></td>
</tr>
<tr>
<td>C2</td>
<td>A</td>
<td>1,3</td>
<td>.0181</td>
<td>3.89</td>
<td>.1432</td>
<td></td>
</tr>
<tr>
<td></td>
<td>B</td>
<td>3,9</td>
<td>.2706</td>
<td>78.31</td>
<td>.0001</td>
<td>B1&gt;B3&gt;B4</td>
</tr>
<tr>
<td></td>
<td>AxB</td>
<td>3,9</td>
<td>.0028</td>
<td>4.24</td>
<td>.0398</td>
<td></td>
</tr>
<tr>
<td></td>
<td>S</td>
<td>3,9</td>
<td>.008</td>
<td>12.37</td>
<td>.0015</td>
<td></td>
</tr>
<tr>
<td>C3</td>
<td>A</td>
<td>1,12</td>
<td>.0295</td>
<td>8.39</td>
<td>.0134</td>
<td>B1&gt;B2&gt;B3&gt;B4</td>
</tr>
<tr>
<td></td>
<td>B</td>
<td>3,36</td>
<td>.7717</td>
<td>80.88</td>
<td>.0001</td>
<td></td>
</tr>
<tr>
<td></td>
<td>S</td>
<td>12,36</td>
<td>.0248</td>
<td>25.32</td>
<td>.0018</td>
<td></td>
</tr>
<tr>
<td>C4</td>
<td>A</td>
<td>1,1</td>
<td>.0014</td>
<td>1.0</td>
<td>.5</td>
<td></td>
</tr>
<tr>
<td></td>
<td>B</td>
<td>3,3</td>
<td>.4033</td>
<td>12.95</td>
<td>.032</td>
<td>B1,B2&gt;B4</td>
</tr>
<tr>
<td></td>
<td>S</td>
<td>1,3</td>
<td>.0638</td>
<td>111.28</td>
<td>.0018</td>
<td></td>
</tr>
</tbody>
</table>

*A negatively biased test, included to give indication of existence of subject differences (Myers, 1976, p. 175).

+Not significant at $\alpha = .05$ under the conservative test.
FIGURE 7. SOURCE CREDIBILITY MEANS BY CLUSTER
provided the first instance of a significant A by B interaction. It involved a divergence of the AB means at lower levels of source credibility - the opposite of the model's prediction. The final two subjects comprising C4 exhibited lower mean responses at each source credibility level.

To summarize the analysis so far, the results found are consistent with the psychology finding that information is discounted as the credibility of the information source falls. The audit managers participating in the experiment indicated a lower probability of the client's internal control system being classified as strong, in the face of evidence suggesting substantial reliance be placed upon it, as the credibility of the senior providing the evidence dropped. Differences between managers appeared to be due to individual differences rather than firm differences. Managers differed in the extent they reduced their belief in the client's system being strong, given the source credibility reduction, and also differed in their mean appraisal of the client's system.

Analysis of Difference Scores

The results of comparing the subjects' responses with the normative model's predictions are now presented. This analysis incorporates information not as yet considered: the subjects' evaluation of the client's system prior to the receipt of compliance test results. This analysis provides insight into the above findings of different mean responses between clusters, the apparent lack of an interaction (multiplicative) effect, and the greater discounting occurring at the lower levels of source credibility. The normative model predicted a source
credibility by confidence level interaction and greater discounting for initial reductions in source credibility.

As explained in the previous chapter, the dependent variable in the difference analysis was the difference between the subject's inferred log likelihood ratio and the model's derived log likelihood ratio. Three points need to be made about the following analysis. First, a positive (negative) difference indicates that the subject attached more (less) diagnosticity to the information than did the model. Second, there are only three levels of source credibility. As explained in the previous chapter, the subject's responses under perfect credibility (Bl) were required to obtain the model's values so there could be no difference between the model and subject at this level. Third, two subjects who were inconsistent in their responses compared to their priors were excluded from the analysis. Three subjects viewed the compliance test results at the 90 percent confidence level as evidence against the client's system being strong. Under these circumstances a decline in source credibility should lead to a movement in the subjects' probabilistic responses towards their priors. Two of the three subjects did not respond in this direction and were excluded from the analysis.

**Hypothesis 4**

The fourth hypothesis concerns the extent to which subjects' responses corresponded to those of the model presented in the previous chapter.

**H4**: The derived likelihood ratios are approximated by the likelihood ratios calculated from the model.
### TABLE 9
SUMMARY OF ANOVA RESULTS - DIFFERENCE SCORES

<table>
<thead>
<tr>
<th>SOURCE</th>
<th>DF</th>
<th>MS</th>
<th>F RATIO</th>
<th>P&gt;F</th>
<th>DUNCAN</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>1</td>
<td>.09</td>
<td>2.16</td>
<td>.152</td>
<td></td>
</tr>
<tr>
<td>B</td>
<td>2</td>
<td>1.782</td>
<td>12.35</td>
<td>.0001</td>
<td>B2, B3&gt;B4</td>
</tr>
<tr>
<td>S</td>
<td>32</td>
<td>.601</td>
<td>124.99</td>
<td>.0001</td>
<td></td>
</tr>
<tr>
<td>AxB</td>
<td>2</td>
<td>.001</td>
<td>.27</td>
<td>.76</td>
<td></td>
</tr>
<tr>
<td>AxS</td>
<td>22</td>
<td>.042</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>BxS</td>
<td>64</td>
<td>.145</td>
<td>.005</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

LOG
LIKELIHOOD
DIFFERENCES

![Graph](image)

**Figure 8.** Mean differences as a function of source credibility for both statistical confidence levels

A1 = .95
A2 = .90
The results of the overall ANOVA run on the log likelihood difference scores are summarized in Table 9 and Figure 8. The results suggest the hypothesis should be rejected. Non-zero differences exist and the source credibility factor is significant \( (p < .0001) \) in explaining these difference scores, while the confidence level is not. Subjects, in general, appear to have revised their prior probabilities less than the model predicted, particularly at lower levels of source credibility. This finding is consistent with the general finding of conservatism in Bayesian studies, and inconsistent with previous cascaded inference research which found that subjects discounted source credibility insufficiently ("best guess" hypothesis). The not significant finding for the confidence level factor means that the vertical distance between the plots in Figure 8 is not significant.

**Firm Effect**

The large subject mean squares indicates that significant subject differences existed. A one-way ANOVA on subjects' mean differences gave a not significant result for the firm factor \( (F[5,27] = 2.26, \ p < .08) \). Neither statistical confidence level nor source credibility is significant in the individual firm ANOVA's at the .05 level (exception: \( F_5, \) source credibility \( F[2,6] = 12.77, \ p < .01 \)), while the subject factor is significant at the .001 level for each firm. An examination of subjects' responses by firm supports the contention that subject differences rather than firm differences are responsible for the observed variability.
Subject Effect

A cluster analysis was performed on the difference scores and again the question of the appropriate number of clusters arose: in the n=3 case one group consisted of two subjects who had large negative mean differences, a second group with less extreme negative mean differences, and a third group with positive mean differences. Inspection of the cluster map, however, indicated that it would be informative to decompose each of the two main groups into separate groups. This was achieved in an n=8 analysis in which four clusters each consisted of a single subject. For ease of presentation, the four single subject clusters were combined to form a fifth cluster (C5). The cluster factor was significant in a one-way ANOVA on subjects' mean differences (F[3,25] = 61.13, p < .0001; Duncan: C4, C3 > C2 > C1).

Results from the analysis of the individual clusters are summarized in Tables 10 and 11 and Figure 9. While possibly the distribution of firms among the clusters is not entirely random, firms' representation in the clusters do not form a pattern and, unfortunately, the cell sizes are too small for statistical testing.

Two relationships are useful in drawing conclusions from the data and interpreting Figure 9. First, non-parallelism of a plot with the horizontal axis means that source credibility contributes to or explains to some extent the difference between the subject's values and the model's values. Therefore, a horizontal plot (no significant differences between source credibility treatment means) implies that the mean group discounting for source credibility approximates the model's discounting. A negatively sloped plot, for example, means that the
### TABLE 10

FIRM REPRESENTATION WITHIN CLUSTERS - DIFFERENCE SCORES

<table>
<thead>
<tr>
<th></th>
<th>F1</th>
<th>F2</th>
<th>F3</th>
<th>F4</th>
<th>F5</th>
<th>F6</th>
<th>Σ</th>
</tr>
</thead>
<tbody>
<tr>
<td>C1</td>
<td>2</td>
<td>--</td>
<td>--</td>
<td>--</td>
<td>2</td>
<td>1</td>
<td>5</td>
</tr>
<tr>
<td>C2</td>
<td>3</td>
<td>4</td>
<td>--</td>
<td>3</td>
<td>2</td>
<td>--</td>
<td>12</td>
</tr>
<tr>
<td>C3</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td>--</td>
<td>--</td>
<td>5</td>
</tr>
<tr>
<td>C4</td>
<td>--</td>
<td>1</td>
<td>3</td>
<td>1</td>
<td>--</td>
<td>2</td>
<td>7</td>
</tr>
<tr>
<td>C5</td>
<td>2</td>
<td>1</td>
<td>--</td>
<td>1</td>
<td>--</td>
<td>--</td>
<td>4</td>
</tr>
</tbody>
</table>

### TABLE 11

SUMMARY OF ANOVA BY CLUSTER - DIFFERENCE SCORES

<table>
<thead>
<tr>
<th>CLUSTER</th>
<th>SOURCE</th>
<th>DF</th>
<th>MS</th>
<th>F RATIO</th>
<th>P &gt; F</th>
<th>DUNCAN</th>
</tr>
</thead>
<tbody>
<tr>
<td>C1</td>
<td>A</td>
<td>1</td>
<td>.002</td>
<td>.06</td>
<td>.823</td>
<td>B2=B3=B4</td>
</tr>
<tr>
<td></td>
<td>B</td>
<td>2</td>
<td>.149</td>
<td>2.0</td>
<td>.198</td>
<td></td>
</tr>
<tr>
<td></td>
<td>S</td>
<td>4</td>
<td>.044</td>
<td>20.58</td>
<td>.001</td>
<td></td>
</tr>
<tr>
<td>C2</td>
<td>A</td>
<td>1</td>
<td>.001</td>
<td>.11</td>
<td>.748</td>
<td>B2&gt;B3,B4</td>
</tr>
<tr>
<td></td>
<td>B</td>
<td>2</td>
<td>.441</td>
<td>11.34</td>
<td>.0004</td>
<td></td>
</tr>
<tr>
<td></td>
<td>S</td>
<td>11</td>
<td>.042</td>
<td>40.33</td>
<td>.0001</td>
<td></td>
</tr>
<tr>
<td>C3</td>
<td>A</td>
<td>1</td>
<td>.052</td>
<td>4.12</td>
<td>.112</td>
<td>B2,B3&gt;B4</td>
</tr>
<tr>
<td></td>
<td>B</td>
<td>2</td>
<td>.472</td>
<td>11.3</td>
<td>.005</td>
<td></td>
</tr>
<tr>
<td></td>
<td>S</td>
<td>4</td>
<td>.019</td>
<td>1.83</td>
<td>.217</td>
<td></td>
</tr>
<tr>
<td>C4</td>
<td>A</td>
<td>1</td>
<td>.01</td>
<td>.36</td>
<td>.569</td>
<td>B2&gt;B3,B4</td>
</tr>
<tr>
<td></td>
<td>B</td>
<td>2</td>
<td>.0359</td>
<td>3.36</td>
<td>.069</td>
<td></td>
</tr>
<tr>
<td></td>
<td>S</td>
<td>6</td>
<td>.111</td>
<td>72.31</td>
<td>.001</td>
<td></td>
</tr>
<tr>
<td>C5 SUBJECT MEANS: S</td>
<td>B2</td>
<td>B3</td>
<td>B4</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>.23</td>
<td>-.49</td>
<td>-.61</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>.04</td>
<td>-.18</td>
<td>-1.05</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>.68</td>
<td>-.06</td>
<td>-.81</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>.23</td>
<td>-.78</td>
<td>-2.52</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
FIGURE 9. MEAN DIFFERENCES AS A FUNCTION OF SOURCE CREDIBILITY BY CLUSTER
mean group discount rate is greater than that of the model. The second relationship also concerns the position of the plot in relation to the horizontal axis through zero. A plot above or below the zero point on the vertical scale implies something about how group members revised their prior beliefs in the face of new information. A plot below zero indicates conservative revision compared to Bayesian revision.

The finding of previous cascaded inference research is that people do not discount reductions in source credibility sufficiently. Only C4 containing seven subjects exhibited behavior consistent with this prior research. Even for C4, source credibility is not significant at the .05 level in explaining difference scores. This result, as evidenced by the reasonably horizontal plot for C4 in Figure 9, suggests that subjects in C4 exhibited more counter-conservatism in the Bayesian sense than insufficient discounting for source credibility. On the other hand, the five subjects in C1 primarily exhibited Bayesian conservatism, while their discounting for source credibility approximated that of the models given the not significant source credibility factor. The significant source credibility factor and the negatively sloped plots for the 17 subjects in C2 and C3 suggests that they discounted reductions in source credibility excessively. Clearly, the previous finding of insufficient adjustment or discounting for source credibility is not supported.

No inferences are possible about subjects' behavior between the 100 percent and 90 percent source credibility levels because, as explained previously, the subjects' and models' values were forced to be equal at the 100 percent level. Nevertheless, another observation concerning
subjects' reaction to source credibility is that the excessive discounting for source credibility occurred at different levels of source credibility. There appears to be a difference between the subjects on how they viewed an initial reduction in source credibility. The model's discounting was shown previously to be the largest for initial reductions in source credibility. Groups C3 and C4 do not appear to take this view while Group C1's discounting is greater than that of the model. Group C3 however, in contrast to group C4, discounted excessively for the reduction in source credibility from 80 percent (B3) to 70 percent (B4). Note that B3 > B4 under Duncan's test for C3 while this same mean comparison is not significant for C1 and C2. Consequently, in terms of group means, the majority of subjects tended to exhibit some degree of excessive discounting for source credibility, but at different levels of source credibility.

The universal finding of previous research is that subjects' behavior when compared to the Bayesian model is conservative; that is, subjects do not take into account the full inferential value of the information in revising their priors. This finding is supported here, with the exception of C4. The largely horizontal plots indicated that, as a group, the five subjects in C5 were conservative, while the seven subjects in C4 were counter-conservative in the revision of their priors. The twelve subjects in C2 appeared to be also conservative, while the behavior of subjects in C3 reflects a strong reaction to the level of source credibility.
Additional Analysis

Previous studies of auditor judgment usually provide some correlation measure of the degree of consensus (inter-subject agreement) among the subjects' responses. Such a measure was computed for the subjects in the present study. The mean Pearson product-moment correlation coefficient for all pairwise comparisons between the 35 subjects is .716, and considering F1-F6 separately, is .70, .69, .48, .76, .92 and .44 respectively. The mean correlation coefficient is similar to that of Ashton (1974; .70) and Hamilton and Wright (1977, overall .66, subjects with more than three years experience .777). Subjects in both these studies evaluated the strength of internal control systems. In other tasks, however, consensus was not so high (Joyce, 1976, audit plan judgments .373; Weber, 1977, judgments of dollar errors in inventory .379).

On the basis of the information presented in the case study and the subject's experience with similar clients, but before learning of the compliance test results, the subject's prior probability of the client's system being strong was elicited. The majority of subjects considered it more likely that the client's system was strong but extreme prior beliefs were recorded (mean = .678, standard deviation = .193, minimum = .25, maximum = .95). To what extent subjects' priors reflected specific items of information in the case study on which subjects focused, and to what extent the priors reflected the subjects' general attitude to a new client, cannot be determined.

The differences in subjects' priors raises the question of whether there is a relationship between subjects' relative priors and relative
mean responses. The cluster analysis of subjects' responses indicated that differences did exist between subjects' mean responses. Thus, do subjects with relatively higher priors also have relatively higher responses? The Pearson product-moment correlation coefficient was calculated between subjects' priors and mean responses. A coefficient of .40 was obtained with a p-value of .02. This result suggests that subjects' priors did affect their responses; two implications follow. First, because subjects took their priors into account, they can be considered, to some extent at least, Bayesian. The difference analysis indicates that subjects were not completely Bayesian, but generally conservative in a Bayesian sense. The second implication is that differences between subjects' responses are, due to differences in the interpretation of the description of the clients' system, reflected in the priors, as well as differences in the interpretation of the compliance test results.

Stage Two: Assumptions of the Analysis

The results of the second stage of the experiment are now presented. The research hypotheses are tested by the F tests associated with the ANOVA and MANOVA models. The assumptions underlying these tests with the implications of their violation were discussed earlier in the chapter under the introduction to the first-stage experiment. For a repeated measures design, the assumptions concern independence, normality, and compound symmetry. To achieve the appropriate independence, the experimental procedures were randomized. Normality was appraised by plotting the sample size responses. The distributions were found to be reasonably symmetric and unimodal. As explained when discussing the
first stage, the assumption of compound symmetry is the critical assumption in a repeated measures design. The conservative degree of freedom test was used in the first stage because of the uncertainty surrounding compound symmetry. Compound symmetry does not appear to be violated in the second stage. This conclusion was reached by inspecting the variance-covariance matrices, and conducting Hartley's test for homogeneity of variance. None of the tests could be rejected at the .05 level.

Analysis of Subjects' Responses

Hypotheses five, six and seven relate to the factors, experience, performance rating and sampling approach respectively, which were manipulated as independent variables. These factors were hypothesized to directly affect the credibility of the audit senior and therefore alter the reliance placed upon him. Hypotheses eight, nine and ten concern refining the influence of source credibility on subjects' behavior based on findings in the relevant psychology literature.

As the subjects made four sample size judgments and provided two self-report responses for each of eight alternative situations, several separate analyses were possible. ANOVAs were run on each dependent variable and a seventh dependent variable which was the total sample selected, consisting of the sum of the four sample sizes. Finally, a MANOVA was run using the four sample size judgments as the multiple dependent variables. A summary of the results is contained in Table 12, while the individual analyses are contained in Tables 13 through 20.
### TABLE 12

SUMMARY OF THE RESULTS OF THE SECOND-STAGE EXPERIMENT

<table>
<thead>
<tr>
<th>Effect</th>
<th>D1</th>
<th>D2</th>
<th>D3</th>
<th>D4</th>
<th>E</th>
<th>G</th>
<th>H</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>.132</td>
<td>.241</td>
<td>.009</td>
<td>.850</td>
<td>.210</td>
<td>.0001</td>
<td>.0001</td>
</tr>
<tr>
<td></td>
<td>A1&lt;A2</td>
<td>A1&gt;A2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>B</td>
<td>.006</td>
<td>.035</td>
<td>.0002</td>
<td>.008</td>
<td>.002</td>
<td>.0001</td>
<td>.0001</td>
</tr>
<tr>
<td>C</td>
<td>.270</td>
<td>.381</td>
<td>.199</td>
<td>.317</td>
<td>.567</td>
<td>.312</td>
<td>.028</td>
</tr>
<tr>
<td></td>
<td>C1&gt;C2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

D1: sample of accounts over 90 days past due
D2: sample of accounts past due between 45 and 90 days
D3: sample of accounts over $2,500
D4: sample of other accounts
E: total sample size
G: weight placed on compliance test results
H: weight placed on sample size recommendations
A: experience - A1: extensive, A2: little
B: performance rating - B1: high, B2: low
C: sampling approach - C1: statistical, C2: judgmental

The table value is the p-value.
### TABLE 13

MANOVA RESULTS FOR D1-D4

<table>
<thead>
<tr>
<th>SOURCE</th>
<th>DF</th>
<th>F RATIO</th>
<th>P&gt;F</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>4,26</td>
<td>3.53</td>
<td>.0198</td>
</tr>
<tr>
<td>B</td>
<td>4,26</td>
<td>5.72</td>
<td>.0019</td>
</tr>
<tr>
<td>C</td>
<td>4,26</td>
<td>1.19</td>
<td>.3391</td>
</tr>
<tr>
<td>A x B</td>
<td>4,26</td>
<td>.87</td>
<td>.4939</td>
</tr>
<tr>
<td>A x C</td>
<td>4,26</td>
<td>.90</td>
<td>.4780</td>
</tr>
<tr>
<td>B x C</td>
<td>4,26</td>
<td>.19</td>
<td>.9409</td>
</tr>
<tr>
<td>A x B x C</td>
<td>4,26</td>
<td>.82</td>
<td>.5233</td>
</tr>
</tbody>
</table>

### TABLE 14

ANOVA RESULTS FOR E

<table>
<thead>
<tr>
<th>SOURCE</th>
<th>DF</th>
<th>MS</th>
<th>F RATIO</th>
<th>P&gt;F</th>
<th>MEANS</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>1</td>
<td>455.22</td>
<td>1.61</td>
<td>.2136</td>
<td></td>
</tr>
<tr>
<td>B</td>
<td>1</td>
<td>91976.52</td>
<td>11.89</td>
<td>.0017</td>
<td>B1&gt;B2</td>
</tr>
<tr>
<td>C</td>
<td>1</td>
<td>329.82</td>
<td>.33</td>
<td>.5672</td>
<td></td>
</tr>
<tr>
<td>S</td>
<td>30</td>
<td>44161.32</td>
<td>.33</td>
<td>.5672</td>
<td></td>
</tr>
<tr>
<td>A x B</td>
<td>1</td>
<td>28.45</td>
<td>.14</td>
<td>.7091</td>
<td></td>
</tr>
<tr>
<td>A x C</td>
<td>1</td>
<td>76.79</td>
<td>.35</td>
<td>.5592</td>
<td></td>
</tr>
<tr>
<td>B x C</td>
<td>1</td>
<td>564.02</td>
<td>.72</td>
<td>.4404</td>
<td></td>
</tr>
<tr>
<td>A x S</td>
<td>30</td>
<td>281.93</td>
<td>.33</td>
<td>.5672</td>
<td></td>
</tr>
<tr>
<td>B x S</td>
<td>30</td>
<td>7737.18</td>
<td>.33</td>
<td>.5672</td>
<td></td>
</tr>
<tr>
<td>C x S</td>
<td>30</td>
<td>995.23</td>
<td>.33</td>
<td>.5672</td>
<td></td>
</tr>
<tr>
<td>A x B x C</td>
<td>1</td>
<td>358.08</td>
<td>1.52</td>
<td>.2272</td>
<td></td>
</tr>
</tbody>
</table>
### TABLE 15

**ANOVA RESULTS FOR D1**

<table>
<thead>
<tr>
<th>SOURCE</th>
<th>DF</th>
<th>MS</th>
<th>F RATIO</th>
<th>P&gt;F</th>
<th>MEANS</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>1</td>
<td>33.75</td>
<td>2.41</td>
<td>.1317</td>
<td></td>
</tr>
<tr>
<td>B</td>
<td>1</td>
<td>98.82</td>
<td>8.94</td>
<td>.0057</td>
<td>B₁&gt;B₂</td>
</tr>
<tr>
<td>C</td>
<td>1</td>
<td>2.82</td>
<td>1.26</td>
<td>.2703</td>
<td></td>
</tr>
<tr>
<td>S</td>
<td>29</td>
<td>180.66</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>A x B</td>
<td>1</td>
<td>1.07</td>
<td>.70</td>
<td>.4109</td>
<td></td>
</tr>
<tr>
<td>A x C</td>
<td>1</td>
<td>0.07</td>
<td>.10</td>
<td>.7547</td>
<td></td>
</tr>
<tr>
<td>B x C</td>
<td>1</td>
<td>1.67</td>
<td>.70</td>
<td>.4089</td>
<td></td>
</tr>
<tr>
<td>A x S</td>
<td>29</td>
<td>14.03</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>B x S</td>
<td>29</td>
<td>11.06</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>C x S</td>
<td>29</td>
<td>2.23</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>A x B x C</td>
<td>1</td>
<td>.42</td>
<td>.55</td>
<td>.4654</td>
<td></td>
</tr>
</tbody>
</table>

### TABLE 16

**ANOVA RESULTS FOR D2**

<table>
<thead>
<tr>
<th>SOURCE</th>
<th>DF</th>
<th>MS</th>
<th>F RATIO</th>
<th>P&gt;F</th>
<th>MEANS</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>1</td>
<td>9.6</td>
<td>1.43</td>
<td>.2408</td>
<td></td>
</tr>
<tr>
<td>B</td>
<td>1</td>
<td>493.07</td>
<td>4.90</td>
<td>.0349</td>
<td>B₁&gt;B₂</td>
</tr>
<tr>
<td>C</td>
<td>1</td>
<td>18.15</td>
<td>.79</td>
<td>.3810</td>
<td></td>
</tr>
<tr>
<td>S</td>
<td>29</td>
<td>786.50</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>A x B</td>
<td>1</td>
<td>19.27</td>
<td>.92</td>
<td>.3462</td>
<td></td>
</tr>
<tr>
<td>A x C</td>
<td>1</td>
<td>14.01</td>
<td>1.88</td>
<td>.1814</td>
<td></td>
</tr>
<tr>
<td>B x C</td>
<td>1</td>
<td>.15</td>
<td>.02</td>
<td>.906</td>
<td></td>
</tr>
<tr>
<td>A x S</td>
<td>29</td>
<td>8.69</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>B x S</td>
<td>29</td>
<td>100.69</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>C x S</td>
<td>29</td>
<td>22.93</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>A x B x C</td>
<td>1</td>
<td>.42</td>
<td>.06</td>
<td>.8077</td>
<td></td>
</tr>
</tbody>
</table>
### TABLE 17

**ANOVA RESULTS FOR D3**

<table>
<thead>
<tr>
<th>SOURCE</th>
<th>DF</th>
<th>MS</th>
<th>F RATIO</th>
<th>P&gt;F</th>
<th>MEANS</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>1</td>
<td>86.4</td>
<td>7.91</td>
<td>.0087</td>
<td>A1&gt;A2</td>
</tr>
<tr>
<td>B</td>
<td>1</td>
<td>1530.15</td>
<td>17.52</td>
<td>.0002</td>
<td>B1&gt;B2</td>
</tr>
<tr>
<td>C</td>
<td>1</td>
<td>109.35</td>
<td>1.73</td>
<td>.1985</td>
<td></td>
</tr>
<tr>
<td>S</td>
<td>29</td>
<td>1150.77</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>A x B</td>
<td>1</td>
<td>77.07</td>
<td>2.55</td>
<td>.1210</td>
<td></td>
</tr>
<tr>
<td>A x C</td>
<td>1</td>
<td>.07</td>
<td>.02</td>
<td>.9030</td>
<td></td>
</tr>
<tr>
<td>B x C</td>
<td>1</td>
<td>2.02</td>
<td>.20</td>
<td>.6559</td>
<td></td>
</tr>
<tr>
<td>A x S</td>
<td>29</td>
<td>10.92</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>B x S</td>
<td>29</td>
<td>87.36</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>C x S</td>
<td>29</td>
<td>63.14</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>A x B x C</td>
<td>1</td>
<td>2.40</td>
<td>.16</td>
<td>.6889</td>
<td></td>
</tr>
</tbody>
</table>

### TABLE 18

**ANOVA RESULTS FOR D4**

<table>
<thead>
<tr>
<th>SOURCE</th>
<th>DF</th>
<th>MS</th>
<th>F RATIO</th>
<th>P&gt;F</th>
<th>MEANS</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>1</td>
<td>12.15</td>
<td>.04</td>
<td>.8504</td>
<td></td>
</tr>
<tr>
<td>B</td>
<td>1</td>
<td>56181.60</td>
<td>8.78</td>
<td>.0060</td>
<td>B1&gt;B2</td>
</tr>
<tr>
<td>C</td>
<td>1</td>
<td>1315.00</td>
<td>1.04</td>
<td>.3166</td>
<td></td>
</tr>
<tr>
<td>S</td>
<td>29</td>
<td>32207.37</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>A x B</td>
<td>1</td>
<td>117.60</td>
<td>.69</td>
<td>.4118</td>
<td></td>
</tr>
<tr>
<td>A x C</td>
<td>1</td>
<td>147.27</td>
<td>.79</td>
<td>.3819</td>
<td></td>
</tr>
<tr>
<td>B x C</td>
<td>1</td>
<td>442.82</td>
<td>.74</td>
<td>.3976</td>
<td></td>
</tr>
<tr>
<td>A x S</td>
<td>29</td>
<td>335.67</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>B x S</td>
<td>29</td>
<td>6398.03</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>C x S</td>
<td>29</td>
<td>1170.12</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>A x B x C</td>
<td>1</td>
<td>312.82</td>
<td>1.31</td>
<td>.2621</td>
<td></td>
</tr>
</tbody>
</table>
### TABLE 19

ANOVA RESULTS FOR G

<table>
<thead>
<tr>
<th>SOURCE</th>
<th>DF</th>
<th>MS</th>
<th>F RATIO</th>
<th>P&gt;F</th>
<th>MEANS</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>1</td>
<td>9.45</td>
<td>20.73</td>
<td>.0001</td>
<td>A1&gt;A2</td>
</tr>
<tr>
<td>B</td>
<td>1</td>
<td>182.16</td>
<td>39.16</td>
<td>.0001</td>
<td>B1&gt;B2</td>
</tr>
<tr>
<td>C</td>
<td>1</td>
<td>0.16</td>
<td>1.06</td>
<td>.3121</td>
<td></td>
</tr>
<tr>
<td>S</td>
<td>27</td>
<td>3.53</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>A x B</td>
<td>1</td>
<td>0.45</td>
<td>1.23</td>
<td>.2773</td>
<td></td>
</tr>
<tr>
<td>A x C</td>
<td>1</td>
<td>0.12</td>
<td>0.08</td>
<td>.7740</td>
<td></td>
</tr>
<tr>
<td>B x C</td>
<td>1</td>
<td>0.12</td>
<td>0.10</td>
<td>.7520</td>
<td></td>
</tr>
<tr>
<td>A x S</td>
<td>27</td>
<td>0.46</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>B x S</td>
<td>27</td>
<td>4.85</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>C x S</td>
<td>27</td>
<td>0.15</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>A x B x C</td>
<td>1</td>
<td>0.45</td>
<td>4.30</td>
<td>.0478</td>
<td></td>
</tr>
</tbody>
</table>

### TABLE 20

ANOVA RESULTS FOR H

<table>
<thead>
<tr>
<th>SOURCE</th>
<th>DF</th>
<th>MS</th>
<th>F RATIO</th>
<th>P&gt;F</th>
<th>MEANS</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>1</td>
<td>37.16</td>
<td>52.24</td>
<td>.0001</td>
<td>A1&gt;A2</td>
</tr>
<tr>
<td>B</td>
<td>1</td>
<td>285.31</td>
<td>79.11</td>
<td>.0001</td>
<td>B1&gt;B2</td>
</tr>
<tr>
<td>C</td>
<td>1</td>
<td>5.23</td>
<td>5.35</td>
<td>.0277</td>
<td>C1&gt;C2</td>
</tr>
<tr>
<td>S</td>
<td>30</td>
<td>9.79</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>A x B</td>
<td>1</td>
<td>7.81</td>
<td>14.03</td>
<td>.0008</td>
<td></td>
</tr>
<tr>
<td>A x C</td>
<td>1</td>
<td>1.95</td>
<td>4.32</td>
<td>.0463</td>
<td></td>
</tr>
<tr>
<td>B x C</td>
<td>1</td>
<td>0.06</td>
<td>0.08</td>
<td>.7758</td>
<td></td>
</tr>
<tr>
<td>A x S</td>
<td>30</td>
<td>0.71</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>B x S</td>
<td>30</td>
<td>3.61</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>C x S</td>
<td>30</td>
<td>0.98</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>A x B x C</td>
<td>1</td>
<td>0.40</td>
<td>0.60</td>
<td>.4439</td>
<td></td>
</tr>
</tbody>
</table>
Hypothesis 5

The fifth hypothesis concerns the effect of the experience independent variable:

\[ \text{H5: The senior's experience level has no affect on the subjects' responses.} \]

The evidence on this hypothesis is mixed. The hypothesis is rejected with a p-value of .0198 by the MANOVA. For the individual sample size ANOVAs, however, experience only accounts for a statistically significant amount of variance for D3 (accounts over $2,500). Both self report measures indicate that subjects put statistically significant more weight on the compliance test results and the sample size recommendations when the senior was more experienced. Thus, the subjects report greater reliance on the senior when he has extensive experience compared to little experience, but the evidence is mixed when it comes to converting this reliance into reduced sample sizes.

Hypothesis 6

The sixth hypothesis concerns the effect of the performance rating independent variable:

\[ \text{H6: The senior's performance rating has no affect on the subjects' responses.} \]

This hypothesis is rejected with a MANOVA p-value of .002. In addition, all the individual sample size judgments have a level of rejection below .04. Subjects reported placing greater weight on the compliance tests and the sample size recommendations when the senior had a higher performance rating. This result, together with the sample size judgments being smaller at the higher rating level, suggests greater reliance is exhibited by the subjects in the senior when he has
a high performance rating compared to a low rating.

**Hypothesis 7**

The seventh hypothesis concerns the effect of the sampling approach independent variable:

**H7**: The sampling approach adopted by the senior has no affect on the subjects' responses.

This hypothesis cannot be rejected. The subjects' sample size judgments did not change according to whether the senior used a statistical or judgmental approach to make his sample size recommendations. As was to be expected, the weight subjects reportedly placed on the compliance tests was not affected by the sampling approach for confirmations. While the subjects' sample size judgments were not affected, the p-value of .028 indicates that the weight subjects gave to the sample size recommendations was influenced to some extent by the sampling approach. Inspection of the treatment means indicates that more weight was given to recommendations based on statistical sampling. This finding is consistent with the earlier suggestion that the weight attached to the sample size recommendations would be the most sensitive of the dependent variables.

**Hypothesis 8**

The eighth hypothesis concerns the nature of the information provided by the senior, and whether the nature of the information interacts with source credibility:

**H8**: There is no interaction between the senior's sampling approach and the senior's experience level or performance rating on the subjects' responses.
This hypothesis cannot be rejected as no interaction effects on any of the sample size judgments are statistically significant. The subjects' self-report responses provide some support for this hypothesis. The weight subjects gave to the sample size recommendations when the senior was experienced with a high performance rating was the same under either sampling approach (5.42). When the senior was described as inexperienced with a poor rating, slightly more weight was given to his recommendation under statistical sampling (2.71 compared to 2.29). This interaction supports the nature of the information having an effect in the case of low source credibility; this effect, however, is not statistically significant.

A likely reason for this hypothesis not being rejected is that subjects saw no difference in the nature of the information presented. The judgment approach described entailed anchoring on last years sample sizes. This is believed to be a common and reasonable practice. Also, the statistical approach entailed some judgmental sampling. The view of no difference in the information presented is consistent with the failure to reject the seventh hypothesis.

Hypothesis 9 and 10

Hypothesis nine and ten concern the importance of the issue reported on by the source.

\[ H_9: \] The independent variables do not affect the subjects' sample size estimates for the relatively unimportant stratum of receivables.

\[ H_{10}: \] The independent variables do not affect the subjects' sample size estimates for the relatively important stratum of receivables.
Both hypotheses are rejected. The relatively unimportant stratum was presented as D1. While it was suggested that because of its unimportance the senior's recommendation would be accepted, the performance rating variable explained a statistically significant amount of variability in the dependent variable (see Table 15). The relatively important stratum was presented as D3. While it was suggested that because of its importance the subjects would discard the senior's recommendation and follow their own sampling policy, both experience and performance rating variables explained a statistically significant amount of variability in the dependent variable (see Table 17).

Yet there is some evidence suggesting that the importance of the issue, or nature of the accounts receivable strata, affected subjects' behavior. Tables 21 and 22 indicate that subjects' behavior differed substantially over strata. In particular, as indicated in Table 21, eighteen subjects did not change their sample size judgments in respect to D1 over levels of the independent variables. This behavior is in conformity with hypothesis nine. The changes in behavior of the other twelve subjects were sufficient to make performance rating statistically significant. Of these eighteen subjects, however, Table 22 indicates that the sample size judgments of nine of the subjects were above those recommended by the senior. Examination of Table 22 suggests that the senior's sample size recommendations were not generally accepted, and that there are apparently a variety of factors affecting subjects' sample size judgments. Additional analysis of subjects' behavior is now undertaken.
### TABLE 21
**BREAKDOWN OF SUBJECTS RESPONSES INTO CHANGE/NO CHANGE CATEGORIES**

<table>
<thead>
<tr>
<th>CATEGORY</th>
<th>D1</th>
<th>D2</th>
<th>D3</th>
<th>D4</th>
<th>G</th>
<th>H</th>
</tr>
</thead>
<tbody>
<tr>
<td>CHANGE</td>
<td>12</td>
<td>21</td>
<td>23</td>
<td>23</td>
<td>27</td>
<td>28</td>
</tr>
<tr>
<td>NO CHANGE</td>
<td>18</td>
<td>9</td>
<td>7</td>
<td>7</td>
<td>3</td>
<td>2</td>
</tr>
</tbody>
</table>

*To be classified as a "no-change," a subject must provide the same sample size to each of the alternate eight situations.

### TABLE 22
**FURTHER BREAKDOWN OF SUBJECTS RESPONSES**

<table>
<thead>
<tr>
<th>CATEGORY</th>
<th>D1</th>
<th>D2</th>
<th>D3</th>
<th>D4</th>
</tr>
</thead>
<tbody>
<tr>
<td>CHANGE-ABOVE</td>
<td>12</td>
<td>7</td>
<td>16</td>
<td>10</td>
</tr>
<tr>
<td>CHANGE-EQUAL</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>CHANGE-BELOW</td>
<td>-</td>
<td>14</td>
<td>6</td>
<td>13</td>
</tr>
<tr>
<td>NO CHANGE-ABOVE</td>
<td>9</td>
<td>2</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>NO CHANGE-EQUAL</td>
<td>9</td>
<td>4</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>NO CHANGE-BELOW</td>
<td>-</td>
<td>3</td>
<td>2</td>
<td>3</td>
</tr>
</tbody>
</table>

*Subjects are classified according to whether their average sample size judgments are above or below those recommended by the senior.*
Additional Analysis

The primary factors that are likely to enter an auditor's judgment process with respect to sample size judgments include: the nature of the client, last year's sample sizes and audit results, the results of this year's audit to date, especially the compliance test results, and the senior's sample size recommendations. These factors are included in the case description. Two additional considerations are the audit technology, particularly the sampling technology of the auditor's firm, and the auditor's own experience and preferred practices. These two considerations may influence the relative weights given to the audit specific factors.

Firm Effect

The firm factor may affect subjects' behavior. Firms are known to differ according to the nature of their audit technologies, for example, in the amount of judgmental versus statistical sampling undertaken, and in the relative emphasis given to attribute and dollar-unit sampling. Such considerations make it more likely to find a firm effect here than in stage one, where likelihood judgments about the state of the client's internal control system were elicited.

To test for a firm effect, subjects' mean responses were calculated and a one-way ANOVA run on each response category. The results are given in Table 23. A firm effect is evident. The significant result for the D4 stratum as well as the total sample size, is consistent with the differences between firms appearing in the main stratum, D4, and also this stratum being sufficiently large to reflect firm differences. Conceivably the significant result for D2 is due to firms
<table>
<thead>
<tr>
<th>DEPENDENT VARIABLE</th>
<th>DF</th>
<th>MS</th>
<th>F RATIO</th>
<th>P&gt;F</th>
</tr>
</thead>
<tbody>
<tr>
<td>D1</td>
<td>5,24</td>
<td>21.29</td>
<td>0.93</td>
<td>.4780</td>
</tr>
<tr>
<td>D2</td>
<td>5,24</td>
<td>226.80</td>
<td>3.17</td>
<td>.0245</td>
</tr>
<tr>
<td>D3</td>
<td>5,24</td>
<td>166.76</td>
<td>1.20</td>
<td>.3394</td>
</tr>
<tr>
<td>D4</td>
<td>5,24</td>
<td>11103.65</td>
<td>5.35</td>
<td>.0058</td>
</tr>
<tr>
<td>E</td>
<td>5,25</td>
<td>14969.81</td>
<td>4.12</td>
<td>.0072</td>
</tr>
<tr>
<td>G</td>
<td>5,22</td>
<td>0.61</td>
<td>0.87</td>
<td>.5185</td>
</tr>
<tr>
<td>H</td>
<td>5,25</td>
<td>2.60</td>
<td>2.74</td>
<td>.0419</td>
</tr>
</tbody>
</table>
having well defined procedures for dealing with accounts past due. The firm effect, however, is not significant for D1 and D2. Accounts past due over 90 days, D1, are distinguished from D2 by the suggested immateriality of the amount.

While accounting firms have to varying degrees written policies in respect to sampling procedures, auditors have considerable flexibility, particularly at the manager level, in conducting audit procedures. Evidently, this flexibility and related judgment differences, as well as the firm effect, is reflected in subjects' responses. Unfortunately, little information is publicly available about firms' sampling procedures and the differences between firms in these procedures. Such information could shed considerable light on the results discussed here.

**Subject Effect**

Subjects had two anchor points available in making their sample size judgments: last year's sample sizes and the senior's recommended sample sizes for this year. Probably last year's sample sizes would not be used again this year due to the improvements in the client's internal control system, which were included in the case. Table 22, which was presented earlier, gave a breakdown of subjects sample size judgments according to whether subjects' sample sizes were above or below those recommended by the senior. Inspection of the raw data and subsequent analysis suggested that this was an informative way of grouping subjects. Table 24 gives the composition based on total sample size of each group by firm. The presence of a firm effect, found to be statistically significant in the last section, is apparent. In Table 25, the subjects comprising the "above" and "below" groups for each
TABLE 24
SUBJECTS' SAMPLE SIZE JUDGMENTS COMPARED TO SENIOR'S RECOMMENDED SAMPLE SIZES BY FIRM

<table>
<thead>
<tr>
<th></th>
<th>F1</th>
<th>F2</th>
<th>F3</th>
<th>F4</th>
<th>F5</th>
<th>F6</th>
</tr>
</thead>
<tbody>
<tr>
<td>ABOVE*</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>5</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>EQUAL</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>BELOW</td>
<td>6</td>
<td>4</td>
<td>3</td>
<td>1</td>
<td>1</td>
<td>0</td>
</tr>
</tbody>
</table>

*classification based on total sample sizes

TABLE 25
INDIVIDUAL SUBJECT CLASSIFICATION OVER ACCOUNTS RECEIVABLE STRATA

<table>
<thead>
<tr>
<th></th>
<th>D2</th>
<th>D3</th>
<th>D4</th>
</tr>
</thead>
<tbody>
<tr>
<td>ABOVE*</td>
<td>5,6,12,14,15,16,18, 19,22,23,24,25,30, 31</td>
<td>5,6,8,9,12,13,14,15, 16,17,18,19,20,21,22, 23,24,26,27,28,29,30, 31</td>
<td>6,9,10,17,19,20,21, 22,24,26,27,28,29, 30,31</td>
</tr>
<tr>
<td>BELOW</td>
<td>1,2,3,4,5,8,*9,10, 11,13,17,20,21,26, 27,28,29</td>
<td>1,2,3,4,7,10,11,25</td>
<td>1,2,3,4,5,7,8,11, 12,13,14,15,16,18, 23,25</td>
</tr>
</tbody>
</table>

*Subjects whose sample sizes equalled those recommended by the senior are included in the "above" category.
# Table 26

## Summary of ANOVA on Subject Groupings

<table>
<thead>
<tr>
<th>Stratum</th>
<th>Source</th>
<th>DF</th>
<th>P&gt;F</th>
<th>Source</th>
<th>DF</th>
<th>P&gt;F</th>
</tr>
</thead>
<tbody>
<tr>
<td>D2</td>
<td>A</td>
<td>1,16</td>
<td>.7200</td>
<td>A</td>
<td>1,12</td>
<td>.0892</td>
</tr>
<tr>
<td></td>
<td>B</td>
<td>1,16</td>
<td>.3579</td>
<td>B</td>
<td>1,12</td>
<td>.0601</td>
</tr>
<tr>
<td></td>
<td>C</td>
<td>1,16</td>
<td>.6140</td>
<td>C</td>
<td>1,12</td>
<td>.0487</td>
</tr>
<tr>
<td>D3</td>
<td>A</td>
<td>1,7</td>
<td>.1114</td>
<td>A</td>
<td>1,21</td>
<td>.0246</td>
</tr>
<tr>
<td></td>
<td>B</td>
<td>1,7</td>
<td>.0727</td>
<td>B</td>
<td>1,21</td>
<td>.0006</td>
</tr>
<tr>
<td></td>
<td>C</td>
<td>1,7</td>
<td>.6527</td>
<td>C</td>
<td>1,21</td>
<td>.1340</td>
</tr>
<tr>
<td>D4</td>
<td>A</td>
<td>1,15</td>
<td>.8365</td>
<td>A</td>
<td>1,13</td>
<td>.5046</td>
</tr>
<tr>
<td></td>
<td>B</td>
<td>1,15</td>
<td>.2459</td>
<td>B</td>
<td>1,13</td>
<td>.0110</td>
</tr>
<tr>
<td></td>
<td>C</td>
<td>1,15</td>
<td>.2451</td>
<td>C</td>
<td>1,13</td>
<td>.4884</td>
</tr>
<tr>
<td>G*</td>
<td>A</td>
<td>1,12</td>
<td>.0096</td>
<td>A</td>
<td>1,14</td>
<td>.0031</td>
</tr>
<tr>
<td></td>
<td>B</td>
<td>1,12</td>
<td>.0005</td>
<td>B</td>
<td>1,14</td>
<td>.0010</td>
</tr>
<tr>
<td></td>
<td>C</td>
<td>1,12</td>
<td>.3665</td>
<td>C</td>
<td>1,14</td>
<td>.6337</td>
</tr>
<tr>
<td>H*</td>
<td>A</td>
<td>1,14</td>
<td>.0006</td>
<td>A</td>
<td>1,15</td>
<td>.0001</td>
</tr>
<tr>
<td></td>
<td>B</td>
<td>1,14</td>
<td>.0001</td>
<td>B</td>
<td>1,15</td>
<td>.0001</td>
</tr>
<tr>
<td></td>
<td>C</td>
<td>1,14</td>
<td>.0261</td>
<td>C</td>
<td>1,15</td>
<td>.4858</td>
</tr>
</tbody>
</table>

*Above/below classification based on total sample sizes*
stratum is provided. Subjects are not consistently classified into one or the other of the groups.

Separate ANOVAs for each group were run on each dependent variable. The results are summarized in Table 26. The interesting observation from the table is that no F ratio is significant at the .05 level for sample size judgments of subjects in the "below" group. In contrast, four of the nine F ratios are significant below the .05 level for the "above" group. Experience and performance rating are significant independent variables, for both groups, in explaining the variability in the weights reportedly given by subjects to the compliance test results and senior's sample size recommendations in making their own judgments. That subjects in the "below" group changed the weights reportedly given to the compliance test results and senior's sample size recommendations, but did not change their sample sizes, at least to a statistically significant level, is not necessarily inconsistent. As explained in the previous chapter, there are reasons to expect the dependent variables to exhibit different degrees of sensitivity.

Greater insight into the differences in behavior between the two groups is obtained from Figure 10. In Figure 10, the treatment means for performance rating are plotted for each dependent variable. Only the plots for this independent variable are presented because no additional insight was gained from the inspection of the other treatment means.

The graphs in Figure 10 show the large differences in mean sample sizes between the two groups (t tests on subjects' mean responses are significant at the .005 level for each stratum). The "below" group
FIGURE 10. PERFORMANCE RATING MEANS FOR EACH RESPONSE CATEGORY BY GROUP

*SSR: senior's sample size recommendation.
G: subjects' self-reported weight given to compliance test results.
H: subjects' self-reported weight given to SSR.
A: 'Above' group.
B: 'Below' group.
is characterized by: (1) sensitivity of sample size judgments to the performance rating independent variable, and (2) a large distance between their sample size judgments and the senior's recommended sample sizes. In contrast, the "above" group is characterized by: (1) sensitivity to the performance rating independent variable, and (2) an increase in their sample size judgments above the recommended sample sizes of the senior when his performance rating falls.

These findings are consistent with the self-report responses. Note that both groups placed more reliance on the compliance test results than on the senior's sample size recommendations ($t[27] = 5.07$, $p > .0005$). However, the "above" group gave more weight to the senior's sample size recommendations than did the "below" group when the senior had a high performance rating ($t[59] = 3.34$, $p > .005$). The "above" group's mean sample sizes approximated those of the senior when he received a high performance rating. With the low performance rating, the group's mean sample size increased due to the lower credibility attached to either, or both, the senior's sample size recommendations and the results of the compliance tests conducted by the senior.

The behavior of both groups is consistent with the psychological theory of social judgment and, more specifically, the assimilation-contrast theory (Anderson, 1971; Rhine and Severance, 1970; Sherif, Sherif and Nebergall, 1965). This theory provides an insightful way of explaining some of the observed behavior. The essence of the theory is that an individual has a region of assimilation around his initial beliefs and, beyond this region, are regions of contrast. It is predicted that a message within the region of assimilation will influence
the individual's beliefs, while a message in the region of contrast will have less, if any influence.

The sample sizes recommended by the senior were apparently in the region of assimilation for the subjects in the "above" group. The result was that when the senior had a high performance rating, his recommendations influenced these subjects. Then, when the senior had a low performance rating, his credibility fell, and his recommendations did not have the same influence; they were given less weight. The reduced influence of the senior's sample size recommendations combined with the added uncertainty surrounding the compliance test results led subjects to increase their sample sizes. In contrast, the senior's sample size recommendations were in the region of contrast for subjects in the "below" group. As a result, the senior's recommendations had little influence on the subjects' sample sizes and, therefore, did not provide an opportunity for source credibility to have a significant effect.

**Multidimensional Scaling**

Multidimensional scaling (MDS) is a set of mathematical techniques that provide a method for uncovering the 'hidden structure' or dimensions of data bases. An advantage of the techniques is that they can treat data as ordinal. For this reason, and also to provide another perspective to subjects' self-report responses, MDS was employed. The analysis of the weights subjects reportedly gave to the compliance test results is not reported on. The results are the same as the ANOVA results reported in Table 19. Basically, both experience level and performance rating are statistically significant in explaining the
variability in subjects' responses.

MDS did provide additional insight into subjects' responses on the weight they reportedly placed on the senior's sample size recommendations. As reported in Table 20, the ANOVA found experience, performance rating and sampling approach, together with the experience two-way interactions, statistically significant. The ALSCAL analysis resulted in performance rating making up the primary dimension and the sampling approach making up the less important second dimension. The goodness-of-fit measure, RSQ, which gives the proportion of total variance of the transformed data accounted for by the model, was a high .999. Figure 11 gives a plot of the results. The eight stimulus points represent the eight alternative descriptions of the senior plotted in the two-dimensional space.

The MDPREF technique is of interest because it provides a preference vector for each subject in the dimension space. It obtains scale values that maximize the correlation between the subject's original preference and the projection on the fitted vector. Consistent with the ALSCAL analysis, the two dimensions extracted are performance rating and sampling approach.

The results are plotted in Figure 12. Each point plotted in the two-dimensional space represents the identified subject's ideal point. Of interest is that some subjects indicated a preference for judgmental over statistical sampling. The majority of subjects appear to have no strong preference for either approach, or a preference for statistical sampling. The finding of this difference in preferences is consistent with the previous finding of a firm effect and the difference in
FIGURE 11. ALSCAL PLOT OF STIMULI ON TWO DIMENSIONS

INTERPRETATION:

- first letter is experience level - high (H) or low (L)
- second letter is performance rating - high (H) or low (L)
- third letter is sampling approach - statistical (S) or judgmental (J)
FIGURE 12. MDPREF PLOT OF SUBJECTS ON TWO DIMENSIONS

INTERPRETATION:
Horizontal dimension is performance rating—higher value represents preference for higher rating.
Vertical dimension is sampling approach—higher (lower) value represents preference for statistical (judgmental) sampling.
emphasis given to judgmental versus statistical sampling by firms. The subjects who indicated a preference for judgmental sampling came essentially from F2 and F5. However, other subjects from both these firms indicated a preference for statistical sampling. This finding also suggests that the importance of the sampling approach independent variable may have been underestimated in the previous analysis. This would have resulted from responses of subjects with different sampling preferences cancelling each other out.

**Subjects' Comments and the Post Experimental Questionnaire**

This section examines the third source of experimental data: the comments provided by subjects. Subjects had two opportunities to make comments about the experiment. First, subjects were given the opportunity to make oral comments to the researcher after the completion of the first-stage test instrument. Second, subjects could make written responses to the post experimental questionnaire attached to the second-stage test instrument. Some of these comments and responses relate directly to the conclusions and limitations of the research and are discussed further in the following concluding chapter.

**Subjects' Comments**

When the subjects had completed the first-stage test instrument, the researcher asked them if they had any problems completing their task and whether they wished to make any comments upon it. No problems were reported. Subjects' comments can be organized around three topics: (1) importance of subordinate credibility, (2) response to variations in credibility, and (3) information sources about a subordinate's
credibility. One subject related the three topics in the following way (paraphrased):

> With several clients going on at the same time, the real problem is where do you spend your time. The difficulty with evaluation is that the personnel department is too superficial, while the grapevine reports out of context.

The first topic is the importance of the source credibility issue in the audit context in general and, specifically, its relevance as captured by the experimental task. As indicated in Table 6, two subjects did not respond differently over levels of source credibility. Both of these subjects stated that their respective firms did not employ unreliable persons. One of them said that he had not personally experienced anyone in the firm on whom he could not rely. The other subject added that if the person conducting the tests had been promoted to senior he could reliably conduct them. Several other subjects commented that the importance of source credibility depended on the particular audit context. They said that they could envisage contexts, other than the one described in the case study, in which source credibility would be of more importance. One example given was the possible existence of management fraud. In general, the ease with which subjects discussed the question of reliance on subordinates suggested that they were familiar and concerned with the issue of source credibility. In addition, two subjects said that they had recently been in the same situation as that described in the case study; that is, taking over an audit already underway and being unfamiliar with both the client and the audit team members who had completed the work to date.
The second topic is the appropriate mode of response to variations in source credibility. The subject paraphrased above was concerned with the most efficient allocation of his time among clients. A determinant of this allocation is the degree of reliance the manager can place on his subordinates. Several subjects said that the normal response to doubt about a subordinate's competence and reliability was at the "front end." That is, the managers would spend more time supervising subordinates, including time at the client's premises, when this doubt existed. This form of response was not captured in the experiment.

The third topic concerns how a manager obtains information about the credibility of his subordinates. Subjects said that the primary source of information was their own personal experience with the particular subordinate. An important source is also the informal communication system within the firm. Several subjects said that it was not uncommon for a senior's reputation to be known through the grapevine before they had personal experience with the senior. One subject said that within his firm, he believed informal communications were the primary source of information concerning a senior's competence. Finally, all firms had some form of formalized evaluation system. The result of such evaluation was often a classification such as "acceptable" and "unacceptable" or "excellent," "good" and "average."

Post Experimental Questionnaire

The last page of the second-stage test instrument consists of a three-question questionnaire (see Appendix D). The first question asked subjects if the task was understandable. Four of the 31 subjects
who returned the test instrument did not respond to the question, the other 27 subjects answered "yes." A variety of complementary comments concerning the clarity of the case study were also included.

The second question asked whether the subjects believed that there were any additional items of information that should have been included in the case study. Five subjects gave no response, 21 subjects said the information provided was sufficient, and 5 subjects made some comment. Two of these subjects said that they would have liked to know whether the parameters used by the senior in making his statistical sample size recommendation were acceptable. A third subject said that, before the previous manager left, he should have reviewed all of the work completed to date and provided the subject with his opinion. A fourth subject said that he would have liked more information on the client and the senior's evaluation. The fifth subject also wanted more information on the senior's evaluation, including examples of why the senior had a particular performance rating, the measures used in performance rating, and whether he passed the CPA exam in the first attempt.

The third question asked if subjects had any other comments about the task. Five subjects gave no response, 8 subjects responded that they had no additional comments, and 18 subjects provided comments. A number of these comments supported the verbal comments given at the end of the first-stage experiment. For example, three subjects repeated that the senior, by the very fact of being a senior, should have been capable of the work required of him in the case study. One subject said that his first response would have been at the "front end," rather
than altering sample sizes. Along the same lines, another subject said that in his experience there was little correlation between sample sizes and the evaluation of the senior. Three subjects said that with respect to the senior's statistical sample size recommendations, they would have liked to have known the parameters utilized.

The other comments related to the approach to the sampling issue. One subject said that on an actual audit he would have used dollar unit sampling. In addition, the use of a stratified sample depended on the materiality of the strata, about which he felt there was insufficient information to make a judgment. Another subject said that an accounting cycle approach to confirmations was appropriate; this would involve taking into account other information such as the condition of cash receipts and various subsidiary ledgers. One subject said that in the situation described he would never use judgmental sampling. In contrast, another subject said that his firm used only judgmental sampling in the situation described so that he found evaluating the statistical sample size recommendations difficult. Finally, several subjects said that the sample sizes employed last year and those recommended by the senior were much larger than those usually used by their firm. These comments indicate the presence of differences between firms and support the finding of a firm effect in the statistical analysis.

The number and nature of comments made are believed to support the researcher's perception that subjects completed the test instruments conscientiously and were interested in the subject-matter. That a number of subjects believed source credibility, in the context presented, did not have as significant an impact on audit risk as in other
possible contexts is interpreted to imply that the context used provided a strong test of source credibility. Subjects' comments can also be viewed as pointing out weaknesses in the experimental design. Probably the major such weakness was not allowing what several subjects referred to as a "front end" response. In general, however, the subjects seemed comfortable with the limitations in the experiment, some of which are inherent in any experiment. Subjects' comments also support the previous finding of firm differences influencing subjects' responses. The existence of both firm and subject differences makes it important in the design of an experiment that a careful compromise is made between, on the one hand, realism and specific detail so that subjects can relate to the task and, on the other hand, abstraction from reality and generality so that firm and individual subject differences do not create problems in the interpretation of the test. The researcher believes the comments provided suggest an acceptable compromise was achieved in both test instruments.

Chapter Summary

This chapter presented the analysis of the data obtained from both stages of the experiment. The results enabled the testing of the hypotheses relating to the final three research objectives. The data was primarily analyzed using an ANOVA approach with the associated F ratios providing the hypothesis tests. Additional analysis provided insight into firm and individual subject behavior in response to the independent variables manipulated.
The next chapter includes a summary of the experimental results, and the conclusions and implications which can be derived from the results.
CHAPTER VI

SUMMARY AND CONCLUSIONS

This chapter summarizes the research undertaken, provides the interpretation of the conclusions and implications of the results, discusses the limitations of the study, and makes suggestions for future research.

The chapter consists of five sections. The first section summarizes the research as described in the first five chapters. This summary provides the basis for drawing, in the second section, conclusions from the results. Implications with respect to auditor judgment research, auditing practice, and experimental psychology are discussed in the third section. The fourth section describes possible limitations to the study. The implications of these limitations for the validity of the study's results and the researcher's opinion on their significance are provided. Finally, in section five, suggestions for future research are provided.

Summary

A synopsis of the study follows: Judgment under uncertainty is a primary ingredient of the audit process. The audit of a client's financial statements is conducted by an audit team; the division of labor
occurring within the audit team means that, in making a judgment, an
auditor has to rely on the work and judgments of other team members.
As is to be expected, auditors make fallible judgments and previous
research has noted specific judgment differences among auditors. It
was hypothesized that an auditor who relies on another auditor as an
information source will be concerned with this other auditor's credi-
bility. In other contexts, source credibility has been found to be a
significant factor entering into an individual's judgment process. A
Bayesian model of the auditor's judgment process, which included a
measure of source credibility, was formulated; it indicated that source
credibility can have a significant impact on the inferential value of
audit evidence. A two-stage experiment was conducted to consider to
what extent audit managers were sensitive to the credibility of the
audit senior, and what factors influence the managers' appraisal of the
senior's credibility. Subjects were found to be sensitive to source
credibility in terms of both their judgments about the state of the
client's internal control system and their judgments about sample sizes
for confirmation of accounts receivable. All the senior-related fac-
tors manipulated had some affect on subjects' behavior.

In Chapter II, descriptions of the audit process were considered.
The dependencies between auditors' judgments that arose because an
audit is conducted by an audit team were recognized. Previous research
on auditor judgment has not explicitly considered that audit judgments
are made in the context of the audit team. A review of this research,
however, did show that, within the limits of the experiments undertaken,
significant differences between individual auditors' judgments were
found. The existence of these judgment differences, in the light of
the judgment dependencies existing within the audit team, suggests
that an auditor should carefully evaluate a subordinate when relying on
the subordinate's work.

The auditor's evaluation of a subordinate as an information source
fits the source credibility scenario investigated in experimental
psychology. A review of this literature indicated that in a variety
of situations, source credibility, when appropriately described,
affected subjects' judgments. The use of the source credibility con­
cept provided several advantages in addition to an existing body of
research on which to draw. One advantage was that the source credi­
bility approach provided a means of concentrating on the user of the
information, the audit manager in this case, and, by using hypothetical
descriptions of the senior, avoided the need to use actual audit teams.
Another advantage was that source credibility could be represented in a
probabilistic context and was subject to modelling. The model developed
provided both a normative analysis of source credibility in an auditing
situation and a convenient benchmark for analyzing the experimental
data.

The overall purpose of the study was to examine the judgmental
dependencies existing in the audit team; in particular, to determine
whether source credibility was a factor which entered into the audi­
tor's judgment and, if so, what effect it had. To achieve this overall
purpose, four research objectives were established. The first research
objective was to model the auditor's judgment process from an audit
team perspective incorporating a measure of source credibility. This
In Chapter III, the probabilistic nature of auditing was recognized and shown to be conducive to a Bayesian representation. A Bayesian model of auditor judgment at each level in the audit team was presented. For the audit manager level, the probabilistic measure of source credibility developed by Schum and DuCharme (op. cit.) was explicitly incorporated into the Bayesian model. The model represented the audit manager's judgment as a function of the relevant audit evidence, and the credibility of the audit senior who was the source of the evidence. The application of the model to an auditing example suggested that even small reductions in source credibility can have a large impact on the inferential value of audit procedures. The literature review in Chapter II presented the psychological finding that, compared to the model's predictions, subjects are relatively insensitive to the impact of source credibility.

In Chapter IV, the research methodology for considering the second, third, and fourth research objectives was described. A laboratory experiment was justified primarily on the grounds of practical considerations, but also with respect to internal validity. Steps to enhance external validity were also described. The experiment consisted of two stages. The first stage considered the second research objective: to test to what degree the auditor's judgment is sensitive to source credibility. The third and fourth hypotheses, to test whether the auditor modifies the audit program to compensate for variations in source credibility and to evaluate what factors about a subordinate influence the auditor's appraisal of the subordinate's credibility, were considered in the second stage. The two-stage
experiment permitted, in the first stage, the explicit presentation of source credibility in a probabilistic form which, in turn, allowed the model to provide benchmark responses and, in the second stage, an implicit presentation of source credibility. The use of the two approaches adds to the reliability and validity of the results.

The choice of audit managers as subjects was based on their position in the audit team hierarchy, having regard to the purpose of the study, and the fact that most of the existing auditor judgment research has used audit assistants and seniors as subjects. The scarcity of audit managers available to act as subjects was one reason for employing a repeated measures design in both stages of the experiment. Another reason was that the primary interest was in how subjects responded over different levels of source credibility; a repeated measures design can capture this change in behavior. Finally, in Chapter IV the specific hypotheses relating to the research objectives were enumerated and their testing, primarily by the F ratios associated with the appropriate ANOVA models, was described and justified.

Chapter V presented the analysis of the data obtained from both stages of the experiment. The results of the hypothesis tests are summarized in Table 27. As well as analysis to specifically test the research hypotheses, additional analysis was undertaken. This additional analysis included investigation at the firm level and, through clustering, at the individual subject level. The results of this analysis are not summarized here. Rather, they are combined with the results of the hypothesis tests in the next two sections, where the researcher's interpretation of the conclusions and their implications
### TABLE 27
**SUMMARY OF HYPOTHESIS TESTS**

<table>
<thead>
<tr>
<th>HYPOTHESIS</th>
<th>RESULT</th>
<th>MODEL AND SIGNIFICANCE LEVEL</th>
</tr>
</thead>
<tbody>
<tr>
<td>H1: The subjects' subjective probability that the client's system has functioned sufficiently well to justify substantial reliance on it is unaffected by the level of source credibility.</td>
<td>Rejected</td>
<td>ANOVA, .0001</td>
</tr>
<tr>
<td>H2: The subjects' subjective probability that the client's system has functioned sufficiently well to justify substantial reliance on it is unaffected by the confidence level of the compliance tests.</td>
<td>Rejected</td>
<td>ANOVA, .0001</td>
</tr>
<tr>
<td>H3: The impact of source credibility is unaffected by the confidence level of the compliance tests.</td>
<td>Not rejected</td>
<td>ANOVA, .9374</td>
</tr>
<tr>
<td>H4: The derived likelihood ratios are approximated by the likelihood ratios calculated from the model.</td>
<td>Rejected</td>
<td>Difference ANOVA, .0001</td>
</tr>
<tr>
<td>H5: The subjects' responses are unaffected by the senior's experience level.</td>
<td>Rejected</td>
<td>MANOVA, .0198</td>
</tr>
<tr>
<td>H6: The subjects' responses are unaffected by the senior's performance rating.</td>
<td>Rejected</td>
<td>MANOVA, .0019</td>
</tr>
<tr>
<td>H7: The subjects' responses are unaffected by the sampling approach adopted by the senior.</td>
<td>Not rejected</td>
<td>MANOVA, .3391</td>
</tr>
<tr>
<td>HYPOTHESIS</td>
<td>RESULT</td>
<td>MODEL AND SIGNIFICANCE LEVEL</td>
</tr>
<tr>
<td>----------------------------------------------------------------------------</td>
<td>--------------</td>
<td>------------------------------</td>
</tr>
<tr>
<td>H8: There is no interaction between the senior's sampling approach and the</td>
<td>Not rejected</td>
<td>MANOVA, .5233</td>
</tr>
<tr>
<td>senior's experience level or performance rating on the subjects' responses.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>H9: The independent variables do not affect the subjects' sample size</td>
<td>Rejected</td>
<td>D1 ANOVA, .01</td>
</tr>
<tr>
<td>estimates for the relatively unimportant stratum of receivables.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>H10: The independent variables do not affect the subjects' sample size</td>
<td>Rejected</td>
<td>D3 ANOVA, .01</td>
</tr>
<tr>
<td>estimates for the relatively important stratum of receivables.</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Conclusions

In a strict sense, the following conclusions apply only to the population sampled: the audit managers of six of the "Big Eight" firms located in Columbus, Ohio. The geographical region, however, is not considered to pose a serious restriction to the conclusions. The issue of reliance on subordinates' work is inherent in the team approach to the audit. Furthermore, the "Big Eight" firms are each believed to have somewhat standardized training programs and audit technologies across their own offices. Office size may affect the generality of the conclusions. It is to be expected that auditors would have extensive personal experience with their fellow auditors in small offices, while more impersonal sources of information about other auditors would be more important in large offices. The offices sampled were medium size offices. Finally, as only "Big Eight" firms were sampled, no generalizations can be made to smaller national or regional firms. The conclusions are organized according to the three research objectives which motivated the experiment, and the additional topic of differences across subjects.

The Second Research Objective

The second research objective was to test to what degree auditors' judgments are sensitive to source credibility. Both stages of the experiment found, as evidenced by the rejection of H1, H4 and H5, that the audit managers were sensitive to the senior's credibility. In stage one, subjects received a message in the form of compliance test
results and the senior's recommendation based on these results. The message was that the client's internal control system has functioned sufficiently well to justify substantial reliance on it. That is, it was strong. Subjects' subjective probability that the client's system was strong declined as the credibility of the senior declined. This result indicates that subjects' reliance on the senior's work declined as his credibility declined. Subjects' discounting for the decline in credibility was essentially linear although, especially for some subjects, there was a tendency for the rate of discounting to increase at the lower levels of source credibility.

The greater discounting at lower levels of source credibility suggests that subjects may not have sufficiently discounted for the initial reductions in source credibility, since, normatively, its greatest impact occurs at the highest credibility levels. Comparing subjects' responses with the model's predictions indicated that for the majority of subjects, however, the discounting was at least equal to that of the model and generally greater at lower levels of source credibility. The greater discounting for source credibility at its lower levels, for example 70 percent, may in part be a reflection of subjects being unfamiliar with such unreliable subordinates. In the methodology chapter, it was suggested that such low levels of source credibility would be unlikely in auditing.

The Third Research Objective

The third research objective also concerns subjects' sensitivity to source credibility. This objective was to test whether auditors modify the audit program to compensate for variations in source
credibility. Modifications to the audit program were considered through subjects' sample size judgments. In general, subjects did alter their sample size judgments in response to different descriptions of the senior, as evidenced by the rejection of hypotheses H5 and H6. Significant differences, however, were found across subjects and over accounts receivable strata. The subjects who were sensitive to source credibility, in the form of the various descriptions of the senior, approximated the senior's sample size recommendations in their own sample size judgments when his credibility was high, and increased their sample sizes when his credibility declined.

The findings under research objectives two and three mutually reinforced the conclusion that the audit managers were sensitive to the credibility of a subordinate auditor as an information source. This conclusion holds over different experimental tasks and different forms of subordinate descriptions. This conclusion is also supported by the comments of subjects and their responses to the two self-report questions. In stage two, subjects were asked to indicate the weight they attached to the compliance test results and the senior's sample size recommendations in making their sample size judgments. As the senior's credibility declined, subjects gave less weight to both the compliance test results and the senior's sample size recommendations.

The Fourth Research Objective

The fourth research objective was to evaluate what factors about a subordinate influence the auditors' appraisal of the subordinate's credibility. The relevant psychology literature primarily focuses on
two factors as significantly affecting a person's evaluation: experience and performance rating. The self-report responses indicated that subjects placed more reliance on the senior when he had the greater experience and higher performance rating. In terms of the sample size judgments, more reliance on the senior resulted in lower sample sizes and, for one group of subjects, greater proximity to the senior's recommendations.

Experience was the less significant of the two factors. Understandably experience did not have the impact of performance rating for, as one subject commented, presumably the more experienced auditor will make a professional judgment, but not necessarily. In respect to other factors influencing the managers' evaluation of subordinates, both the informal communications system's rating and personal experience were referred to in subjects' comments. Neither of these factors, nor sex and race, which have been extensively studied in performance rating literature, were allowed for in the experiment.

In the psychology literature dealing with source credibility, the interaction of other variables with source credibility has been considered. An attempt was made in the second stage to see if the impact of source credibility was affected by its interaction with other variables, in particular, the type of information presented, by hypothesis H8, and the importance of the issue, by hypotheses H9 and H10. Nothing definitive could be concluded from the results. Regarding the type of information, any effect may have been cancelled out to some extent by subjects having differing preferences between sampling approaches. However, subjects may have considered that there was no relevant difference
in the type of information based on the different sampling approaches. The possible interaction between source credibility and the importance of the issue to the subject is also clouded by the question of how subjects interpreted the relevant information. Subjects responded differently to the manipulation of the independent variables over the various accounts receivable strata, but subjects' responses suggest that subjects did not view the relative importance of the strata as originally hypothesized by the researcher.

The Subject Effect

Differences across subjects' judgments were found in both stages of the experiment. To put these differences into perspective, the average correlation coefficient between subjects' responses in the first stage was .7. As a measure of consensus, this is a reasonably high value. The majority of subjects were described as conservative in a Bayesian sense, exhibiting somewhat excessive discounting for source credibility. In the second stage, the firm factor appeared to account for a large part of the differences in sample size judgments. Subjects were described, in a general sense, as responding according to whether the senior's sample size recommendations were within their regions of acceptance or rejection.

Several interesting differences did exist between subjects. Subjects were approximately equally divided on whether they saw a difference between conducting compliance tests at 90 percent and 95 percent confidence levels. The classification of some subjects as conservative and others, a minority, as counter-conservative indicates that differences existed in the relative weights given to prior beliefs and
additional information in the form of compliance test results. Subjects also differed in their relative preferences for statistical and judgmental sampling. Finally, subjects differed in how they responded over the various accounts receivable strata.

No definitive explanation can be given for these differences. However, three factors probably contributed to the differences. One factor is differences in subjects' cognitive processes. The conservative and counter-conservative form of Bayesian revision is a possible example of such a difference. The existence of differences in cognitive processes even among expert judges is well documented in psychology. A second factor is differences between subjects in audit practices and policies. An example of this type of difference is the preference for statistical versus judgmental sampling. While the firm factor likely explains a large part of these differences, the results suggest it does not provide a complete explanation. At the managerial level in the audit team, presumably auditors have some flexibility to follow their personal preferences based on their own experience. The third factor is differences between subjects induced by the experiment. The necessary abstraction from reality in the case study could cause subjects to interpret aspects of the case study differently. Thus, some of the differences between subjects might not arise in an actual audit. Unfortunately, the relative importance of the three factors can not be measured.
Implications

The implications of the above conclusions are now discussed under three headings: auditor judgment research, auditing practice, and experimental psychology.

Auditor Judgment Research

The study's implications with respect to existing auditor judgment research should be considered since this research motivated the study. One implication concerns the quality of auditors' judgment. Simply stated, the telling characteristic of expert judgment is that, given the same set of data, expert judges will come to the same conclusions about the data. The nature of expert judgment was described in more detail in Chapter II. Existing auditor judgment research, for example Weber (op. cit.), has not always found that auditor-subjects make the responses expected from expert judges. Here, an argument for a certain behavior was made based on descriptions of the audit process, a normative model, and experimental findings from psychology. The audit managers responded in the appropriate direction with reasonably high consensus.

A related implication concerns the nature of the judgment differences among auditors found in previous studies. These differences have led to the auditors' "expertness" being questioned. In an attempt to recognize the complexity of audit judgments, this study made a descriptive analysis of the audit process, collected data from a two-stage experiment, and provided subjects with an opportunity to make additional comments. An implication from this comprehensive approach is that care is required in evaluating auditors' judgments. Differences
between auditors' judgments were shown to be explained by a firm effect as well as a subject effect. In respect to the subject effect, it was suggested that to some extent the differences were artificial in that they resulted from different interpretations of the initial information provided in the case study. More importantly, there is an inherent flexibility in the audit process in that a different effort in part of the process can be offset by a different effort in a subsequent part. Thus, comparing auditors' judgments on one part of the audit process in isolation may be misleading. Interpreting the nature of auditor judgment differences is further complicated by the experimental design required. A meaningful experiment on auditor judgment requires the use of a representative task so that the expert judges acting as subjects have an opportunity to apply their expertise. However, a consequence of this design is that the subjects can utilize their variety of experiences and preferences in responding to the task. This adds to the potential for observing differences.

The third and most important implication of the conclusions for auditor judgment research concerns the potential impact of auditors' judgment differences. A common response of researchers on finding variability among auditors' judgments is to question the impact of the judgment differences on the audit process. The audit review process is one form of quality control of auditors' work and judgments. Furthermore, the extent of the judgment dependencies existing in the audit team means that a judgment made by one auditor will often be an input into another auditor's judgment. Therefore, there is a continual review process occurring, provided the auditor does not automatically accept the
subordinate's work as accurate. The conclusion of the experiment is that the audit managers sampled do not automatically accept the subordinate's work as accurate.

The audit process provides many opportunities for the auditor to modify his own judgment and the audit process for doubt over another auditor's reliability. The subjects' comments indicated that one form of this adjustment occurs at the "front end" in terms of increased supervision. The experiment indicated that the audit program will also be modified. Since the ability of the subjects to evaluate their subordinates so that an appropriate adjustment is made was not investigated, that the adjustment always takes place was not established. But, the study's conclusions suggest that to the extent that an accurate evaluation is made, an audit manager will recognize the decline in the informativeness of the work associated with a less than perfectly credible senior, and act appropriately.

This implication is in contrast to a secondary implication of the study by Mock and Turner (op. cit.). As the only other study to consider the impact of the audit team on auditor judgment, they found that the review by the audit manager in actual two-member audit teams did not reduce the variability in the audit senior's judgments. As previously discussed, there are a variety of possible explanations for their finding. They acknowledge that the audit managers may not have had sufficient information to make an evaluation. In this study, in contrast, the evaluation was made for the manager. Also, subject motivation may be questioned in their study.
The final implication related to existing research concerns the relationship between compliance and substantive tests. At issue is whether substantive tests are modified to reflect the internal control strength of the client. Mock and Turner (op. cit.) found that sample size recommendations for substantive tests by subjects were smaller when internal control systems were stronger. While this issue was not specifically addressed in this study, the conclusions do support the existence of a relationship. The subjects increased their sample sizes when the state of the internal control system was more uncertain because of doubt about the reliability of the audit senior conducting the evaluation of the system.

**Auditing Practice**

While the above implications for auditor judgment research are clearly relevant to auditing practice, the second set of implications directly relate to auditing practice. The first implication is that accounting firms should be reassured by the conclusions of the study. In the most general sense, the audit objective is to keep audit risk within an acceptable level; here audit risk is the possibility that a material misstatement in the client's financial statements will escape detection. Source credibility, as a representation of the ability of an auditor to conduct an audit procedure and accurately report the results, is an important element of audit risk. Managers were not only aware of this element of audit risk, but responded in the appropriate direction to its manipulation.

In regard to the comparison of subjects' responses with the predictions of the model, subjects generally responded by somewhat
excessive discounting for source credibility and by conservative Bayesian revision of their priors. That is, subjects attached less inferential value to the information than did the model. A possible consequence of this behavior is overauditing. In underestimating the informativeness of the results of an audit procedure, more work than necessary may be undertaken by auditors in order to satisfy their minimum levels of assurance.

Subjects' conservative behavior can be given an interesting alternative interpretation. In statistical testing, risk is assessed in terms of type I (α) and type II (β) error rates. In auditing the type II error rate is the most important. The observed conservatism can be argued to result from the auditors' sensitivity to β and their desire not to underestimate it. For example, consider the internal control example from Chapter III. If the error rates are substituted for the basic probabilities, the state-evidence matrix is:

<table>
<thead>
<tr>
<th></th>
<th>$H_1$</th>
<th>$H_2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$D_1$</td>
<td>$1-\alpha$</td>
<td>$\beta$</td>
</tr>
<tr>
<td>$D_2$</td>
<td>$\alpha$</td>
<td>$1-\beta$</td>
</tr>
</tbody>
</table>

Given: $A_{12} = \frac{P(D_1 | H_1)}{P(D_1 | H_2)} = \frac{1-\alpha}{\beta}$

The empirical result $A_{12}(S) < A_{12}(M)$ implies $\frac{1-\alpha}{\beta}(S) < \frac{1-\alpha}{\beta}(M)$

Assuming subjects' subjective α equals the α at which the tests were conducted then subjects' subjective β is greater than the actual one.
Thus, what was initially called overauditing may be a cost effective level of auditing once the costs associated with the components of audit risk are considered.

The conclusions of the study also have implications for the training of auditors and for communications within the accounting firm. Auditor training should have the effect of increasing auditors' source credibility. Even at initially high levels of source credibility, the model indicated that subsequent increases in source credibility have a significant affect on the inferential value of information. This result supports the investment in training programs firms make. Subjects' comments indicated that primary sources of information about subordinates are the managers' personal experience with the subordinate and the informal communication system's evaluation of the subordinate. One aspect of auditor training should be instruction in how to evaluate subordinates. The mixed results for the influence of experience together with subjects' comments suggest that, as to be expected, performance entails more than simply experience.

A consequence of the division of labor within the audit team is the importance of communication. Communication is important across as well as within the audit team so that an auditor is familiar with those he works with. Subjects commented about their reliance on the informal communication system. Accounting firms need to consider whether this reliance should be encouraged or discouraged, and whether it is an indication that the formalized performance rating system is inadequate. The formalized performance rating of the senior was statistically significant across all dependent variables. For this rating to be
informative in practice, it must be sufficiently inclusive.

A final implication concerns the significant firm factor found in respect to subjects' sample size judgments. Given the pressures on accounting firms to find increasingly cost efficient audit procedures, the differences between firms suggest that, for some firms at least, improvements may be available. At the very least, the differences between firms are worthy of investigation.

**Experimental Psychology**

The study's conclusions also have implications for experimental psychology. The study provides an application of cascaded inference theory, which has undergone considerable theoretical development, but little experimental application. This application involved a realistic task setting and expert judges as subjects. In psychology there is a call for greater realism in experimental designs (Petrinovich, *op. cit.*). Of interest is that the subjects in the study excessively discounted for source credibility, while in the existing psychology studies the discounting was inadequate. No specific conclusion can be drawn about the conflicting findings because of the different experimental designs employed. However, the use of actual auditors in a realistic task setting may be the reason. The finding of sensitivity to source credibility and the results suggesting Bayesian conservatism are consistent with the relevant psychology research.

**Limitations**

This section discusses limitations of the study which may confound or invalidate the conclusions and their implications. Limitations with
respects to the experiment as a whole are discussed first, followed by limitations specifically relating to each stage of the experiment.

Three limitations relate to the experiment as a whole. The first is an internal validity issue. It is the possible "demand effects" induced by the repeated measures design. In Chapter IV the reasons for the use of a repeated measures design were given, together with the steps taken to guard against serious "demand effects" occurring. The experimental results suggest that "demand effects" were not significant. In the first stage, 15 subjects did not change their responses over confidence levels. This result suggests manipulation of this independent variable did not lead to any "demand effects," and supports the contention that subjects' responses to the manipulation of source credibility were not simply experimentally induced. Similarly, in the second stage subjects discriminated between accounts receivable strata in responding to manipulations of the independent variables.

The other two limitations concern external validity. The first is the restriction placed on subjects' mode of response to the uncertainty introduced by the senior's credibility. The experiment did not allow for a "front end" response in the form of increased supervision. Use of managers' estimate of budgeted hours was originally considered as a dependent variable. It was not pursued because estimating budgeted hours is difficult even at the best of times; it was believed impractical to present sufficient information in the case study to enable subjects to provide an estimate with which they would be comfortable. Note that recognition of this alternative response does not make the evaluation of the subordinate's credibility any less important. Only the timing of the response is affected. The manager would have to
observe, or in effect do, all the work performed by the subordinate to avoid all reliance on the subordinate's work.

The nature of the experiment also made it impossible for the manager to consult with the senior on any aspect of the senior's work the manager felt was critical. Again, however, neither the evaluation of, nor importance of, source credibility is eliminated by this ex post alternative. The documentation in auditing is insufficient for managers to be always able to consult the working papers to satisfy a doubt. In addition, as was referred to in Chapter II, face to face consultation is not always satisfactory because of a person's generally limited ability to explain the rationale for his judgment.

The second limitation relating to the external validity of the experiment is that the subjects' ability to evaluate the senior's credibility was not considered. This limitation is somewhat mitigated by the existence of formalized performance rating procedures in firms. Their usefulness, however, was questioned. The variety of factors mentioned by subjects as having influenced their evaluation of a subordinate and the finding that this evaluation was not simply tied to experience do suggest that subjects make more than a superficial evaluation of subordinates.

Two potential limitations specifically dealing with the first stage of the experiment require comment. The first limitation is the use of a probabilistic framework. While the nature of the audit process makes it conducive to a probabilistic representation, to what extent auditors actually think in terms of probabilities is unknown. It is unlikely that auditors think about the reliability of their
subordinates in terms of probabilities. The significance of this limitation is unknown. All that can be said is that subjects appeared comfortable completing the test instrument, and none of them commented about this aspect of it. The second limitation is the assumptions that were required to conduct the difference analysis. These assumptions were referred to in Chapter IV and discussed in detail in Appendix C. The limited sensitivity analysis conducted suggests that the experimental results are robust to violations of these assumptions.

Several potential limitations are associated with the second stage of the experiment. One is the possibility that subjects' responses to the second-stage test instrument were influenced by the administration of the first-stage test instrument. One way to have controlled for this possibility would have been to have some subjects complete stage two first. Practical considerations pursuaded the researcher not to do this. The carry-over effect was hopefully minimized by the two stages focusing on different aspects of the audit process, and by having subjects complete the second stage at a later date. The receipt of the second-stage test instrument was approximately a week or more after administration of the first stage. The impact of any carry-over effect would probably be to make subjects more sensitive to the independent variables manipulated in the second stage.

The other limitations concern the extent of the data analysis conducted. One limitation is the inability to individually compare each subject's responses to the two test instruments. This inability arose because of lack of foresight by the researcher; second-stage test instruments were identified only by firm and not by subject. The other
limitations are beyond the control of the researcher. Additional analysis is clearly warranted on the nature of the firm effect and the different effect of the various accounts receivable strata. As previously explained, the lack of publicly available information on firm specific audit technologies makes this analysis impossible at the present time.

In summary, the limitations discussed in this section generally arose from the process of adapting a complex auditing situation to a laboratory experiment. The researcher believes the essence of this situation was captured by the experiment and that, even taken together, the limitations do not threaten the main conclusions of the study.

Suggestions For Future Research

Since the study represents a first attempt at investigating the context effect of the audit team on auditors' judgments, replications of the experiment are worthwhile. The study by Mock and Turner (op. cit.) is the only other study to shed some light on this issue. Furthermore, the conflict in the findings of this study with certain findings of Mock and Turner and certain experimental psychology findings makes replications of interest. Replications could strengthen the study's conclusions by using different levels of the independent variables, by including a dependent variable that measures the "front end" response and, possibly, by employing a between subject design to eliminate any lingering doubts about "demand effects."

Other extensions relate to the areas of limited analysis in the study. The investigation of the nature of the firm effect is an obvious extension. What is required to conduct this investigation is
information on the specific practices of accounting firms, for example, the extent of different types of sampling conducted, the firm guidelines available for deciding what type of sampling to employ under what circumstances, and the firm guidelines or firm practices in respect to the critical values to utilize in sampling procedures. The researcher is currently in the process of attempting to secure some of this information which, as was previously explained, is not publicly available.

The ability of auditors to evaluate their subordinates also requires investigation. Possibly, real life audit teams could be used and members' evaluation of their subordinates compared to some "normative rating" conducted by the researcher. Alternatively, various descriptions of auditors could be presented to subjects, in a more detailed form than was done here, to determine which factors were relied upon by subjects, and what was the consensus between subjects.

Two analytical extensions are of interest. One is the extension of the model to include preposterior Bayesian analysis. Such an extension would enable the consideration of the effect of varying source credibility on the expected utility of the audit manager and possibly accounting firm. In such an extension, special cases of propositions concerning the consequences of making decisions pursuant to a model that did not allow for source credibility could be developed. The other extension is to adopt an agency theory approach and characterize the manager/senior relationship as a principle/agent relationship. The conflict of interest would be in the senior's disutility for diligence and the manager's indifference to the senior's diligence beyond its affect on payoffs. Source credibility would arise in the form of
uncertainty regarding signals received about the senior's actions.

There are extensions involving both analytical and experimental research. One can be called the flexibility issue in audit judgment. The flexibility available to the auditor in selecting audit procedures and effort was referred to in several places in the study. Part of this flexibility arises from the sequential dependencies existing in the audit process and the availability of multiple-objective tests. While some modelling of these dependencies has already been undertaken the complexities involved have limited progress. Investigation of simpler approaches is warranted. On the descriptive side, how auditors make interdependent judgments, and to what extent auditors recognize the flexibility available is of interest. Finally, research on the audit review process is warranted. Although the conclusions of this study are relevant to understanding the effectiveness of the review process, it was not specifically examined here. The audit review process has not been investigated yet.
A Bayesian Representation of the Audit Team

Members' Judgments
Bayes' Theorem is a statement of conditional probability that exploits the dependence between two or more variables. Its computational form is derived from the definition of a conditional probability:

\[ P(A|B) = \frac{P(AnB)}{P(B)} \]

given that,

\[ P(AnB) = P(B|A)P(A) \]

\[ P(B) = \sum_{i}P(B|A_i)P(A_i) \]

Then

\[ P(A_i|B) = \frac{P(B|A_i)P(A_i)}{\sum_{i}P(B|A_i)P(A_i)} = \frac{P(A_i)}{P(B)} \cdot P(B|A_i) \]

The ratio form of Bayes' Theorem is often used:

\[ \frac{P(A_1|B)}{P(A_2|B)} = \frac{P(B|A_1)P(A_1)}{P(B|A_2)P(A_2)} \]

or

\[ \Omega_1 = A_{12} \Omega_0 \]

where

\[ A_{12} = \frac{P(B|A_1)}{P(B|A_2)} \]

The ratio form emphasizes the likelihood ratio, \( A_{12} \), which is a measure of the diagnosticity of the message (B). That is, it measures the ability of the message to distinguish between competing hypotheses \( A_1 \) and \( A_2 \).

**Audit Senior**

The audit senior is only concerned with his own prior beliefs and the audit evidence so that the standard Bayes formula is sufficient. To
aid exposition two states \( i=0,1 \) are assumed to exist.

\[
PO_1(H_0|D) = \frac{P_1(D|H_0)PR_1(H_0)}{PR_1(D)}
\]

\[
= \frac{P_1(D|H_0)PR_1(H_0)}{P_1(D|H_0)PR_1(H_0) + P_1(D|H_1)PR_1(H_1)}
\]

**Audit Manager**

The manager not only has the accumulated audit evidence to weight in forming his judgment, but also the recommendation of the senior and his implicit appraisal of the quality of the senior's work. The manager's judgment can be represented as:

\[
PO_2(H_0|PO_1,D) = \frac{P_2(PO_1,D|H_0)PR_2(H_0)}{P_2(PO_1,D|H_0)PR_2(H_0) + P_2(PO_1,D|H_1)PR_2(H_1)}
\]

The senior's evaluation \( (PO_1) \) and the audit evidence \( (D) \) are conditionally dependent, that is, their joint occurrence contains information in and of itself to the manager. Therefore

\[ P_2(PO_1,D|H_0) \neq P_2(PO_1|H_0)P_2(D|H_0) \]

rather

\[ P_2(PO_1,D|H_0) = P_2(D|H_0)P_2(PO_1|H_0,D) \]

With this condition, the model of the manager's judgment becomes:

\[
PO_2(H_0|PO_1,D) = \frac{P_2(PO_1,H_0,D)P_2(D|H_0)PR_2(H_0)}{P_2(PO_1,H_0,D)P_3(D|H_0)PR_2(H_0) + P_2(PO_1,H_1,D)P_2(D|H_1)PR_2(H_1)}
\]

**Audit Partner**

The same operation can be carried out for the model of the partner

\[
PO_3(H_0|PO_2,PO_1,D) =
\]
\[
\frac{P_3(PO_2, PO_1, D|H_0)P_3(H_0)}{P_3(PO_2, PO_1, D|H_0)P_3(H_0) + P_3(PO_2, PO_1, D|H_1)P_3(H_1)}
\]

Where

\[
P_3(PO_2, PO_1, D|H_0) = P_3(D|H_0)P_3(PO_2, PO_1|H_0, D)
\]

\[
= P_3(D|H_0)P_3(PO_1|H_0, D)P_3(PO_2|H_0, D, PO_1)
\]

therefore

\[
P_03(H_0|PO_2, PO_1, D) =
\]

\[
\frac{P_3(PO_2|H_0, D, PO_1)P_3(PO_1|H_0, D)P_3(D|H_0)P_3(H_0)}{P_3(PO_2|H_0, D, PO_1)P_3(PO_1|H_0, D)P_3(D|H_0)P_3(H_0) + P_3(PO_2|H_1, D, PO_1)P_3(PO_1|H_1, D)P_3(D|H_1)P_3(H_1)}
\]
APPENDIX B

Extension of the Bayesian Model to Include a Measure of Source Credibility
The following manipulation is involved in extending the Bayesian model to include a measure of source credibility (Schum and DuCharme, op. cit.).

Given
\[
P(D_{\xi} \mid H_1) = \frac{P(D_{\xi} \mid nH_1)}{P(H_1)}
\]
and
\[
P(D_{\xi} \mid nH_1) = \sum_{k} P(D_{\xi} \mid D_k^* \mid nH_1)
\]

This statement says that \(P(D_{\xi} \mid nH_1)\) can be written as a disjoint union of several compound events, where the \(D_k^*\) form a mutually exclusive and exhaustive set of data states. This means that in the evaluation of the client's internal control system, for example, the evaluation provides a message that the system is either strong or weak (not strong), although the signal is not restricted to only two messages.

and
\[
P(D_{\xi} \mid nD_k^* \mid nH_1) = P(D_{\xi} \mid H_1 nD_k^*) P(D_k^* \mid H_1) P(H_1)
\]
then
\[
P(D_{\xi} \mid H_1) = \sum_{k} P(D_{\xi} \mid H_1 nD_k^*) P(D_k^* \mid H_1)
\]

Assuming that credibility is independent of the particular hypothesis under consideration, then
\[
P(D_{\xi} \mid H_1 nD_k^*) = P(D_{\xi} \mid D_k^*)
\]
so
\[
P(D_{\xi} \mid H_1) = \sum_{k} P(D_{\xi} \mid D_k^*) P(D_k^* \mid H_1)
\]

This step involves a conditional independence assumption, that is, the dependence between \(H_1\) and \(D_{\xi}\) is only due to the dependence between \(H_1\) and \(D_k^*\) and \(D_k^*\) and \(D\). In an auditing context, presumably the credibility
of an auditor does not change over hypothesized states of the client.

The measure of source credibility is $P(D_{\lambda} | D^*_{k})$ and the likelihood ratio becomes

$$A_{oj} = \frac{P(D_{\lambda} | H_o)}{P(D_{\lambda} | H_j)} = \frac{\sum_k P(D_{\lambda} | D^*_{k}) P(D^*_{k} | H_o)}{\sum_k P(D_{\lambda} | D^*_{k}) P(D^*_{k} | H_j)}$$
APPENDIX C

Steps Involved in Calculating the Subject's Inferred Likelihood Ratio and the Model's Derived Likelihood Ratio
Subject's Inferred Likelihood Ratio

The subjects provided their prior probability and posterior probability that the client's internal control system functioned sufficiently well to allow substantial reliance on it. The task was set in a two-state framework in that the system functioned or did not function sufficiently well. When only two alternative states are involved, knowing the probability of one state's existence implies knowledge of the other probability. Therefore, the subjects' inferred likelihood ratio could be calculated from the two responses provided by the subjects in the following manner.

Given \[ \Omega_1 = A_{12} \Omega_0 \]
then \[ A_{12} = \frac{\Omega_1}{\Omega_0} \cdot \frac{1}{1-\Omega_0} \]

\[ = \frac{P(H_1 | D_1)}{P(H_2 | D_1)} \cdot \frac{P(H_2)}{P(H_1)} \]

Model's Derived Likelihood Ratio

The model's derived likelihood ratio can be represented as:

\[ A = \frac{P_1(r) + (1-P_1)(1-r)}{P_2(r) + (1-P_2)(1-r)} \]

Where: \[ r = P(D_1 | D_1^*) = P(D_2 | D_2^*) \]; \[ 1-r = P(D_1 | D_2^*) = P(D_2 | D_1^*) \]
\[ P_1 = P(D_1^* | H_1); 1-P_1 = P(D_2^* | H_1) \]
\[ P_2 = P(D_1^* | H_2); 1-P_2 = P(D_2^* | H_2) \]
\[ i = \ell = k = 1,2 \]
In the case of perfect source credibility (r=1) the likelihood ratio refers to the standard Bayesian likelihood ratio:

\[ A = \frac{P_1}{P_2} \]

Now \( P_1 \), which is the level of confidence at which the compliance tests are conducted, is specific in the case. The subject's likelihood ratio \( A(S) \) is inferred. Consequently, substituting \( A(S) \) for \( A(M) \) enables solving for \( P_2 \). Knowing both \( P_1 \) and \( P_2 \), we can consider the three cases of less than perfect source credibility, comparing the subject's inferred likelihood ratio with the model's prediction.

In order to make this comparison, three assumptions are implied:

1. In the case of perfect source credibility, the subjects' likelihood ratios were the same as those of the model.

2. The derived conditional probability \( (P_2) \) was stable over difference source credibility levels.

3. The level of confidence at which the individual compliance tests were conducted \( (P_1) \) equaled the confidence level for the overall test of system compliance.

Assumption one involves the basic step explained above for estimating \( P_2 \). This assumption was not expected to hold and, based on the results, did not. Nevertheless, the primary concern was how subjects adjusted for variations in source credibility compared to the model, not whether the subjects were really Bayesian. For the second assumption, (potential) event impact and source credibility are independent constructs. \( P_2 \) is related to \( P_1 \) which was controlled in the experiment. Also relevant to both the first two assumptions is the sensitivity
analysis conducted in relation to assumption three. The results of the
difference analysis were remarkably robust to alternative specifications
of the model's likelihood ratio.

The third assumption requires more detailed discussion because it
raises a familiar issue in statistics, but an issue that, as far as the
researcher is aware, has not been addressed in auditing: the difference
between the error rate per comparison and the experimentwise error rate.
The assumption was necessary because internal control evaluation has no
standard evaluatin criteria, rather the conclusion on the system is based
on the auditor's own professional judgment. More specifically, the
assumption was necessary because of the nature of compliance testing. An
appraisal of the functioning of the system does not depend on the outcome
of one particular test, but a subjective interpretation of the overall
compliance test results. Since compliance tests are designed to test
specific aspects of a client's system they are somewhat independent, but
multiple objective tests and the dependence between aspects of a client's
system mean that the tests are also somewhat dependent.

In the case study, seven separate compliance tests were conducted
by the senior. The researcher cannot really specify how many of these
tests must fail for the result to be interpreted as a message that the
system is not strong, that is, substantial reliance cannot be placed on
it. If, for the system to be judged strong no rejections among the seven
tests are required, and the tests are independent, the experimentwise
confidence levels are:

\[(P_1 = .95)^7 = .70\]
\[(P_1 = .90)^7 = .48\]
Alternatively, if either no rejections or one rejection indicates a strong system then under the binomial distribution:

\[(.95)^7 + \binom{7}{1} (.95)^6 (.05)^1 = .95\]
\[(.90)^7 + \binom{7}{1} (.90)^6 (.05)^1 = .85\]

Consequently, an approximation of the experimentwise error rate was unavoidable. The assumption made is considered reasonable. It needs to be emphasized that the related issues of how compliance tests are utilized and what is the role of the experimentwise error rate in auditing have not been researched.

What is the effect of this assumption not holding? If the confidence level for the overall test of system compliance is less than \(P_1\), then the precision of the test declines. It is to be expected that the test's diagnosticity (L) declines. In turn, source credibility has a moderating influence so that the decline in A would be less in L. Therefore, if the finding was one of conservatism under the error rate per comparison (\(P_1\)) then the conservatism would be less pronounced under the experimentwise error rate. There would be an upward shift in the plots of the treatment means. These effects were in fact observed when the difference analysis was conducted using the extreme experimentwise confidence levels of .70 and .48 in place of .95 and .90. The results of the difference analysis presented in Chapter V all held. In the new cluster analysis only two subjects changed clusters. Figures 13 and 14 are provided for comparison purposes. Figure 13 includes mean differences under both confidence levels, while Figure 14 should be compared with Figure 9. The comparison indicates that the assumption is reasonable. Whether auditors have intuitively reached the same conclusion is an interesting question.
FIGURE 13. MEAN DIFFERENCES AS A FUNCTION OF SOURCE CREDIBILITY UNDER TWO DIFFERENT STATISTICAL CONFIDENCE LEVEL ASSUMPTIONS*

*This Figure extends Figure 8
A1 and A2 are the alternative confidence levels
EC is the individual comparison confidence level
EW is the experimentwise confidence level
FIGURE 14. MEAN DIFFERENCES AS A FUNCTION OF SOURCE CREDIBILITY BY CLUSTER, BASED ON EXPERIMENTWISE CONFIDENCE LEVEL*

*This figure can be compared to Figure 9
The numbers in the brackets are the subjects in the respective clusters.
APPENDIX D

The Test Instrument for Stage One of the Experiment
This casestudy deals with the review, by the manager-in-charge, of an audit senior's recommendations relating to reliance on internal control in determining the extent of confirmation of accounts receivable. The casestudy focuses specifically on compliance tests the senior has conducted.

You are asked to assume the role of the newly-assigned manager-in-charge of the Erwin audit for the year ending December 31, 1979. The total staff time for the 1979 audit should approximate 400 hours. Except for yourself and the manager you are replacing, the audit personnel are the same as on last year's audit. The previous manager resigned while the interim audit work was in progress.

Your task in the casestudy is to evaluate recommendations made by the audit senior, Neil Tacket, using different assumptions about his competence.

To keep the casestudy manageable while as realistic as possible, we have focused it entirely on the audit procedures relating to part of Erwin's revenue cycle - the recording of sales and accounts receivable. In the attached materials, you will find:

- a description of the client,
- a description of the client's accounting system for sales and receivables,
- an appraisal of the system,
- a preliminary task for you to perform,
- a description of the senior, Neil Tacket, and the compliance tests he conducted,
- your main task.

Description of the Client

The Erwin Hardware Company is a wholesale distributor of hardware to quality hardware stores and hardware chains in the mid-west. The owners do not take an active part in running the company. Your firm's assessment of Erwin's management is that it is honest and competent. Erwin's accounting personnel are regarded by your firm as reasonably competent and highly trustworthy.

Accounts receivable is a critical audit area representing 24% of total assets at September 30, 1979. For the nine months ended that date a net income of $740,000 was earned on net sales of $15,000,000.
Description of the Client's Accounting System for Sales and Receivables

A description and a flowchart of the client's accounting system for sales and receivables follow.

Customers' orders are received by salesmen who forward them to the sales order clerk. Since all sales are credit sales, the orders must be approved by the credit manager. If approved, the customer's order is accepted and a 'sales order' is prepared.

A copy of each sales order is sent to the shipping department and to the billing department. When the goods are shipped, a copy of the bill of lading is sent to the billing department; there, a billing clerk uses both documents and a standard price list to prepare an invoice. Before the original of the invoice is mailed to the customer, a clerk from the accounting department checks both the description of the goods and the quantities to both the sales order and the bill of lading, checks the prices to the standard price list, and checks the extensions and additions; this clerk then initials the duplicate copy of the invoice.

The duplicate copies are sent to the accounting department, where a bookkeeper enters them in the sales journal (sequentially by invoice number) and in the accounts receivable subsidiary ledger.

The general ledger bookkeeper summarizes the sales journal monthly and posts the totals to the general ledger accounts affected. The chief accountant accounts for sales order numbers, bill of lading numbers, and invoice numbers weekly and reconciles the accounts receivable ledger with the corresponding account in the general ledger monthly.

Note: Cash collections and sales returns and allowances are not considered in this case study.
Appraisal of the System

The appraisal of the client's accounting system for sales and accounts receivable was conducted under the supervision of the previous audit manager. The conclusion (subject to the compliance test results) was that the internal control (accounting) system for sales and receivables was sufficiently strong to justify substantial reliance on it in planning substantive tests, particularly the extent of confirmation of accounts receivable. Compliance tests to see whether the system functioned sufficiently well to justify the reliance contemplated on the basis of this appraisal are the focus of this case study. The compliance test results are given later.

In last year's audit no reliance was placed on the client's internal control system. This decision was made because the system was appraised as only fair and the compliance tests found numerous exceptions. As a result of last year's 'management letter' to the client, two changes were made in the system, effective January 1, 1979; (1) sales invoices were to be recorded in the sales journal according to numerical sequence, rather than in the sequence of the 'sales orders' to which they relate, and (2) an accounting clerk was to check the quantity, pricing, extension and footing of each invoice before the original was mailed to the customer.

You are asked to accept the conclusion that the internal control system for sales and receivables is sufficiently strong to justify substantial reliance on it for the purpose of confirmation of accounts receivable. Of concern in this case study is whether the system has functioned sufficiently well to justify substantial reliance on it for the above purpose.

1. In future when a reference is made to substantial reliance being placed on the client's system, it should be interpreted that the substantial reliance refers to determining the extent of confirmation of accounts receivable. The effect on other audit procedures is not part of this case study.
Preliminary Task

Before the compliance test results are considered, please answer the following question by drawing on (1) your own experience with similar clients and audits, and on (2) the information presented in the case study.

What do you regard as the likelihood that the following statement is true?

The client's internal control system for sales and accounts receivable has functioned sufficiently well to justify substantial reliance on it in determining the extent of confirmation of accounts receivable.

Please indicate your response on the following line scale.

Example: Marking the scale at .49 would indicate a belief that there is a 49 percent likelihood that the system has functioned sufficiently well to justify substantial reliance on it, and a 51 percent (1-.49) likelihood that it has not functioned sufficiently well to justify substantial reliance on it.
Description of the Senior, Neil Tacket, and the Compliance Tests

The compliance tests (described below) were conducted by the senior, Neil Tacket. You have worked with Tacket on other audits, and your evaluation of him is shared by other audit managers. You (and the other managers) regard Tacket as a competent administrator, especially adept at dealing with people, both audit and client. His technical ability is considered later. Four different evaluations of his technical ability, in particular, his consistency and care in conducting detailed audit work and making audit judgments are given.

Tacket is out of town on another engagement and you cannot discuss with him in person the details of the interim audit procedures he performed. Consequently, in forming your own opinion on the reliance to be placed on the client's system, the weight you give to the compliance tests conducted and the recommendation made by Tacket will be the weight that you consider your evaluation of him warrants.

The compliance tests performed by Tacket follow and the test results are given on the next page. You are asked to accept these tests as being adequate for the purpose of determining how the system has functioned. The question about the tests, as you will learn later, is whether you believe Tacket has conducted them competently and has presented the results accurately.

a. Account for a sequence of sales invoices in the sales journal.

b. Trace selected invoices from the sales journal to
   1) Accounts receivable subsidiary ledger, and test for amount, date, and invoice number.
   2) Duplicate sales invoice, and test for the amount recorded in the journal, date, customer name, and classification. Check the quantity, pricing, extensions, footings, and evidence of internal verification on the duplicate sales invoice.
   3) Bill of lading, and test for customer name, product description, quantity, and date.
   4) Duplicate sales order, and test for customer name, product description, quantity, and date.

c. Trace selected bills of lading to duplicate sales invoices and the invoices to the sales journal. Check for customer name, product description, quantity and date of shipment.
**ATTRIBUTE SAMPLING DATA SHEET**

**Client:** Erwin Hardware  
**Audit Area:** Compliance Tests - Sales  
**Interim Audit 1979**  
**Pop. Size:** 34,647

**Definition of Objectives:** Examine duplicate sales invoices and related documents to determine if the system has functioned as intended.

**Definition of the population:** Sales invoices for the 1/1/’79 to 9/31/’79. First invoice number = 30,000. Last invoice number = 64,646.

**Definition of the sampling unit, organization of population items and random selection procedures:** Sales invoice number, recorded in the sales journal sequentially, random number table.

<table>
<thead>
<tr>
<th>DESCRIPTION OF ATTRIBUTE</th>
<th>AUDIT PLAN*</th>
<th>TEST RESULTS</th>
<th>ANALYSIS OF ERRORS</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Expect. error rate</td>
<td>Upper error prec. rate</td>
<td>Sample error rate</td>
</tr>
<tr>
<td>1. Existence of the sales invoice number in the sales journal.</td>
<td>0%</td>
<td>2%</td>
<td>0%</td>
</tr>
<tr>
<td>2. Amount and other data in the subs. ledger agree with the sales journal entry.</td>
<td>1</td>
<td>4</td>
<td>0.5</td>
</tr>
<tr>
<td>3. Amount and other data in the sales invoice agree with the sales journal entry.</td>
<td>1</td>
<td>4</td>
<td>0</td>
</tr>
<tr>
<td>4. Evidence that quantity, pricing, extensions, and footings are checked (initials and correct amount).</td>
<td>1</td>
<td>4</td>
<td>1</td>
</tr>
<tr>
<td>5. Quantity and other data on bill of lading agree with the duplicate sales invoice.</td>
<td>1</td>
<td>4</td>
<td>0.5</td>
</tr>
<tr>
<td>6. Quantity and other data on the sales order agree with the duplicate sales invoice.</td>
<td>1</td>
<td>6</td>
<td>0</td>
</tr>
<tr>
<td>7. Quantity and other data on the customer order agree with the duplicate sales invoice.</td>
<td>2</td>
<td>8</td>
<td>0</td>
</tr>
</tbody>
</table>

*Sample sizes were obtained from sample size tables for confidence level used. The confidence level is provided later.

The other tests are not part of this casestudy.

**Recommendations:**

1. System has functioned sufficiently well to allow substantial reliance on it in determining the extent of confirmation of accounts receivable.
2. Specific recommendations concerning the errors found should be discussed with management.

**Note:** The effect of the errors found on audit procedures other than confirmation of accounts receivable is not considered in the casestudy.
Description of Your Main Task

Four situations are provided. In each situation you are given two additional items of information:

1. The confidence level at which Tacket performed the compliance tests.
2. A specific evaluation of Tacket.

These items of information are then varied across situations. In all, you are asked to make eight responses.

Your task is to re-appraise the likelihood that the client's system has functioned sufficiently well to justify substantial reliance upon it, given:

A. The results of the compliance tests conducted by Tacket.
B. Tacket's recommendation about reliance on the system.
C. The two items of 'additional information'.
SITUATION

The senior, Neil Tacket, recommends that substantial reliance be placed on the client's system in determining the extent of confirmation of accounts receivable.

Additional Information:

1. Compliance Tests: The compliance test were conducted with a 90% confidence level.
2. Evaluation of Tacket: The consensus among the audit managers, including yourself, who have worked with Tacket is that he is completely reliable in conducting audit work and in presenting the results. In terms of the rating scale given below, he is considered to be 100% reliable—which means that, for example, in the past the results he has presented have been found in review to be always acceptable.

<table>
<thead>
<tr>
<th>50%</th>
<th>60%</th>
<th>70%</th>
<th>80%</th>
<th>90%</th>
<th>100%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Completely</td>
<td>Highly</td>
<td>Fairly</td>
<td>Fairly</td>
<td>Highly</td>
<td>Completely</td>
</tr>
<tr>
<td>Unreliable</td>
<td>Unreliable</td>
<td>Unreliable</td>
<td>Reliable</td>
<td>Reliable</td>
<td>Reliable</td>
</tr>
</tbody>
</table>

What do you regard is the likelihood that the following statement is true given the above additional information, Tacket's recommendation, and the other information contained in the case study?

The client's internal control system for sales and accounts receivable functioned sufficiently well to justify substantial reliance on it in determining the extent of confirmation of accounts receivable.

No | 1 | 3 | 4 | 6 | 8 | 10
Likelihood | Fifty-Fifty | Likelihood | Certain | Likelihood

If the evaluation of Tacket's reliability is still the same, but the compliance tests were conducted with a 95% rather than a 90% confidence level, would your response change?

Yes or No (circle one)

If yes indicate the new likelihood that the statement is true.

No | 1 | 3 | 4 | 6 | 8 | 10
Likelihood | Fifty-Fifty | Likelihood | Certain | Likelihood
The senior, Neil Tacket, recommends that substantial reliance be placed on the client's system in determining the extent of confirmation of accounts receivable.

Additional Information:

1. Compliance Tests: The compliance tests were conducted with a 90% confidence level.
2. Evaluation of Tacket: The consensus among the audit managers, including yourself, who have worked with Tacket is that he is fairly unreliable in conducting audit work and in presenting the results. In terms of the rating scale given below, he is considered to be 70% reliable—which means that, for example, in the past the results he has presented have been found in review to be acceptable 70% of the time and unacceptable the other 30%.

What do you regard is the likelihood that the following statement is true given the above additional information, Tacket's recommendation, and the other information contained in the case study?

The client's internal control system for sales and accounts receivable functioned sufficiently well to justify substantial reliance on it in determining the extent of confirmation of accounts receivable.

If the evaluation of Tacket's reliability is still the same, but the compliance tests were conducted with a 95% rather than a 90% confidence level, would your response change?

Yes or No (circle one)

If yes indicate the new likelihood that the statement is true.
SITUATION 3

The senior, Neil Tacket, recommends that substantial reliance be placed on the client's system in determining the extent of confirmation of accounts receivable.

Additional Information:

1. Compliance Tests: The compliance test were conducted with a 90% confidence level.
2. Evaluation of Tacket: The consensus among the audit managers, including yourself, who have worked with Tacket is that he is fairly reliable in conducting audit work and in presenting the results. In terms of the rating scale given below, he is considered to be 80% reliable-which means that, for example, in the past the results he has presented have been found in review to be acceptable 80% of the time and unacceptable the other 20%.

What do you regard is the likelihood that the following statement is true given the above additional information, Tacket's recommendation, and the other information contained in the case study?

The client's internal control system for sales and accounts receivable functioned sufficiently well to justify substantial reliance on it in determining the extent of confirmation of accounts receivable.

If the evaluation of Tacket's reliability is still the same, but the compliance tests were conducted with a 95% rather than a 90% confidence level, would your response change?

Yes or No (circle one)

If yes indicate the new likelihood that the statement is true.
SITUATION

The senior, Neil Tacket, recommends that substantial reliance be placed on the client's system in determining the extent of confirmation of accounts receivable.

Additional Information:

1. Compliance Tests: The compliance tests were conducted with a 90% confidence level.
2. Evaluation of Tacket: The consensus among the audit managers, including yourself, who have worked with Tacket is that he is highly reliable in conducting audit work and in presenting the results. In terms of the rating scale given below, he is considered to be 90% reliable—which means that, for example, in the past the results he has presented have been found in review to be acceptable 90% of the time and unacceptable the other 10%.

<table>
<thead>
<tr>
<th>50%</th>
<th>60%</th>
<th>70%</th>
<th>80%</th>
<th>90%</th>
<th>100%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Completely Unreliable</td>
<td>Highly Reliable</td>
<td>Fairly Unreliable</td>
<td>Fairly Reliable</td>
<td>Highly Unreliable</td>
<td>Completely Reliable</td>
</tr>
</tbody>
</table>

What do you regard is the likelihood that the following statement is true given the above additional information, Tacket's recommendation, and the other information contained in the case study?

The client's internal control system for sales and accounts receivable functioned sufficiently well to justify substantial reliance on it in determining the extent of confirmation of accounts receivable.

If the evaluation of Tacket's reliability is still the same, but the compliance tests were conducted with a 95% rather than a 90% confidence level, would your response change?

Yes or No (circle one)

If yes indicate the new likelihood that the statement is true.
APPENDIX E

The Test Instrument for Stage Two of the Experiment
This casestudy deals with the review, by the manager-in-charge, of an audit senior's recommendations relating to reliance on internal control in determining the extent of confirmation of accounts receivable.

You are asked to assume the role of the newly-assigned manager-in-charge of the Erwin audit for the year ending December 31, 1979. The total staff time for the 1979 audit should approximate 400 hours. Except for yourself and the manager you are replacing, the audit personnel are the same as on last year's audit. The previous manager resigned while the interim audit work was in progress.

Your task in the casestudy is to decide on the sample sizes for confirmation of accounts receivable. Several situations are presented. In each situation, different assumptions are made about the audit work background of the senior, Neil Tacket, who has conducted the compliance tests and recommended confirmation sample sizes.

In the attached materials you will find:

- a description of the client,
- a description of the client's accounting system for sales and receivables,
- an appraisal of the system,
- a description of the senior, Neil Tacket, and the compliance tests he conducted,
- your task,
- a post experimental questionnaire.

Description of the Client

The Erwin Hardware Company is a wholesale distributor of hardware to quality hardware stores and hardware chains in the mid-west. The owners do not take an active part in running the company. Your firm's assessment of Erwin's management is that it is honest and competent. Erwin's accounting personnel are regarded by your firm as reasonably competent and highly trustworthy.

Accounts receivable is a critical audit area representing 24% of total assets at September 30, 1979. For the nine months ended that date a net income of $740,000 was earned on net sales of $15,000,000.
Description of the Client's Accounting System for Sales and Receivables

A description and a flowchart of the client's accounting system for sales and receivables follow.

Customers' orders are received by salesmen who forward them to the sales order clerk. Since all sales are credit sales, the orders must be approved by the credit-manager. If approved, the customer's order is accepted and a 'sales order' is prepared.

A copy of each sales order is sent to the shipping department and to the billing department. When the goods are shipped, a copy of the bill of lading is sent to the billing department; there, a billing clerk uses both documents and a standard price list to prepare an invoice. Before the original of the invoice is mailed to the customer, a clerk from the accounting department checks both the description of the goods and the quantities to both the sales order and the bill of lading, checks the prices to the standard price list, and checks the extensions and additions; this clerk then initials the duplicate copy of the invoice.

The duplicate copies are sent to the accounting department, where a bookkeeper enters them in the sales journal (sequentially by invoice number) and in the accounts receivable subsidiary ledger.

The general ledger bookkeeper summarizes the sales journal monthly and posts the totals to the general ledger accounts affected. The chief accountant accounts for sales order numbers, bill of lading numbers, and invoice numbers weekly and reconciles the accounts receivable ledger with the corresponding account in the general ledger monthly.

Note: Cash collections and sales returns and allowances are not considered in this case study.
Appraisal of the System

In last year's audit no reliance was placed on the client's internal control system. This decision was made because the system was appraised as only fair and the compliance tests found numerous exceptions. Last year's confirmation results are given later.

As a result of last year's 'management letter' to the client, two changes were made in the system, effective January 1, 1979: (1) sales invoices were to be recorded in the sales journal according to numerical sequence, rather than in the sequence of the 'sales orders' to which they relate, and (2) an accounting clerk was to check the quantity, pricing, extension and footing of each invoice before the original was mailed to the customer.

The appraisal of the client's accounting system for sales and accounts receivable was conducted under the supervision of the previous audit manager. The conclusion (subject to the compliance test results) was that the internal control (accounting) system for sales and receivables was sufficiently strong to justify substantial reliance on it in planning substantive tests, particularly the extent of confirmation of accounts receivable. You are asked to accept this conclusion. Of concern in this casestudy is whether the system has functioned sufficiently well to justify substantial reliance on it in determining the extent of confirmation of accounts receivable.
Description of the Senior, Neil Tacket, and the Compliance Tests

The compliance tests (described below) were conducted by the senior, Neil Tacket. You have worked with Tacket on other audits, and your evaluation of him is shared by other audit managers. You (and the other managers) regard Tacket as a competent administrator, especially adept at dealing with people, both audit and client. Information will be given on his technical ability later.

Tacket is out of town on another engagement and you cannot discuss with him in person the details of the interim audit procedures he performed, nor his rationale for his sample size recommendations. Consequently, in forming your own opinion on the appropriate sample sizes for confirmation of accounts receivable, the weight you give to the compliance tests conducted and the sample size recommendations made by Tacket will be the weight that you consider your evaluation of him warrants.

The compliance tests performed by Tacket follow and the test results are given on the next page. You are asked to accept these tests as being adequate for the purpose of determining how the system has functioned. The question about the tests, as you will learn later, is whether you believe Tacket has conducted them competently and has presented the results accurately.

a. Account for a sequence of sales invoices in the sales journal.

b. Trace selected invoices from the sales journal to
   1) Accounts receivable subsidiary ledger, and test for amount, date, and invoice number.
   2) Duplicate sales invoice, and test for the amount recorded in the journal, date, customer name, and classification. Check the quantity, pricing, extensions, footings, and evidence of internal verification on the duplicate sales invoice.
   3) Bill of lading, and test for customer name, product description, quantity, and date.
   4) Duplicate sales order, and test for customer name, product description, quantity, and date.

   c. Trace selected bills of lading to duplicate sales invoices and the invoices to the sales journal. Check for customer name, product description, quantity and date of shipment.
ATTRIBUTE SAMPLING DATA SHEET

Client: Ervin Hardware
Audit Area: Compliance Tests - Sales
Definition of Objectives: Examine duplicate sales invoices and related documents to determine if the system has functioned as intended.
Definition of the population: Sales invoices for the 1/1/79 to 9/30/79. First invoice number = 33,000. Last invoice number = 64,464.
Definition of the sampling unit, organization of population items and random selection procedures: Sales invoice number, recorded in the sales journal sequentially, random number table.

DESCRIPTION OF ATTRIBUTE

<table>
<thead>
<tr>
<th></th>
<th>AUDIT PLAN</th>
<th>TEST RESULTS</th>
<th>ANALYSIS OF ERRORS</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Existence of the sales invoice number in the sales journal.</td>
<td>0% 2% 0% 2%</td>
<td>---</td>
<td></td>
</tr>
<tr>
<td>2. Amount and other data in the sales journal agree with the sales invoice.</td>
<td>1 4 0.3 3</td>
<td>Posting to the wrong customer's account. Total amount involved less than $400.</td>
<td></td>
</tr>
<tr>
<td>3. Amount and other data in the sales invoice agree with the sales journal entry.</td>
<td>1 4 0 2</td>
<td>---</td>
<td></td>
</tr>
<tr>
<td>4. Evidence that quantity, pricing, extensions, and footings are checked (initials and correct amount).</td>
<td>1 4 1 4</td>
<td>The accounting clerk's initials were not recorded. No dollar errors found.</td>
<td></td>
</tr>
<tr>
<td>5. Quantity and other data on bill of lading agree with the duplicate sales invoice.</td>
<td>1 4 0.3 3</td>
<td>The date on the duplicate sales invoice was several days later than the shipping date.</td>
<td></td>
</tr>
<tr>
<td>6. Quantity and other data on the sales order agree with the duplicate sales invoice.</td>
<td>1 6 0 3.8</td>
<td>---</td>
<td></td>
</tr>
<tr>
<td>7. Quantity and other data on the customer order agree with the duplicate sales invoice.</td>
<td>2 8 0 3.8</td>
<td>---</td>
<td></td>
</tr>
</tbody>
</table>

Sample sizes were obtained from sample size tables using a 95% confidence level.

The other tests are not part of this case study.

Recommendations:
1. System has functioned sufficiently well to allow substantial reliance on it in determining the extent of confirmation of accounts receivable.
2. Specific recommendations concerning the errors found should be discussed with management.

Note: The effect of the errors found on audit procedures other than confirmation of accounts receivable is not considered in this case study.
ENVIN HARDWARE COMPANY

The results of last year's confirmation work was as follows (as of 10/31/’79):

<table>
<thead>
<tr>
<th>TYPE</th>
<th>NUMBER</th>
<th>Mailed %</th>
<th>NUMBER</th>
<th>Received %</th>
<th>TOTAL DOLLARS</th>
<th>Mailed %</th>
<th>Received %</th>
</tr>
</thead>
<tbody>
<tr>
<td>Past Due:</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>90 days</td>
<td>15</td>
<td>100.0%</td>
<td>6</td>
<td>40.0%</td>
<td>10,400</td>
<td>100.0%</td>
<td>4,462</td>
</tr>
<tr>
<td>45-90 days</td>
<td>34</td>
<td>100.0%</td>
<td>33</td>
<td>61.1%</td>
<td>51,110</td>
<td>100.0%</td>
<td>31,689</td>
</tr>
<tr>
<td>OVER 51,500</td>
<td>15</td>
<td>100.0%</td>
<td>29</td>
<td>82.9%</td>
<td>106,499</td>
<td>100.0%</td>
<td>85,199</td>
</tr>
<tr>
<td>UNDER 51,500</td>
<td>2,001</td>
<td>16.8%</td>
<td>149</td>
<td>78.8%</td>
<td>1,443,713</td>
<td>16.5%</td>
<td>186,477</td>
</tr>
<tr>
<td>TOTAL</td>
<td>2,146</td>
<td>20.0%</td>
<td>333</td>
<td>74.2%</td>
<td>1,614,762</td>
<td>25.4%</td>
<td>307,728</td>
</tr>
</tbody>
</table>

Alternative work was performed on all non-replies.

There were no exceptions noted of audit significance on either the replies or as a result of the alternative
test work.

The confirmation work was reviewed by the previous audit manager. You are asked to accept its accuracy.

DETAILS OF ACCOUNTS RECEIVABLE

AS OF 10/31/’79

<table>
<thead>
<tr>
<th>Range</th>
<th>No.</th>
<th>Amount</th>
</tr>
</thead>
<tbody>
<tr>
<td>$ 0 - 500</td>
<td>720</td>
<td>$ 250,130</td>
</tr>
<tr>
<td>500 - 1000</td>
<td>680</td>
<td>322,700</td>
</tr>
<tr>
<td>1000 - 1500</td>
<td>342</td>
<td>423,622</td>
</tr>
<tr>
<td>1500 - 2000</td>
<td>150</td>
<td>242,609</td>
</tr>
<tr>
<td>2000 - 2500</td>
<td>75</td>
<td>166,651</td>
</tr>
<tr>
<td>2500 &amp; above</td>
<td>45</td>
<td>111,310</td>
</tr>
<tr>
<td>TOTAL</td>
<td>2,008</td>
<td>1,740,932</td>
</tr>
</tbody>
</table>

Included in the above amounts were 33 accounts amounting to $122,319 which were
past due between 45 and 90 days, and 14 accounts amounting to $219,102 which
were past due 90 days or more.

TACKET'S RECOMMENDATIONS

On the basis of the results of the compliance tests be conducted, Tacket recommends that substantial reliance
be placed on the system. Tacket also recommends the following sample sizes for the confirmation of accounts
receivable.

Confirmation of accounts receivable to be done as of October 31, 1979. Send out positive confirmations as
follows:

1. 0 accounts over 90 days past due.
2. 20 accounts past due between 45 and 90 days.
3. 200 accounts over $1,500.
4. 100 other accounts.
Description of Your Task

On the basis of the compliance test results he conducted, Tacket also provides sample size recommendations for the confirmation of accounts receivable.

Eight alternative situations are presented on the following pages. In each situation you are asked to indicate:

A. The sample sizes that in your professional judgment are appropriate.
B. The weight you have placed, in making your sample size judgments, on the compliance tests conducted by Tacket.
C. The weight you have placed, in making your sample size judgments, on the sample size recommendations of Tacket.

As previously stated, of concern is your belief about the quality of Tacket's work - whether he has conducted the compliance tests competently, has presented the results accurately, and has made his sample size recommendations soundly. In order to form an opinion on these matters, you are given three items of information about Tacket:

1. Experience level.
2. Performance level.
3. Approach to sample size recommendations.
These three items are varied across situations.

On the opposite page you are given last year's confirmation results, details of this year's accounts receivable, and Tacket's sample size recommendations.
SITUATION _

INFORMATION ABOUT TACKET:

EXPERIENCE - Little - Tacket has been a senior for only about a year. Erwin is his first wholesaling client.

EVALUATION - High - Tacket has received excellent ratings in the firm's formalized performance evaluations. He is regarded as competent in supervising field work, adept at dealing with audit and client personnel, qualified in technical matters, careful in his audit work, and sound in his audit judgments. Your own appraisal of the quality of his work supports these ratings.

SAMPLING APPROACH - Statistical - In a memo, Tacket explains that he derived the sample sizes he has recommended from sample size tables, having regard for the risk levels he judged acceptable.

YOUR TASK:

A. Indicate below the sample sizes you would likely select based upon your professional judgment, considering the audit work conducted by Tacket and the impact of the above information about Tacket on your sample size judgments.

1. ___ accounts over 90 days past due.
2. ___ accounts past due between 45 and 90 days.
3. ___ accounts over $2,500.
4. ___ other accounts.

B. Indicate on the line scale below the weight you have placed, in making your sample size judgments, on the compliance tests conducted by Tacket.

\[ \begin{array}{ccccccc}
1 & 2 & 3 & 4 & 5 & 6 & 7 \\
\text{Very Low} & \text{Moderate} & \text{Very High}
\end{array} \]

C. Indicate on the line scale below the weight you have placed, in making your sample size judgments, on the sample size recommendations of Tacket.

\[ \begin{array}{ccccccc}
1 & 2 & 3 & 4 & 5 & 6 & 7 \\
\text{Very Low} & \text{Moderate} & \text{Very High}
\end{array} \]
SITUATION

INFORMATION ABOUT TACKET:

EXPERIENCE - Extensive - Tacket has been a senior for four years. He has had extensive commercial experience with both small and large clients, including wholesalers.

EVALUATION - High - Tacket has received excellent ratings in the firm's formalized performance evaluations. He is regarded as competent in supervising field work, adept at dealing with audit and client personnel, qualified in technical matters, careful in his audit work, and sound in his audit judgments. Your own appraisal of the quality of his work supports these ratings.

SAMPLING APPROACH - Statistical - In a memo, Tacket explains that he derived the sample sizes he has recommended from sample size tables, having regard for the risk levels he judged acceptable.

YOUR TASK:

A. Indicate below the sample sizes you would likely select based upon your professional judgment, considering the audit work conducted by Tacket and the impact of the above information about Tacket on your sample size judgments.

1. ___ accounts over 90 days past due.
2. ___ accounts past due between 45 and 90 days.
3. ___ accounts over $2,500.
4. ___ other accounts.

B. Indicate on the line scale below the weight you have placed, in making your sample size judgments, on the compliance tests conducted by Tacket.

[Very Low Weight] [Moderate Weight] [Very High Weight]

C. Indicate on the line scale below the weight you have placed, in making your sample size judgments, on the sample size recommendations of Tacket.

[Very Low Weight] [Moderate Weight] [Very High Weight]
INFORMATION ABOUT TACKET:

EXPERIENCE - Little - Tacket has been a senior for only about a year. Erwin is his first wholesaling client.

EVALUATION - Low - Tacket has received only mediocre ratings in the firm's formalized performance evaluations. Although he is regarded as competent in administration and personnel matters, his technical skill, his care in carrying out audit work, and the soundness of his audit judgments are considered barely acceptable. Your own appraisal of the quality of his work supports these ratings.

SAMPLING APPROACH - Judgmental - In a memo, Tacket explains that he arrived at the sample sizes he has recommended, by adjusting last year's sample sizes to recognize the improvement he judged had taken place in the client's internal controls.

YOUR TASK:

A. Indicate below the sample sizes you would likely select based upon your professional judgment, considering the audit work conducted by Tacket and the impact of the above information about Tacket on your sample size judgments.

1. ___ accounts over 90 days past due.
2. ___ accounts past due between 45 and 90 days.
3. ___ accounts over $2,500.
4. ___ other accounts.

B. Indicate on the line scale below the weight you have placed, in making your sample size judgments, on the compliance tests conducted by Tacket.

C. Indicate on the line scale below the weight you have placed, in making your sample size judgments, on the sample size recommendations of Tacket.
INFORMATION ABOUT TACKET:

EXPERIENCE - Little - Tacket has been a senior for only about a year. Erwin is his first wholesaling client.

EVALUATION - High - Tacket has received excellent ratings in the firm's formalized performance evaluations. He is regarded as competent in supervising field work, adept at dealing with audit and client personnel, qualified in technical matters, careful in his audit work, and sound in his audit judgments. Your own appraisal of the quality of his work supports these ratings.

SAMPLING APPROACH - Judgmental - In a memo, Tacket explains that he arrived at the sample sizes he has recommended, by adjusting last year's sample sizes to recognize the improvement he judged had taken place in the client's internal controls.

YOUR TASK:

A. Indicate below the sample sizes you would likely select based upon your professional judgment, considering the audit work conducted by Tacket and the impact of the above information about Tacket on your sample size judgments.

1. ___ accounts over 90 days past due.
2. ___ accounts past due between 45 and 90 days.
3. ___ accounts over $2,500.
4. ___ other accounts.

B. Indicate on the line scale below the weight you have placed, in making your sample size judgments, on the compliance tests conducted by Tacket.

C. Indicate on the line scale below the weight you have placed, in making your sample size judgments, on the sample size recommendations of Tacket.
INFORMATION ABOUT TACKET:

EXPERIENCE - Little - Tacket has been a senior for only about a year. Erwin is his first wholesaling client.

EVALUATION - Low - Tacket has received only mediocre ratings in the firm's formalized performance evaluations. Although he is regarded as competent in administration and personnel matters, his technical skill, his care in carrying out audit work, and the soundness of his audit judgments are considered barely acceptable. Your own appraisal of the quality of his work supports these ratings.

SAMPLING APPROACH - Statistical - In a memo, Tacket explains that he derived the sample sizes he has recommended from sample size tables, having regard for the risk levels he judged acceptable.

YOUR TASK:

A. Indicate below the sample sizes you would likely select based upon your professional judgment, considering the audit work conducted by Tacket and the impact of the above information about Tacket on your sample size judgments.

1. ___ accounts over 90 days past due.
2. ___ accounts past due between 45 and 90 days.
3. ___ accounts over $2,500.
4. ___ other accounts.

B. Indicate on the line scale below the weight you have placed, in making your sample size judgments, on the compliance tests conducted by Tacket.

C. Indicate on the line scale below the weight you have placed, in making your sample size judgments, on the sample size recommendations of Tacket.
INFORMATION ABOUT TACKET:

EXPERIENCE - Extensive - Tacket has been a senior for four years. He has had extensive commercial experience with both small and large clients, including wholesalers.

EVALUATION - Low - Tacket has received only mediocre ratings in the firm's formalized performance evaluations. Although he is regarded as competent in administration and personnel matters, his technical skill, his care in carrying out audit work, and the soundness of his audit judgments are considered barely acceptable. Your own appraisal of the quality of his work supports these ratings.

SAMPLING APPROACH - Judgmental - In a memo, Tacket explains that he arrived at the sample sizes he has recommended, by adjusting last year's sample sizes to recognize the improvement he judged had taken place in the client's internal controls.

YOUR TASK:

A. Indicate below the sample sizes you would likely select based upon your professional judgment, considering the audit work conducted by Tacket and the impact of the above information about Tacket on your sample size judgments.

1. ___ accounts over 90 days past due.
2. ___ accounts past due between 45 and 90 days.
3. ___ accounts over $2,500.
4. ___ other accounts.

B. Indicate on the line scale below the weight you have placed, in making your sample size judgments, on the compliance tests conducted by Tacket.

C. Indicate on the line scale below the weight you have placed, in making your sample size judgments, on the sample size recommendations of Tacket.
SITUATION 7

INFORMATION ABOUT TACKET:

EXPERIENCE - Extensive - Tacket has been a senior for four years. He has had extensive commercial experience with both small and large clients, including wholesalers.

EVALUATION - Low - Tacket has received only mediocre ratings in the firm's formalized performance evaluations. Although he is regarded as competent in administration and personnel matters, his technical skill, his care in carrying out audit work, and the soundness of his audit judgments are considered barely acceptable. Your own appraisal of the quality of his work supports these ratings.

SAMPLING APPROACH - Statistical - In a memo, Tacket explains that he derived the sample sizes he has recommended from sample size tables, having regard for the risk levels he judged acceptable.

YOUR TASK:

A. Indicate below the sample sizes you would likely select based upon your professional judgment, considering the audit work conducted by Tacket and the impact of the above information about Tacket on your sample size judgments.

1. __ accounts over 90 days past due.
2. __ accounts past due between 45 and 90 days.
3. __ accounts over $2,500.
4. __ other accounts.

B. Indicate on the line scale below the weight you have placed, in making your sample size judgments, on the compliance tests conducted by Tacket.

C. Indicate on the line scale below the weight you have placed, in making your sample size judgments, on the sample size recommendations of Tacket.
INFORMATION ABOUT TACKET:

EXPERIENCE - Extensive - Tacket has been a senior for four years. He has had extensive commercial experience with both small and large clients, including wholesalers.

EVALUATION - High - Tacket has received excellent ratings in the firm's formalized performance evaluations. He is regarded as competent in supervising field work, adept at dealing with audit and client personnel, qualified in technical matters, careful in his audit work, and sound in his audit judgments. Your own appraisal of the quality of his work supports these ratings.

SAMPLING APPROACH - Judgmental - In a memo, Tacket explains that he arrived at the sample sizes he has recommended, by adjusting last year's sample sizes to recognize the improvement he judged had taken place in the client's internal controls.

YOUR TASK:

A. Indicate below the sample sizes you would likely select based upon your professional judgment, considering the audit work conducted by Tacket and the impact of the above information about Tacket on your sample size judgments.

1. ___ accounts over 90 days past due.
2. ___ accounts past due between 45 and 90 days.
3. ___ accounts over $2,500.
4. ___ other accounts.

B. Indicate on the line scale below the weight you have placed, in making your sample size judgments, on the compliance tests conducted by Tacket.

C. Indicate on the line scale below the weight you have placed, in making your sample size judgments, on the sample size recommendations of Tacket.
Post Experimental Questionnaire

1. Was the task understandable?

2. Are there any items of information about the senior which you believe should have been included in the casestudy but were not?

3. Do you have any other comments about the task?
BIBLIOGRAPHY


226.


Biddle, G. C. and E. J. Joyce, "Heuristics and Biases: Their Implications for Probabilistic Inference in Auditing," (in progress), University of Chicago.


Hovland, C. I. "Reconciling Conflicting Results Derived from Experiments and Survey Studies of Attitude Change," American Psychologist, 1959, 14, pp. 8-17.


