

THE EFFECT OF CONDITIONING EVIDENCE ON AUDITORS'  
EVIDENCE SUFFICIENCY ASSESSMENTS

by

James Joseph Donegan

---

Copyright © James Joseph Donegan 1995

A Dissertation Submitted to the Faculty of the  
COMMITTEE ON BUSINESS ADMINISTRATION

In Partial Fulfillment of the Requirements  
For the Degree of

DOCTOR OF PHILOSOPHY  
WITH A MAJOR IN MANAGEMENT

In the Graduate College

THE UNIVERSITY OF ARIZONA

1995

## INFORMATION TO USERS

This manuscript has been reproduced from the microfilm master. UMI films the text directly from the original or copy submitted. Thus, some thesis and dissertation copies are in typewriter face, while others may be from any type of computer printer.

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

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

Oversize materials (e.g., maps, drawings, charts) are reproduced by sectioning the original, beginning at the upper left-hand corner and continuing from left to right in equal sections with small overlaps. Each original is also photographed in one exposure and is included in reduced form at the back of the book.

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

# UMI

A Bell & Howell Information Company  
300 North Zeeb Road, Ann Arbor, MI 48106-1346 USA  
313/761-4700 800/521-0600



THE EFFECT OF CONDITIONING EVIDENCE ON AUDITORS'  
EVIDENCE SUFFICIENCY ASSESSMENTS

by

James Joseph Donegan

---

Copyright © James Joseph Donegan 1995

A Dissertation Submitted to the Faculty of the  
COMMITTEE ON BUSINESS ADMINISTRATION

In Partial Fulfillment of the Requirements  
For the Degree of

DOCTOR OF PHILOSOPHY  
WITH A MAJOR IN MANAGEMENT

In the Graduate College

THE UNIVERSITY OF ARIZONA

1995

UMI Number: 9603714

Copyright 1995 by  
Donegan, James Joseph  
All rights reserved.

---

UMI Microform 9603714  
Copyright 1995, by UMI Company. All rights reserved.

This microform edition is protected against unauthorized  
copying under Title 17, United States Code.

---

UMI

300 North Zeeb Road  
Ann Arbor, MI 48103

As members of the Final Examination Committee, we certify that we have

read the dissertation prepared by James Joseph Donegan

entitled THE EFFECT OF CONDITIONING EVIDENCE ON

AUDITORS' EVIDENCE SUFFICIENCY ASSESSMENTS

and recommend that it be accepted as fulfilling the dissertation  
requirement for the Degree of Doctor of Philosophy

William Waller  
William Waller

July 25, 1995  
Date

William L. Felix, Jr.  
William L. Felix, Jr.

July 26, 1995  
Date

Jeffrey W. Schatzberg  
Jeffrey W. Schatzberg

July 27, 1995  
Date

\_\_\_\_\_  
\_\_\_\_\_

\_\_\_\_\_  
Date

\_\_\_\_\_  
Date

Final approval and acceptance of this dissertation is contingent upon  
the candidate's submission of the final copy of the dissertation to the  
Graduate College.

I hereby certify that I have read this dissertation prepared under my  
direction and recommend that it be accepted as fulfilling the dissertation  
requirement.

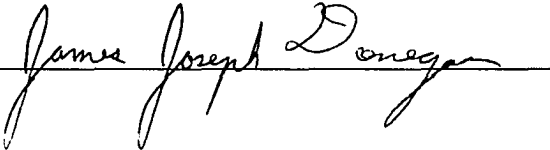
William L. Felix, Jr.  
Dissertation Director  
William L. Felix, Jr.

July 26, 1995  
Date

## STATEMENT BY AUTHOR

This dissertation has been submitted in partial fulfillment of requirements for an advanced degree at the University of Arizona and is deposited in the University Library to be made available to borrowers under rules of the library.

Brief quotations from this dissertation are allowable without special permission, provided that accurate acknowledgement of source is made. Requests for permission for extended quotation from or reproduction of this manuscript in whole or in part may be granted by the copyright holder.

SIGNED: A handwritten signature in cursive script, reading "James Joseph Donegan", is written over a horizontal line. The signature is positioned to the right of the word "SIGNED:".

## ACKNOWLEDGEMENTS

I wish to thank my dissertation chairman William Felix, Jr. who has provided intellectual stimulation and encouragement through the years. I also thank William Waller, Jeffrey Schatzberg, Travis Hirschi and Terence Connolly for their help as members of my committee. I owe thanks to the other faculty and doctoral students who provided valuable comments and advice. Finally, I wish to express my appreciation to Joan Wild, Ann Kersten, and Marion Levy.

**DEDICATION**

This work is dedicated to Michele W. Ganon, fellow doctoral student and wife.

## TABLE OF CONTENTS

<b>LIST OF FIGURES</b> .....	8
<b>LIST OF TABLES</b> .....	9
<b>ABSTRACT</b> .....	10
<b>CHAPTER 1 INTRODUCTION</b> .....	11
<b>1.1 Evidence Sufficiency Judgments</b> .....	12
<b>1.2 Prior Research</b> .....	13
<b>1.3 Contribution</b> .....	14
<b>CHAPTER 2 LITERATURE REVIEW</b> .....	17
<b>2.1 Second-Order Uncertainty</b> .....	17
<b>Subjective Expected Utility</b> .....	18
<b>Ellsberg</b> .....	19
<b>Yates and Zukowski</b> .....	20
<b>Curley, Yates and Abrams</b> .....	21
<b>Einhorn and Hogarth</b> .....	22
<b>Information Completeness and Second-Order Uncertainty</b> .....	23
<b>Sanctions and Second-Order Uncertainty</b> .....	25
<b>2.2 Revision of Second-Order Uncertainty</b> .....	26
<b>Evidence Sufficiency</b> .....	28
<b>2.3 Causes and Implications of Decreased Confidence</b> .....	29
<b>2.4 Auditor Assessment of First-Order and Second-Order</b>	
<b>Uncertainty</b> .....	32
<b>Morton and Felix</b> .....	33
<b>Messier and Quilliam</b> .....	33
<b>Bedard</b> .....	34
<b>Waller</b> .....	35
<b>Revision of Auditor's First-Order Judgments</b> .....	36
<b>2.5 Conclusion</b> .....	38
<b>CHAPTER 3 THEORY AND HYPOTHESES</b> .....	41
<b>3.1 The Waller and Felix (1984) Model of Opinion Formulation</b> .....	43
<b>3.2 Planning Risk and the Expected Sufficiency of Evidence</b> .....	45
<b>3.3 Evidence Consistency and the Revision of Audit Risk</b>	
<b>and Sufficiency Assessments</b> .....	47
<b>3.4 Hypotheses</b> .....	53
<b>CHAPTER 4 RESEARCH DESIGN</b> .....	57
<b>4.1 Subjects</b> .....	59
<b>4.2 Procedure</b> .....	60
<b>4.3 Task and Materials</b> .....	61
<b>4.4 Conclusion</b> .....	65

**TABLE OF CONTENTS-Continued**

<b>CHAPTER 5 DATA ANALYSIS AND DISCUSSION</b> .....	66
<b>5.1 Preliminary Analysis</b> .....	66
Posterior Assessments of Bel(R) and Bel(S) .....	69
<b>5.2 Experimental Results</b> .....	72
Test of Risk Revision .....	73
Test of Sufficiency Revision .....	74
Joint Test of Audit Risk and Evidence Sufficiency Revision .....	76
<b>5.3 Other Results</b> .....	78
Additional Tests of Audit Risk and Evidence Sufficiency Revision ..	78
Analysis of Flowchart .....	82
<b>5.4 Manipulation Effectiveness</b> .....	85
<b>5.5 Demographic Variables</b> .....	88
Homogeneity of Subject Attributes Across Treatment Conditions ...	89
Association Between Demographic Attributes and Dependent Variables .....	90
Summary of Analysis of Demographic Variables .....	93
<b>5.6 Discussion</b> .....	93
Summary of Results .....	94
Limitations and Implications .....	96
Conclusion .....	99
<b>APPENDIX A: RESEARCH INSTRUMENT</b> .....	100
<b>REFERENCES</b> .....	107

**LIST OF FIGURES**

Figure 1, Effect of the Consistency of Conditioning Evidence on Bel (R) and Bel (S) .....	50
Figure 2, Experimental Design .....	58
Figure 3, Comparison of Actual and Hypothetically Adjusted Means for Sufficiency Revision .....	82

### LIST OF TABLES

Table 1, Elements of the Waller and Felix Model . . . . .	44
Table 2, Effect of the Consistency of Conditioning Evidence on Revision of Bel(R) and Bel(S) . . . . .	54
Table 3, Summary Statistics for Initial Assessments of Risk and Sufficiency . . . . .	68
Table 4, Correlations Between Initial Assessments of Risk and Sufficiency . . . . .	69
Table 5, Summary Statistics for Posterior Assessments of Risk and Sufficiency . . . . .	70
Table 6, Correlations Between Posterior Assessments of Risk and Sufficiency . . . . .	71
Table 7, Sign of Risk Revision by Cell . . . . .	74
Table 8, Sign of Sufficiency Revision by Cell . . . . .	75
Table 9, Joint Test of Hypotheses 1 and 2 . . . . .	77
Table 10, Mean Risk Revision by Cell . . . . .	79
Table 11, Mean Sufficiency Revision by Cell . . . . .	80
Table 12, Expectation of Increased Sample Size by Cell . . . . .	83
Table 13, Association Between PRIOR RISK and Exit Questions . . . . .	86
Table 14, Distribution of Demographic Variables Across Treatments . . . . .	89
Table 15, Correlations between Demographic and Dependent Variables . . . . .	90
Table 16, Decomposition of Risk Assessments by Gender . . . . .	93

### ABSTRACT

The auditor's report (AICPA 1988) generally contains two statements of belief. The first is an opinion as to whether the financial statements examined are free of material misstatement. The second is an assertion that the audit provides a reasonable basis for the opinion expressed. If the auditor judges the examination to have been inadequate, then the report should either be suitably qualified or no opinion offered. Evaluating the sufficiency of evidence obtained is thus an important part of the auditor's task.

Considerable research exists describing auditor knowledge of error frequencies and causes, as well as the cognitive factors that affect error judgments. In comparison, less is known about evidence sufficiency judgments and their impact on the opinion formulation process. This monograph: (1) further examines the relationship between audit risk and evidence sufficiency judgments first hypothesized by Waller and Felix (1984); and (2) reports an experiment which investigates the effect of consistent and inconsistent conditioning evidence on the revision of evidence sufficiency assessments. The results suggest that conditioning evidence increases sufficiency assessments, except when unexpected audit differences are discovered.

**CHAPTER 1****INTRODUCTION**

The auditor's report (AICPA 1988) generally contains two statements of belief. The first is an opinion as to whether the financial statements examined are free of material misstatement. The second is an assertion that the audit provides a reasonable basis for the opinion expressed. If the auditor judges the examination to have been inadequate, then the report should either be suitably qualified or no opinion offered. Evaluating the sufficiency of evidence obtained is thus an important part of the auditor's task.

Considerable research exists describing auditor knowledge of error frequencies and causes, as well as the cognitive factors that affect error judgments. In comparison, less is known about evidence sufficiency judgments and their impact on the opinion formulation process. This monograph: (1) further examines the relationship between audit risk and evidence sufficiency judgments first hypothesized by Waller and Felix (1984); and (2) reports an experiment which investigates the effect of consistent and inconsistent conditioning evidence on the revision of evidence sufficiency assessments. The results suggest that conditioning evidence increases sufficiency assessments, except when unexpected audit differences are discovered.<sup>1</sup>

---

<sup>1</sup>The term "differences" rather than "errors" is used since not all discrepancies between auditor expectations and test results are due to misstatements. Some results that initially seem to indicate misstatement will ultimately be attributed to nonerror causes.

### **Evidence Sufficiency Judgments**

Evidence sufficiency judgments are made for two reasons. First, deciding the amount and type of evidence that should be collected is an important element in forming a correct opinion (AU-326). Second, although the auditor attests to the fairness of the financial statements, the report warrants the adequacy of the examination, not the correctness of the opinion. When a report has been issued which appears, ex post, to have been incorrect, the auditor may avoid sanctions by demonstrating that sufficient competent evidential matter was obtained. Consideration of evidence sufficiency may both improve audit effectiveness and allow the auditor to assess the justifiability of the opinion.

Evidence sufficiency judgments are made throughout the engagement. Client acceptance or continuation is predicated on the probable existence of adequate evidence on which to base an opinion. In developing the audit program, the auditor selects the specific procedures that he believes will reduce the interrelated risks of issuing an unjustified or incorrect opinion to acceptably low levels. As procedures are performed, the auditor considers whether the evidence obtained is adequate to justify the conclusions reached concerning specific audit objectives. After fieldwork has been completed, a final evaluation is made as to whether sufficient evidence has been obtained to support an overall opinion.

If the results obtained during fieldwork confirm the risk assessments made during planning, the audit program may be executed without revision and the auditor is likely to believe that sufficient evidence has been obtained to justify the opinion. SAS-53, **The Auditor's Responsibility to Detect and Report Material Errors and Irregularities**

(AICPA 1988), requires that when unexpected results are obtained the auditor consider the causes of such outcomes and evaluate the implications for the audit as a whole. Previously obtained evidence may appear less persuasive than originally believed and additional procedures may be necessary to resolve the uncertainty. In contrast, expected differences should not induce reconsideration of the audit plan, since their detection serves only to confirm prior belief.

### **Prior Research**

Evidence sufficiency assessments are a type of second-order uncertainty judgment. Cognitive research has demonstrated that choices or likelihood assessments made under conditions of significantly incomplete information (ambiguity) often violate the axioms of subjective probability (Ellsberg 1961, Yates and Zukowski 1976, Curley, Yates and Abrams 1986, Einhorn and Hogarth 1985a). Risky choice appears to be affected not only by preferences and beliefs, but also by second-order judgments as to the completeness of the information on which first-order beliefs are based. Perceptions of second-order uncertainty may cause ambiguity averse cognizers to postpone choice and extend the search for decision relevant information.

According to Frisch and Baron (1988), second-order uncertainty judgments are intended primarily to improve decision-quality. Other researchers believe that ambiguity aversion is greater when choices may have to be justified to others (Brown 1990, Ashton 1992). The degree of second-order uncertainty experienced has been reported to depend

on the amount, credibility and consistency of evidence (Einhorn and Hogarth 1985a); the perceived possibility of social sanctions as a result of incorrect choice (Curley, Yates and Abrams 1986); and the ability of cognizers to recognize the omission of task relevant information (Waller 1990, Sanbonmatsu, Karder, and Herr 1992).

Waller and Felix (1984) introduced the concept of second-order uncertainty to the audit literature, asserting that opinion justification is a critical aspect of the audit task. Two recent studies have investigated factors affecting the development and progression of second-order judgments in the audit environment. Waller (1990) has found that confidence in control risk judgments is affected by task knowledge, information completeness, and demand for task precision. Morton and Felix (1990) presented evidence suggesting that sufficiency judgments are increased as credible audit evidence accumulates.

### **Contribution**

This study extends the cognitive auditing literature in two ways. Primarily, it explicates and tests Waller and Felix's (1984) assertion that evidence drives audit risk and evidence sufficiency judgments "usually jointly but at times severally." The findings suggest that the consistency of conditioning evidence affects the revision of both first-order and second-order judgments when initial evidence implies fair presentation. However, only audit risk revision was found to be influenced by consistency when initial evidence implied material misstatement. Secondly, arguments are developed and preliminary evidence provided suggesting that the decision to begin diagnosis of unusual financial statement

symptoms (Einhorn 1976, Libby 1985) may be triggered by increases in second-order uncertainty.

The results also contribute to the cognitive psychology literature. Empirical research has focused on the degree of second-order uncertainty experienced at a single point in time. Except for the theory and evidence provided by Waller and Mitchell (1991), little is known about the revision of second-order uncertainty as conditioning evidence is evaluated. An unresolved issue identified by Waller and Mitchell is whether second-order uncertainty decreases monotonically as evidence is collected, or whether it may at times increase as new information is acquired. The present research suggests that conditioning information inconsistent with the implications of prior evidence can induce increased perceptions of second-order uncertainty.

Finally, this study is of potential benefit to practice units seeking to increase the likelihood that error findings will be adequately investigated. The "expectation gap" standards were a response to Congressional pressure for greater auditor effectiveness and accountability. Alleged audit failures have often revealed that audit teams encountered unexpected findings which could have been explained by plausible error hypotheses which were either not considered or prematurely discarded (eg., Accounting Series Releases 153, 157, and 173, SEC 1977). Failure to expand testing and correctly diagnose the cause of unexpected results suggest that auditors may not always revise sufficiency requirements to reflect current knowledge. Obtaining further understanding of evidence sufficiency

judgments may aid firms in better assessing potential liability and perhaps improve audit practice.

The remainder of the paper is organized as follows. Chapter 2 reviews relevant literature from the domains of psychology and auditing. Chapter 3 presents the Waller & Felix (1984) model and develops the substantive hypotheses. Chapter 4 describes the research design and experimental procedures. Chapter 5 reports the empirical results and discusses the implications and limitations of the research.

## **CHAPTER 2                      LITERATURE REVIEW**

This chapter reviews the literature relevant to investigating auditors' evidence sufficiency judgments. Section 1 discusses perceptions of second-order uncertainty at a single point in time. Section 2 considers revision of second-order uncertainty judgments as conditioning events are experienced. Section 3 focuses on initial responses to inconsistent conditioning evidence. Section 4 reviews the relevant cognitive auditing research. The chapter concludes with a brief summarization of the literature reviewed.

### **2.1                                      Second-Order Uncertainty**

The existence, effects, and causes of uncertain probabilities are topics of dispute in a broad spectrum of literatures. Economists and normative decision theorists are concerned with implications as to the descriptive and prescriptive validity of the subjective expected utility (SEU) model, as well as possible axiomatic representations of second-order judgments. Psychologists have investigated factors that affect uncertainty about probabilities and developed cognitive accounts of its role in decision-making. Finally, some epistemologists and psychologists have attempted to link second-order uncertainty to methods of formally assessing systems of belief justification. This section briefly summarizes SEU theory and then reviews theory and evidence concerning the relationship between contextual ambiguity, second-order uncertainty and risky choice.

### **Subjective Expected Utility**

In order for SEU theory (Savage 1954/1972) to be an optimal model for decision-making: (1) cognizers' preferences must be uniquely defined up to a ratio transformation across all possible pairs of choices and events; and (2) subjective probabilities concerning the occurrence of events must be complete, nonnegative and additive. The expected utility associated with each choice-event pair is the product  $U(E_i)P(E_i|C_j)$ , where  $U(E_i)$  is the utility associated with an event  $E$  and  $P(E_i|C_j)$  represents the conditional probability that  $E$  will occur, given that action  $j$  was selected. The sum of the expected utilities associated with each action are then calculated and the choice which maximizes utility selected.

A necessary condition for the descriptive or prescriptive validity of the SEU model is the existence of beliefs that satisfy the axioms of probability. According to Savage, a personal probability takes into account all of the evidence available to a cognizer when the judgment is made. Variations in the reliability or relevance of evidence are impounded in a single assessment. The subjective probability of rain today is the only measure of risk needed to decide whether or not to carry an umbrella. Separate second-order assessments, such as an expression of confidence in the rain judgment, are not meaningful. But, as subsequent research has shown, contexts exist in which probabilities cannot be inferred from observed choices. Considerable evidence exists to indicate that cognizers perceive some probability judgments as less certain than others.

## Ellsberg

Ellsberg (1961) provided the first evidence that belief about an uncertain event may not be fully described by a single probability judgment. He reported a study in which subjects indicated their preferences for a series of gambles requiring the correct specification of the color of a ball to be drawn from a hypothetical urn. Urn I contained an unknown proportion of red and black balls totalling 100, whereas Urn II contained an equal mix. Applying the principle of insufficient reason to events involving Urn I (Lindgren 1976) a probability of .5 should have been assigned to the selection of red or black from either urn. Subjects were sequentially offered choices between 4 pairs of gambles: (1) Red<sub>I</sub> or Black<sub>I</sub>, (2) Red<sub>II</sub> or Black<sub>II</sub>, (3) Red<sub>I</sub> or Red<sub>II</sub>, (4) Black<sub>I</sub> or Black<sub>II</sub>. Most respondents were indifferent to the choices in gambles 1 and 2, but preferred Red<sub>II</sub> in gamble 3 and Black<sub>II</sub> in gamble 4. Since it is inconsistent with the additivity axiom to prefer Red<sub>II</sub> to Red<sub>I</sub> and  $\sim$ Red<sub>II</sub> to  $\sim$ Red<sub>I</sub>, it follows that probabilities are not inferable from these choices.

Ellsberg ascribed his paradoxical results to differential knowledge of the event generating mechanisms for gambles 3 and 4. Even if probability estimates for all outcomes were judged 50%, those involving Urn I were based on ignorance of the underlying distribution whereas those involving Urn II were based on direct confirmation.<sup>2</sup> Cognizers

---

<sup>2</sup>The range of possible distributions (100 for Urn I, 1 for Urn II) was hypothesized to be an element of uncertainty, but Ellsberg also speculated that respondents were distrustful of the description of Urn I. Real world gambles involving mechanisms similar to Urn I are often "sucker" bets and subjects may have avoided them on such grounds.

were hypothesized to be averse to actions when the occurrence of the preferred state depended on an imperfectly understood process. According to Ellsberg:

What is at issue might be called the [relative] **ambiguity** of this information, a quality depending on the amount, type, reliability, and "unanimity" of information, and giving rise to one's degree of "confidence" in an estimate of relative likelihoods.

Ellsberg's analysis raises issue as to whether subjects who embrace the paradox (including Savage) violate the Savage axioms due to cognitive limitations or because conformity does not maximize their utility under conditions of ambiguity. Subsequent research has centered on two questions: what are the contextual and cognitive determinants of ambiguity and (when) is ambiguity aversion a normatively appropriate response?

### **Yates and Zukowski**

Yates and Zukowski (1976) report an experiment which investigated whether ambiguity aversion is a function of the range of possible values of distributional parameters. Subjects were asked to: (1) choose between a single pair of gambles; (2) designate the color of the winning chip; (3) play the selected gamble. Two gambles corresponded to Ellsberg's observable  $U_{m_{II}}$  and unobservable  $U_{m_I}$ , except that they used ten poker chips (red and blue) mixed in a book bag. A third gamble consisted of: (a) a preliminary draw from a uniform distribution to determine the number of winning chips placed in the bag (0-10) and (b) the actual drawing. The expectation of drawing a red or blue chip was .5 for all three gambles, which differed only in the range of possible values for parameter  $p$ .

According to the range hypothesis: (1) the simple lottery would have no ambiguity since  $p$  was constant; (2) the compound lottery would have the maximal ambiguity for a 10 chip draw, since  $p$  could assume 11 values; and (3) the unobservable lottery might be perceived as less ambiguous than the complex lottery, as the range for  $p$  might be less than 11. Contrary to their hypothesis, Yates and Zukowski found that preferences for the simple (range=1) and complex (range=11) lotteries did not significantly differ, but that both were strongly preferred to the unobservable alternative (range  $\leq$  11). They conclude that, "The results provide clear evidence against the conclusion that aversive ambiguity is completely reducible to the ranges of subjective second order probability distributions."

### **Curley, Yates and Abrams**

Curley, Yates and Abrams (1986) report six experiments in which subjects were offered a choice between lotteries similar to Ellsberg's Urn I (unknown composition) and Urn II (equal mix). Each experiment measured whether the proportion of subjects who were averse to the ambiguous lottery was affected by a possible cause of ambiguity aversion. The explanations tested included risk aversion, regret minimization, concern that the lottery was actually a hostile 2-person game, and avoidance of social sanctions. The general level of avoidance was high, with selection of the unambiguous lottery averaging 70% across the experiments. Support was found only for the social sanction hypothesis. When the lottery was to be played before an audience and its composition publicly revealed ex post, less than 20% of the subjects selected the alternative with the unknown proportions. Subjects' risk

preferences did not affect ambiguity aversion, lending additional support to the distinction between first-order and second-order judgments.

### **Einhorn and Hogarth**

Einhorn and Hogarth (1985a) proposed and tested a descriptive model of choice which explains second-order uncertainty in terms of the characteristics of the evidence upon which judgment is based. They agreed with Yates and Zukowski (1976) that ambiguity is inversely related to the number of distributions that can be excluded before making a choice. However to replace the simple range hypothesis they proposed that the amount, credibility and consistency of evidence affects confidence in the estimate of parameter  $p$ .

Einhorn and Hogarth (1985a) hypothesize that choices are made using an anchoring and adjustment strategy in which the anchor is provided by an initial probability judgment based on the consistency of the available evidence. For example, if eight of ten witnesses to an accident agreed that the offending vehicle was green, the anchor would be .8. Because assessments of  $p$  may be incorrect, cognizers perform mental simulations to determine the likelihood of other parameters. These simulations are influenced by the amount and credibility of evidence available to estimate  $p$ . Credibility is determined by the competence of the source and the feature similarity of  $E$  and  $\sim E$ . Also, subject optimism/pessimism may combine with the perceived favorability of the most likely outcome to influence judgment.

One reported experiment manipulated evidence amount and consistency. Subjects assessed the likelihood that a taxi cab which caused an accident was green, after being

informed that there were  $k$  witnesses who reported that the taxi was green, and  $n-k$  that it was blue. Mean assessments were close to the expected value of  $k/n$  when  $n$  was large. However, estimates tended to regress towards .5 when  $n$  was small, reflecting an apparent aversion to reporting extreme probabilities based on small samples. Because they could not report both probabilities and confidence, subjects may have combined both judgments by estimating a confidence interval and reporting the conservative bound.

A second experiment varied the credibility of evidence sources, as well as the similarity of the events being reported by observers. Einhorn and Hogarth (1985) expected to find that reported observations would be perceived as less credible when outcomes ( $E$  or  $\sim E$ ) were difficult to distinguish. In one scenario subjects were informed that a word recognition task had been performed either by a fifteen year old or six year old child (more credible vs less credible witness) who had distinguished between either similar (BED or BID) or dissimilar (BED or ROT) word-pairs. Sources low in credibility and easily confusable events were expected to increase ambiguity. Likelihood estimates were affected by perceived differences in source credibility, but not by the degree of feature similarity.

### **Information Completeness and Second-Order Uncertainty**

The perceived completeness of information has been suggested as the general factor responsible for perceptions of second-order uncertainty. Frisch and Baron (1988) in their review of empirical research found incomplete information to be the common denominator across empirical investigations. They define ambiguity as "the subjective experience of

missing information relevant to a prediction" and assert that obtaining more evidence is a sound heuristic for improving decision-making whenever second-order uncertainty is high.

Similarly, Smith, Benson, and Curley (1991) have, concluded that:

First-order probability assessments are often based on information that is critically incomplete, gaps in related arguments having been filled in by assumptions. Second-order assessments highlight these assumptions by calling first-order probabilities into question....A key issue for probability assessment is the interface between the reasoning that produces a belief and the judgmental assessment that qualifies it.

Judgments of information completeness compare the present evidence set with that which a cognizer believes should be available before making a decision. As such, differential perceptions of completeness are likely to be affected by task experience and expertise. Two recent studies have investigated the relationship between information completeness and experience. Waller (1990) reports that experienced auditors were more confident in their risk assessments than journeymen when presented with a relatively complete data set.<sup>3</sup> Deletion of task relevant data decreased the confidence only of the experienced subjects. Sanbonmatsu, Karder and Herr (1992) reported a study investigating the effect of incomplete information on the confidence of highly knowledgeable, moderately knowledgeable and ignorant subjects. When evaluating the quality of a bicycle using an incomplete data set, highly knowledgeable judges were found to be more neutral in their evaluations and lower in expressed confidence, as compared to moderately knowledgeable

---

<sup>3</sup>Section 2.3 provides a more detailed discussion of Waller (1990).

judges. In both studies, the absence of relevant information apparently decreased confidence only for subjects with sufficient knowledge of the judgment task to identify the omissions.

### **Sanctions and Second-Order Uncertainty**

Some researchers have proposed that ambiguity aversion is intensified when choices must be justified to third-parties (Curley, Yates, and Abrams 1986, Baron and Hershey 1988, Brown 1990). In this view second-order uncertainty measures the likelihood of sanctions being applied to a decision-maker who acts without warrant. Brown (1990) adopts the extreme position, arguing that ambiguity affects decisions only when third-party justification is required. Frisch and Baron (1988) have taken a more moderate stance, stating that:

The issue of blame, responsibility and regret are more salient in situations of ambiguity. In particular, a person may be blamed (rightly or wrongly, depending on whether the missing information could have been usefully obtained) for not obtaining the missing information. This also argues against not making a choice.

As noted supra, Curley, Yates and Abrams (1986) found that ambiguity aversion increased when gambles selected by subjects were played out in front of an audience and Ashton (1992) has reported that the demand for task justification increased judgment accuracy. Otherwise little is known about the extent to which measures of second-order uncertainty or confidence are affected by aversion to social or economic sanctions.

## 2.2 Revision of Second-Order Uncertainty

Waller and Mitchell (1991) provide an account of belief revision which reconciles the Bayesian model (and by extension the Savage axioms) with the existence of second-order uncertainty. They describe second-order uncertainty as a cognitive state that arises due to differences between imagining and experiencing an event  $E$ . They consider the following situation. A cognizer at point  $t$  is required to provide an assessment of  $P_t(E_t|C_t)$ , where  $C_t$  is some conditioning event that will, or will not, occur after the initial assessment, but before  $E \vee \sim E$  is observed at time  $t$ . Application of Bayes theorem leads to the requirement that  $P_t(E_t|C_t) = P_t(E_t)$ , should  $C$  occur. If other events are foreseen which may conjoin with  $C$  to affect  $P_t(E_t)$ , Bayesian conditionalization requires that they be accounted for in the initial estimate, which then becomes  $P_t(E_t|C_t, \dots, n_t)$ . If the association between  $E$  and  $C$  that is imagined at time  $t$  is unchanged at  $t'$ , this relation will hold. If experiencing  $C$  is different than imagining it, then  $P_t(E_t|C_t) \neq P_t(E_t)$ .

Individuals may also believe that their models of the event generating process are incomplete and that "extraconditional" information relevant to assessing  $P_t(E_t)$  may be observed "without being able to specify ex ante the exact conditions to which the information will pertain" (Waller and Mitchell 1991). Rather than adjusting  $P_t(E_t|C_t)$  to account for currently unknown data that may be incorporated in  $P_t(E_t|C_t, \dots)$ , the assessment at time  $t$  may be based solely on current knowledge and a separate confidence judgment made to reflect the extent to which  $C_t$  is believed to be a full instantiation of  $P_t(E_t|.)$ .

According to this account, equal judgments, based on more or less complete probability schema, will differ in the extent to which they are predictive of future belief.

Waller and Mitchell hypothesize that: (1) when extraconditional evidence is experienced in conjunction with  $C_t$ ,  $P_t(E_{t'})$  will differ from  $P_t(E_{t'}|C_t)$ ; (2) experiencing extraconditional evidence generally increases confidence; and (3) the greater confidence at time  $t$ , the less the impact of extraconditioning evidence on  $P_t(E_{t'})$ . In addition they suggest that the affective state induced by  $C \vee \sim C$  will increase or decrease  $P_t(E_{t'})$  relative to  $P_t(E_{t'}|C_t)$ . The results of their empirical investigation generally support the hypotheses, however some doubt remained as to whether increases in confidence between  $t$  and  $t'$  were attributable to extraconditional evidence or to the observation of  $C_t$ .

An issue raised, but not addressed, by Waller and Mitchell (1991) is whether confidence judgments monotonically increase as evidence is collected or whether new information can lower confidence. The research reviewed in this section has identified three attributes of evidence (quantity, credibility, and consistency) that may combine to determine confidence at a moment in time. In order for confidence to decline, the contribution made by at least one of these factors must be perceived as less at time  $t'$  than at  $t$ . Evidence quantity cannot be less at  $t'$  than at  $t$  and may be dismissed as a potential cause of decreased confidence. Confidence could be decreased if the credibility of prior evidence is impeached between  $t$  and  $t'$ , or if the implications of conditioning evidence reduced the overall consistency of an information set at time  $t$ , as compared to  $t'$ . The latter presumably

common occurrence is the focus of the present study, since it may explain how evidence jointly affects audit risk and evidence sufficiency judgments.

### **Evidence Consistency**

When making a single judgment based on simultaneously evaluated information, the second-order uncertainty literature has used "consistency" to denote the extent to which a set of evidence uniformly points to the occurrence of a single state. For contexts in which the judgment task is simply to predict  $E$  v  $\sim E$ , as in Einhorn and Hogarth (1985a), two items of evidence are consistent if each increases the judged likelihood that the same state will occur. Consistency, as it applies to second-order judgments, has not been examined in a belief revision context.

In investigating revision of first-order judgments, the term "consistent" has been applied to a set of conditioning information with uniform implications (evaluated sequentially or simultaneously), and contrasted with evidence described as mixed or conflicting (Einhorn and Hogarth 1985b, Ashton and Ashton 1988). In the same studies, "confirmatory" denoted conditioning evidence that supported the same belief as a prior probability assessment. These distinctions may have resulted partially from simplifying assumptions in the model being tested. The initial probability anchors were provided by the researchers and did not represent subject's beliefs based on prior evidence. Subsequent information could confirm or disconfirm the inherited anchor, but the design did not permit a consistency comparison between prior and conditioning evidence. In contrast, the

sequences of conditioning evidence could be characterized both as to their mutual consistency and as to whether they confirmed or disconfirmed the initial anchor. When initial probabilities denote cognizers' actual evaluation of evidence, the distinction between confirmation and consistency may be less important. Consistent initial evidence is likely to result in extreme likelihood judgments. For example, subjects in Einhorn and Hogarth (1985a) who reported probability judgments greater than .85 (less than .15) generally had evaluated evidence that was uniformly consistent with the occurrence (nonoccurrence) of the target event. Those who received conflicting information reported more moderate judgments. In a belief revision task, the consistency between the implications of initial and conditional evidence may closely parallel the agreement between prior and posterior belief. The degree to which conditional evidence confirms prior belief may proxy for a direct assessment of the consistency between the initial and conditioning evidence.

### **2.3 Causes and Implications of Decreased Confidence**

For confidence to increase as information is experienced,  $P(E|.)$  should be perceived as more complete at time  $t'$ , when compared with  $t$ . If inconsistent conditioning evidence causes a decrease in confidence, the completeness of  $P(E|.)$ , even after full instantiation, may be in doubt. Beach, Smith, Lundell, and Mitchell (1988) have posited that, decision-makers consider "whether the projected image [anticipated states] are sufficiently compatible with the trajectory image [goals] to warrant continuing with the status quo. Plans are altered only when the projected image clearly fails to reach the intended goal. Since confidence

should normally increase as evidence is collected, decreases suggest that the projected image may no longer be targeted towards an acceptably low level of second-order uncertainty by the time a decision must be made.

Decreases in confidence may cause suspension of the existing information search plan and prompt activities intended to diagnosis the cause of the decline. Although the second-order uncertainty literature has not addressed the effects of decreased confidence, some insight into possible implications can be obtained from research concerned with responses to unexpected or anomalous events. Since conditioning evidence that reduces consistency will generally disconfirm prior belief, it is plausible that responses to both phenomena are similar.

According to Piaget (1952), events whose observation disconfirms prior belief require the restructuring of cognitive schema or accommodation in order to be understood. Similarly, Einhorn and Hogarth (1986) assert that causal reasoning is precipitated by the perception of a difference in a field or background. Causal hypotheses are constructed to explain "events that are unusual, abnormal, or unlikely... one rarely seeks the cause of why one feels `average.'" (Einhorn and Hogarth 1986). However if the context is such that one would expect to feel abnormal, such as after a car crash, normal behavior may require explanation.

Question asking may be the first response to unexpected or anomalous events and has been accorded a central role in diverse cognitive processes such as knowledge acquisition, creativity and self-control (Graessner and McMahan 1993). Graessner and

McMahan report a study in which subjects read and solved quantitative problems taken from textbooks. Subjects believed that the purpose of the experiment was to test their problem-solving abilities. Questions were presented either in complete form or with the following changes: (1) deletion of information necessary to solve the problem; (2) addition of a contradictory statement of fact; (3) addition of subtly irrelevant information which blended with the problem solving context; and (4) addition of saliently irrelevant information.

In one experiment the dependent variable was the number of questions subjects generated while solving the problem. The greatest number of questions occurred in the deletion and salient irrelevancy versions. Unexpectedly, the contradictory version did not generate significantly greater questions than the original. In a second experiment subjects gave each passage an anomaly detection rating. Passages containing deletions or contradictions were perceived to be more anomalous than either the unaltered version or those passages containing subtle or salient irrelevancies.

Pyszynski and Greenberg (1981) reported an experiment that tested whether interest in obtaining information useful for explaining observed behavior was affected by whether the behavior was expected or anomalous. Subjects were told that they would be observing the behavior of a second participant (actually a confederate) in a variety of situations. The participants were privately informed that the experimenter would ask the confederate to perform either a small or large favor related to the study. A manipulation check revealed that subjects thought it very likely that the small favor would be acceded to,

and very unlikely that the large favor would be performed. The request was made and the confederate either complied or refused. When given a subsequent opportunity to obtain information about the confederate that was either "interesting" or relevant to forming an explanation for the observed behavior, subjects whose expectations had been confounded were more likely to choose the latter.

Graessner and McMahan (1993) and Psyszczynski and Greenberg (1981) both report that cognizers initial response to inconsistent outcomes is to seek information relevant to explaining the cause of the anomaly.<sup>4</sup> Viewed from a second-order uncertainty perspective, the results suggest that the anomalous outcomes may have lead to lower subject confidence in the schema  $P(E|.)$ . Asking questions and reviewing causally relevant information may have been intended to restructure the schema and restore confidence in judgment accuracy. Therefore, decreased confidence may be the cognitive phenomenon that prompts causal diagnosis.

#### **2.4 Auditor Assessment of First-Order and Second-Order Uncertainty**

This section reviews primarily the empirical auditing research investigating second-order uncertainty. Discussion of Waller and Felix (1984) has been deferred to Chapter 3. The section concludes with a brief summary of the audit risk revision literature relevant to the present research.

---

<sup>4</sup>See Nisbett and Ross (1980) or Einhorn and Hogarth (1986) for reviews of the extensive literature on causal judgments.

**Morton and Felix**

Morton and Felix (1990) investigated changes in second-order uncertainty as part of a study that tested the effects of SAS-55 (1988) on control risk assessments. Prior to SAS-55 the professional standards differentiated between: (1) subjective assessments of control risk and; (2) the degree to which such beliefs were supported by evidence. Only well-justified beliefs could be relied upon to reduce substantive testing. As revised, the standards eliminate the distinction between subjective risk assessment and justified reliance by mandating that the assessment of control risk be no lower than that which is adequately supported by evidence.

The experiment tested whether auditors conform to SAS-55 by assessing control risk no lower than the level they believe to be completely supported by evidence. Contrary to the standard, they found that auditors reduced control risk assessments even when evidence sufficiency was assessed at far less than maximum. Of relevance to the present study, their results show that assessments of evidence sufficiency are increased by the evaluation of consistent credible evidence. Subjects who first read a scenario describing an apparently effective control structure and then received test results consistent with effective controls recorded higher evidence sufficiency ratings than those who only read the scenario.

**Messier and Quilliam**

Messier and Quilliam (1992) rely on Beach and Mitchell's (1978) contingency theory to argue that greater accountability leads to the invocation of more complex decision

models. They also review avoidance tactics that may be employed depending on the source of possible sanctions (other auditors, clients, third-parties). Although not directly addressed by the present research, a contingency approach to audit strategy is compatible with the arguments that are developed in Chapter 3.

### **Bedard**

Bedard (1992) tested the Einhorn and Hogarth (1985a) ambiguity model using an audit judgment task. Subjects were informed that controls were apparently adequate for the sales cycle and that a test of controls was to be performed to determine the extent to which the confirmation of receivables could be restricted. The subjects were asked to provide a preliminary control risk judgment (which was not used to assess ambiguity), were given the results of the control test, and then asked to make a final assessment.

As with Einhorn and Hogarth (1985a), inferences concerning the extent of ambiguity were based on differences between probability assessments and an assumed probability anchor. For example, if the case stated that an assistant auditor performing a control test found 2 deviations in a sample of 50, the anchor for control risk was presumed to be .04. The amount of evidence was manipulated by changing the sample size from 10 to 80. Source credibility was varied by supplying information which indicated that the assistant was either highly competent or inept and careless.

The results obtained did not generally support the validity of the Einhorn and Hogarth (1985a) model. As predicted, sample size was inversely related to ambiguity when

source credibility was high. However, when source credibility was low, high error rates and large samples interacted to increase ambiguity. Also contrary to the results reported by Einhorn and Hogarth (1985a), ambiguity was greater when the evidence source was more credible.<sup>5</sup>

Only limited reliance should be placed on Bedard's findings. According to the Einhorn and Hogarth (1985a) model, the anchor in the assessment would be deviations/sample size. However, an initial judgment of control risk was elicited by Bedard prior to reviewing the outcome of the control test so that an alternative explanatory model could be tested. The result may have been to inadvertently convert the experiment into a belief revision task, instead of the single judgment process described by Einhorn and Hogarth (1985a).

### **Waller**

Waller (1990) investigates the effects of experience, demand for task precision and information completeness on the confidence judgments of auditors. Inexperienced auditors can be expected to have relatively low confidence in their risk judgments. On the other hand, experienced auditors should be more confident as long as most of the evidence usually relied on is available. It follows that a significantly incomplete evidence set should lower the confidence of experienced auditors more than inexperienced auditors, since only the

---

<sup>5</sup>A second experiment in which students performed a different accounting task provided results consistent with Einhorn and Hogarth (1985a).

former group would be aware of the importance of the omission. Finally, Waller predicts that the demand for task precision will be inversely associated with second-order uncertainty. The risk assessment practices of accounting firms generally require the use of ordinal scales using such terms as low, medium, and high. Because auditors are, in effect, able to specify a probability range rather than a point-estimate, they may be more confident when rating risk on ordinal as compared to continuous scales.

Waller's (1990) results corroborate the hypotheses, indicating that auditor confidence varies with knowledge of the decision environment and the demand for task precision. Consistent with Sanbonmatsu, Karder, and Herr (1992), the less experienced auditors made similar confidence judgments across differing levels of information completeness. The experienced subjects were more confident in the complete information condition, but less confident when information they considered relevant to the risk assessment task was omitted.

### **Revision of Auditor's First-Order Judgments**

Hypotheses concerning evidence sufficiency judgments should take into account knowledge of the audit risk assessment process. The present research requires only that audit risk judgments be predictable with respect to the sign of revision.. Ashton and Ashton (1988), Asare (1992), and McMillan and White (1993), among others, have investigated patterns of audit risk revision. All report high levels of agreement concerning the objective implications of audit evidence and therefore the direction of audit risk revision implied by

an item of conditioning evidence. These studies have found that the magnitude of revision depends on the degree to which conditioning evidence confirms prior belief (Ashton and Ashton 1988, Asare 1992, and McMillan and White 1993).<sup>6</sup> Accordingly, the magnitude of risk revision in the present experiment was expected to be amplified or dampened depending on treatment condition. The model tested by Ashton and Ashton (1988) provides a useful basis for an initial comparison of audit risk and evidence sufficiency revisions.

Ashton and Ashton (1988) report a series of experiments which used experienced auditors and an auditing scenario to test an early version of Einhorn and Hogarth's surprise/contrast model (1985b). They found that the direction of belief revision, relative to an inherited anchor, was in accordance with the objective implications of the evidence presented. For example, the statement that, "The tasks of timekeeping and payment of employees are adequately segregated from the task of payroll preparation" decreased control risk assessments whereas the statement that such separation was lacking increased control risk assessment. However, the anchor from which initial belief was adjusted had a significant effect on the magnitude of revision. There was little upward revision from high control risk anchors (.8) when evidence consistent with poor controls was evaluated. Similarly, decreases from low control risk anchors (.2) were insignificant when evidence

---

<sup>6</sup>Church (1991) has reported contrary results and McMillan and White (1993) report that the magnitude of change does not differ when using a scale which measures proportional as opposed to absolute magnitudes.

consistent with low control risk was evaluated. For both high and low anchors, large revisions in belief occurred when disconfirming evidence was sequentially evaluated.

These results are at least partially explainable by end-of-scale effects. For example, subjects who inherited a control risk anchor of .2 could reduce their assessment by a maximum of .2, regardless of the persuasiveness of additional evidence. Investigating the impact of confirmatory conditioning evidence on audit risk, but not evidence sufficiency, may lead to the incorrect conclusion that auditor belief is not affected by such evidence. If changes in evidence sufficiency, as well as audit risk, are measured, evidence that is confirmatory, and hence consistent, might be found to have an important role in the opinion formulation process.

## 2.5

### Conclusion

This chapter has reviewed the psychology and audit literature concerned with the causes and effects of second-order uncertainty. Ample evidence exists to conclude that second-order uncertainty at times influences judgment and choice. Although debate as to the normative status of second-order uncertainty continues, two plausible arguments support the decision-usefulness of such judgments.

Smith, Benson and Curley (1991) and Frisch and Baron (1988) have argued that basic cognitive processes are structured towards improving choice through ambiguity avoidance. Attending to second-order uncertainty may induce cognizers to avoid action when probabilities are based on incomplete information. Of course, the usefulness of such

a strategy depends on the feasibility of deferring choice and the potential for acquiring additional relevant information. Ambiguity avoidance may also be a function of the perceived threat of social and economic sanctions, whether or not responsibility to third-parties is an explicit element in the decision context (Curley, Yates and Abrams 1986, Brown 1990, Ashton 1992). These two functions are not mutually exclusive, rather it is likely that they jointly explain the role of second-order belief in the decision process.

The main findings of the empirical literature investigating second-order uncertainty may be summarized as follows: (1) subjects prefer gambles whose expected values are based on observable processes (Ellsberg 1961, Yates and Zukowski 1976, Curley, Yates and Abrams 1986); (2) small sample sizes and low source credibility increase ambiguity (Einhorn and Hogarth 1985a); (3) incomplete information lowers the judgment confidence of knowledgeable subjects (Waller 1990, Sanbonmatsu, Karder, and Herr 1992); (4) Confidence generally increases as incremental information is obtained (Morton and Felix 1990, Waller and Mitchell 1991).

A topic that has not been addressed by the existing literature is whether judgment confidence may be reduced by conditioning evidence and, if so, how do cognizers respond to increased second-order uncertainty? At issue is whether an evidence set can appear less complete as new information is obtained. The chapters that follow examine auditors' second-order uncertainty judgments and report an experiment which suggests that inconsistent conditioning evidence can decrease confidence. Since the consistency of conditioning evidence also affects first-order belief, insight is also provided into the

relationship between first-order and second-order judgments hypothesized by Waller and Felix (1984).

### **CHAPTER 3                      THEORY AND HYPOTHESES**

The primary purpose of the auditor's report is to communicate belief as to whether the statements examined are free from material misstatement (AU-110). Accordingly, cognitive auditing research has focused on the contextual and mental factors that affect judgments of audit risk. Empirical findings suggest that the formation and revision of audit risk assessments are consistent with the objective implications of conditioning evidence, but that models based on Bayes rule or other normative criteria incompletely describe auditor judgment (Waller and Felix 1987, Ashton and Ashton 1988).

Whereas a substantial body of knowledge exists concerning the determinants of audit risk judgments, less is known about evidence sufficiency assessments. The third standard of fieldwork states that "sufficient competent evidential matter" must be obtained prior to issuing an opinion, and AU-504 requires that a disclaimer be issued when reasonable assurance cannot be attained. When unusual relationships or results are noted the auditor is required to consider modifying the nature, extent and timing of planned procedures (AU-312.26). These prescriptions are aimed at avoiding opinions based on significantly incomplete information and suggest that evidence sufficiency judgments measure second-order uncertainty.

By viewing evidence sufficiency assessments as measures of judgment confidence, conjectures concerning their formation and revision can be adapted from the cognitive literature. The judged sufficiency of evidence may therefore depend on the amount, credibility and consistency of evidence obtained (eg., Ellsberg 1961; Smith, Benson and

Curley 1991) and may increase as conditional evidence is experienced (Waller and Mitchell 1991). The performance of additional or extended procedures following unexpected results may be viewed as a means of reducing ambiguity by obtaining more complete information (Frisch and Baron 1988).

The empirical auditing research has been largely confined to factors influencing a single confidence judgment (Waller 1990, Bedard 1992). Although it appears plausible that confidence should generally increase as procedures are performed, only one empirical study has addressed this issue (Morton and Felix 1990). Further investigation of the factors which affect sufficiency revision is needed. This chapter extends Waller and Felix (1984) by considering the joint effect of conditioning evidence on revision of audit risk and evidence sufficiency assessments. The hypotheses developed in the final section are aimed at investigating a fundamental property of evidence sufficiency assessments: do they necessarily increase as procedures are performed, or is positive revision contingent on obtaining results consistent with the implications of prior evidence?

The remainder of the chapter is divided into five sections. Section 1 presents the Waller & Felix (1984) model of auditor judgment. Section 2 discusses evidence sufficiency assessment during planning. Section 3 considers the effect of consistent and inconsistent conditioning evidence on sufficiency judgments. The substantive hypotheses are stated in Section 4.

### 3.1 The Waller and Felix (1984) Model of Opinion Formulation

Waller and Felix (1984) introduced consideration of second-order uncertainty into the auditing literature. They proposed that audit evidence is acquired to reduce:

- (1) a "first-order" uncertainty about whether the client's financial statements contain a material error and
- (2) a "second-order" uncertainty about whether a set of audit evidence provides adequate support for a conclusion about the "first-order" uncertainty.

The first role of evidence is characterized as opinion formulation (audit risk assessment), whereas the second is viewed as opinion justification (evidence sufficiency assessment). When evaluating whether an item of information affects auditor cognition its impact on both uncertainties must be considered. The two functions are said to be driven by evidence, "usually jointly but at times severally."

The model's variables and constraints are defined in Table 1. Waller and Felix (1984) adopt the notation  $Bel(.|.)$  to indicate belief, noting that such functions are not necessarily numerical and may not conform to any a priori axiomatic system. Materiality ( $M$ ) denotes the amount of misstatement in the financial statements that would probably affect the judgment of a reasonable user (AU-312).  $Bel(M|E)$  (audit risk) denotes the assessed risk of  $M$  based on conditioning evidence ( $E$ ) and is subsequently referred to as  $Bel(R)$ . The lower  $M$ , the greater the likelihood that misstatements in excess of  $M$  will exist and therefore, materiality is inversely related to the value of  $Bel(R)$ .  $R^*$  is the maximum value of  $Bel(R)$ , that the auditor is willing to tolerate at the conclusion of the engagement

without modifying the opinion. The value set for  $R^*$  may be based on the reputation the auditor wishes to maintain and the incidence of litigation deemed acceptable.

Table 1  
Elements of the Waller and Felix Model

M =	amount equal to material misstatement
E =	audit evidence obtained
R =	risk of an uncorrected material error (audit risk)
$R^*$ =	the maximum allowable value of R
S =	the degree to which the evidence collected justifies R
$S^*$ =	the minimum value of S required to justify expressing an opinion as to the value of R.
C =	the auditor's cost.
$C^*$ =	the maximum cost that the auditor is willing to bear to obtain evidence.

$Bel(S)$  is the assessed sufficiency of the evidence used to form  $Bel(R)$  with  $S^*$  representing the minimum value of  $Bel(S)$  that the auditor considers necessary in order to issue an unqualified report.  $Bel(S)$  is similar to a confidence assessment; it represents the belief that sufficient evidence has been obtained, rather than the risk that the evidence is inadequate. As such it would generally be expected to increase as a function of competent evidence collected.

$C$  and  $C^*$  refer to actual and maximum intended costs. Cost constraints provide incentive to limit evidence collection by restricting tests to the minimum level needed to induce  $Bel(R) \leq R^*$  and  $Bel(S) \geq S^*$ . Without economic constraints auditors would routinely conduct all possible tests so that both first-order and second-order uncertainty

would be reduced to negligible levels at the conclusion of most audits.<sup>7</sup> The audit task can be succinctly described as the collection of evidence until  $\text{Bel}(R) \leq R^*$ ,  $\text{Bel}(S) \geq S^*$ , subject to  $C \leq C^*$ .

### 3.2 Planning Risk and the Expected Sufficiency of Evidence

The review of the psychology and auditing literature suggests that second-order uncertainty is a function of the quantity, credibility and consistency of the evidence obtained. This characterization is congruent with the professional auditing standards which discuss the persuasiveness of evidence in terms of its sufficiency and competence (AU-326) and require that testing be extended when unexpected (inconsistent) results are obtained (AU-316). This section discusses the relationship between planning risk assessments and the quantity, credibility and consistency of evidence expected to be needed to attain  $\text{Bel}(S) \geq S^*$ .

According to the professional standards the value of  $\text{Bel}(R)$  used to plan detailed tests is composed of two independent and sequential elements: (1) inherent risk, or the risk that a material misstatement would occur in the absence of controls; and (2) control risk, or the risk that the existing control structure would neither prevent nor detect such an occurrence (AU-312, AU-350). Research suggests that these constructs are, in fact, neither independent (Cushing and Loebbecke 1983, Waller 1993), nor sequential (Waller 1993).

---

<sup>7</sup>The economics of auditing is beyond the scope of the research. The discussion only assumes that auditors are responsive to economic incentives and attempt to conduct an efficient audit. Whether auditors adjust  $S^*$  to avoid exceeding  $C^*$  is not considered.

The terms planning risk or  $Bel(R_p)$  are adopted herein to denote the value of  $Bel(R)$  used for developing the detailed audit program without assumption as to nature of decomposition.<sup>8</sup>

Planning risk judgments vary across engagements, accounts and assertions. One effect of differing planning risk assessments is that more persuasive evidence is necessary to induce  $Bel(S) \geq S^*$  when  $Bel(R_p)$  is high as compared to low. Since clients generally desire unqualified opinions, accepting an engagement implies that the auditor will attempt to obtain reasonable assurance that all material misstatements have been detected and corrected. The amount of conditioning evidence necessary to drive posterior  $Bel(R)$  below  $R^*$  will be greatest when  $Bel(R_p)$  is maximum. This relationship is incorporated in the audit risk model (AU-350) which specifies an inverse relationship between  $Bel(R_p)$  and the allowable risk that audit procedures will fail to detect material misstatements.<sup>9</sup> High planning risk assessments require that extensive effective procedures be performed in order to attain  $S^*$ , and low assessments may allow for the restriction of tests and the substitution of less effective procedures.

---

<sup>8</sup> Waller (1993) uses the term "occurrence risk" to denote the auditor's assessment of the likelihood that a material misstatement existed prior to beginning the audit. Since misstatements may have been detected prior to developing the detailed audit program, the values of  $Bel(R)$  used in tactical planning might differ from occurrence risk, hence the use of the term "planning risk".

<sup>9</sup>Using PDR to indicate planned detection risk, the model can be restated as  $PDR = R^*/Bel(R_p)$ . However, PDR is not equivalent to  $S^*$ , since  $Bel(R_p)$  and  $R^*$  comprise only two of the factors which determine  $S^*$ . Audit procedures may be performed even when  $R^*/Bel(R_p) > 1$  indicating that PDR does not fully determine evidence sufficiency.

Once planning risk for an audit area has been formed, the auditor must choose the type of procedures to be applied. The procedures selected provide a mix of evidence quantity, credibility and consistency that the auditor believes will induce  $Bel(R) \leq R^*$  and  $Bel(S) \geq S^*$  without exceeding  $C^*$ . The trade-off between justification and cost restricts quantity and credibility, since larger samples are more expensive and the credibility of evidence is often directly related to the cost of its acquisition.<sup>10</sup> Thus, auditors must usually be satisfied by evidence that is persuasive rather than convincing (AU-326). Evidence consistency, unlike quantity or credibility, does not directly affect audit cost and, ceteris paribus, the auditor would prefer to rely more on consistency and less on quantity and credibility to achieve  $S^*$ .<sup>11</sup> However, since the results of a procedure are not known in advance, relying on expected consistency entails a risk that the realized increase in  $Bel(S)$  will be less than intended.

### **3.3 Evidence Consistency and the Revision of Audit Risk and Sufficiency Assessments**

Evidence consistency has been defined (Sec 2.2) as the extent to which all elements in an information set point towards the same outcome. If the evidence obtained during

---

<sup>10</sup>For example, vouching is less costly than confirmation, but is also less persuasive.

<sup>11</sup>Consistency cannot entirely offset the need for adequate amounts of credible evidence (Einhorn and Hogarth 1985a), a point also recognized by the professional standards (AU 326, AU-350). Thus, when  $Bel(R_p) \leq R^*$ , the procedures planned will be expected to yield consistent but limited evidence, that will be sufficient to support the opinion.

planning is uniformly consistent, then  $Bel(R_p)$  will assume an extreme value. In contrast, evidence with mixed implications will result in moderate values for planning risk. In a belief revision context such as auditing it may be inefficient to evaluate the consistency of the full evidence set each time new data is acquired. The extremeness of  $Bel(R_p)$  may instead proxy for ex ante consistency and ex post consistency may be judged by the extent to which conditioning evidence confirms planning risk. Conditioning evidence that is **consistent** with prior evidence should also **confirm** prior belief and thus consistency jointly affects the revision of  $Bel(S)$  and  $Bel(R)$ . The circumstances under which expectations of high consistency can be traded-off to reduce the quantity or credibility of conditioning evidence and decrease cost are considered next.

If the extremeness of  $Bel(R_p)$  can proxy for the consistency of the initial evidence, it can also serve to estimate ex post consistency. If evidence acquired during planning uniformly indicates that an account is fairly presented and  $Bel(R_p)$  is low, conditioning evidence will be expected to be confirmatory and therefore consistent. The quantity and credibility of conditioning evidence that the auditor imagines will be needed to achieve  $Bel(S) \geq S^*$  will be lower than if less consistent results were expected. Sample sizes may be reduced and less costly (and less effective) procedures selected.

If initial evidence implies the existence of material misstatement,  $Bel(R_p)$  will be high and procedures will again be expected to yield consistent and confirmatory conditioning evidence. If it were acceptable to issue an adverse or qualified opinion after obtaining sufficient evidence supporting  $Bel(R) > R^*$ , expected consistency could again be traded-off

against quantity and credibility to reduce audit effort. However, because the auditor has been posited to desire a final value of  $Bel(R) \leq R^*$ , expectations of consistent evidence confirming  $Bel(R) > R^*$  cannot be used to reduce evidence quantity or credibility.

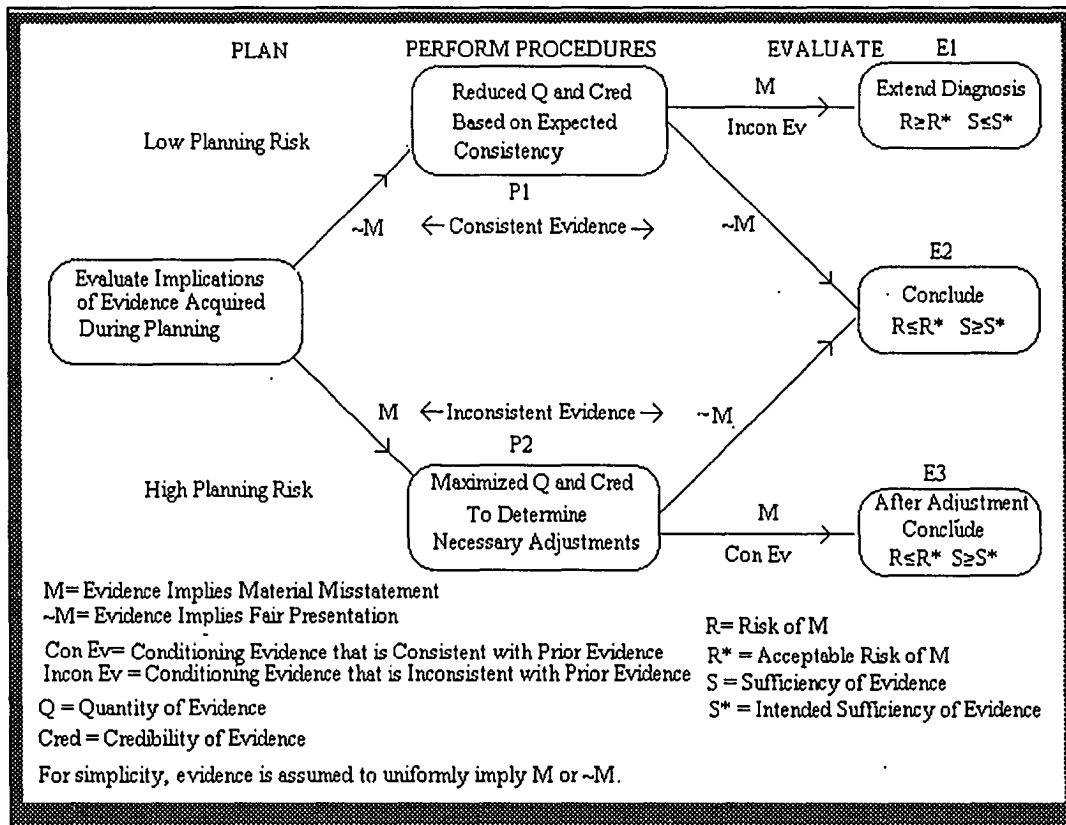
The tests performed when planning risk is high are expected to do more than provide reasonable assurance that material misstatements exist. If errors are detected, the evidence must be sufficient to indicate the adjustments necessary for fair presentation. An examination intended to ascertain necessary adjustments is likely to be far more intensive than would be required simply to establish the presence of error. To conclude  $R > R^*$  the auditor need only detect a single material misstatement ( $M$ ), a result which might be obtained by testing one transaction. However, to conclude  $Bel(R) \leq R^*$ , after discovering  $M$ , the auditor may have to continue procedures until the remaining unexamined items are immaterial. Thus, when  $Bel(R_p)$  is high, audit effort cannot be reduced based on an expectation of consistent conditioning evidence.

Finally, when initial evidence has mixed implications, no expectation of consistency between prior and conditioning evidence can exist.  $Bel(R_p)$  will be moderate and the audit program will rely on the quantity and credibility of evidence to achieve  $Bel(S) \geq S^*$ .

The foregoing arguments suggest that an asymmetry exists between evidence sufficiency requirements when initial evidence is consistent with fair presentation, as opposed to material misstatement. Initial evidence implying fair presentation  $\sim M$  may lead to the restriction of tests based on the expected consistency of conditioning evidence. No such reduction is available when initial evidence suggests the existence of material

misstatement. Given the goal of reporting  $Bel(R) \leq R^*$ , the auditor must rely more extensively on evidence quantity and credibility to satisfy  $S^*$  when  $Bel(R_p)$  is high as compared to low.

Figure 1  
Effect of the Consistency of Conditioning Evidence on Bel (R) and Bel (S)



The postulated relationship between evidence consistency and assessment of  $Bel(R)$  and  $Bel(S)$  is depicted in Figure 1 (above). In order to determine the nature and extent of testing, the auditor first evaluates the implications of evidence acquired during planning. Initial evidence indicating fair presentation ( $\sim M$ ) leads to a low value for  $Bel(R_p)$  and

reduced procedures (Box P1). Limited, but consistent, conditioning evidence implying  $\sim M$  will allow the auditor to conclude  $Bel(R) \leq R^*$ ,  $Bel(S) \geq S^*$  (E2). Evidence implying material misstatement (M) will induce the auditor to maximize the quantity and credibility of evidence (P2). Consistent conditioning evidence implying M will be sufficient to confirm  $Bel(R) > R^*$ , but a opinion reflecting  $Bel(R) \leq R^*$  will be issued if the client accedes to the required adjustments (E3). When inconsistent conditioning evidence is obtained,  $Bel(R)$  will be revised in the implied direction. The effect on  $Bel(S)$  will depend on whether the initial evidence supported  $\sim M$  or M and therefore, whether achieving  $S^*$  was predicated on obtaining consistent results.

An audit program considered adequate to justify a  $\sim M$  conclusion (P1) may be inadequate to persuasively support M. If expected consistency is used to reduce the quantity and credibility of evidence when  $Bel(R_p)$  is low, results that fail to clearly confirm  $\sim M$  will cause ex post  $Bel(S)$  will be less than its expected value. Outcomes that imply M may reduce  $Bel(S)$  below its ex ante value, if the inconsistency outweighs the increase in quantity and credibility. Additional and perhaps more credible types of evidence will be needed to attain  $Bel(R) \leq R^*$  and  $Bel(S) \geq S^*$  (E1, Figure 1).

Before performing additional procedures the auditor is likely to attempt to diagnose the cause of the inconsistency. Determining probable cause helps the auditor obtain a better understanding of the client's operating environment and control structure and reduces the likelihood that additional procedures will be inconclusive, or fail to identify all misstatements. Initial diagnosis will be performed quickly without much regard for the

formal generation or evaluation of hypotheses. The auditor is likely to verify the correct performance of the procedure and review the audit program and related memoranda for errors in logic.<sup>12</sup> If probable cause is not determined by these actions, the auditor may make a second iteration through hypothesis generation and assessment which is likely to be more structured and may be formally documented by modifications of the audit program. Hypotheses may be self-generated or suggested by the client. Candidate hypotheses can be tested by expanding sample size or by performing procedures not originally considered necessary. The process concludes when  $Bel(S) \geq S^*$  and  $Bel(R) \leq R^*$ , either because an error cause has been judged probable and an adjustment made, or a nonerror explanation has been accepted.

When initial evidence indicates the likely presence of material misstatement and the auditor begins with a presumption of error, the audit plan will provide for the collection of large amounts of credible evidence (P2, Figure 1). Expected consistency will not be used to reduce evidence quantity or credibility since the examination must be sufficiently comprehensive to support adjustments and confirm the existence of material misstatement. Conditioning evidence that fails to identify misstatements would be deemed inconsistent. However, not having reduced testing based on expected consistency, the auditor would be unlikely to significantly extend diagnosis. The extensiveness of the planned procedures would render it unlikely that additional relevant information could be obtained. After

---

<sup>12</sup> Logical errors are an aspect of nonsampling risk, which according to the professional standards "can be reduced to a negligible level" through proper quality control (AU-350.11).

scrutinizing the work papers and questioning assistants to determine whether procedures were improperly performed or evaluated, the unexpected result is likely to be accepted and an unqualified report issued (E3, Figure 1).

In the absence of consistent prior evidence and therefore extreme values for  $Bel(R_p)$ , the consistency of conditioning evidence will not be expected to contribute to second-order uncertainty reduction. The audit plan will rely on evidence quantity and credibility to achieve opinion justification.  $Bel(R)$  will be revised in the appropriate direction upon observing evidence implying  $M$  or  $\sim M$  and  $Bel(S)$  should increase to its minimum acceptable value. If the auditor concludes  $Bel(R) > R^*$  testing may be expanded to determine the adjustments necessary for fair presentation.

### **3.4 Hypotheses**

The discussion has suggested that auditors purposefully structure their tests so that obtaining sufficient justification from a procedure may depend, in part, on the consistency of evidence obtained. Sample sizes are minimized and costly, but effective, procedures foregone only when the evidence acquired during planning consistently suggests that an account balance is fairly stated. If significant audit differences are subsequently identified, the quantity and credibility of evidence may be inadequate to persuasively establish the existence of error or indicate the required adjustments. In such instances realized  $Bel(S)$  will be less than its expected value and the auditor must determine whether revision of the audit program is needed to attain  $Bel(S) > S^*$ . Contrarily, evidence collection is maximized

when initial evidence implies material misstatement, since the examination must be sufficient both to confirm  $Bel(R) > R^*$  and to determine the corrections needed to permit an unqualified report. Inconsistent audit evidence may have a differing effect on sufficiency revision when initial evidence implies fair presentation as opposed to material misstatement.

The hypotheses are developed in a simple audit environment (See Table 2 below) in which: (1) initial assessments of audit risk and evidence sufficiency are based on evidence that suggests either material misstatement or fair presentation and; (2) revised assessments are conditioned on subsequent findings that are either consistent or inconsistent with the implications of the initial evidence.

Table 2  
Effect of the Consistency of Conditioning Evidence on  
Revision of  $Bel(R)$  and  $Bel(S)$

Implications of Prior Evidence	Implications of Audit Test	
	Fair Presentation	Material Misstatement
Fair Presentation	A    R- * S+	B    R+ S-
Material Misstatement	C    R- S+	D    R+ * S+
* Revision may be insignificant in cells A and D due to the limited impact of consistent evidence.		

The present study requires that audit risk judgments be predictable with respect to the sign of revision. As noted in Section 2.4, Ashton and Ashton (1988), Asare (1992), and McMillan and White (1993) have investigated patterns of audit risk revision, finding that

high levels of agreement exist concerning the objective implications of audit evidence and the direction of audit risk revision. It was therefore expected that  $Bel(R)$  would be negatively revised in Cells A and B, and positively in Cells C and D. Hypothesis 1 is stated in the alternative form as:

H1:  $Bel(R)$  will be increased (decreased) by conditioning evidence implying material misstatement (fair presentation).

The researchers cited in the preceding paragraph also found that the magnitude of revision depended on the degree to which conditioning evidence confirmed prior belief. Although not the subject of a formal hypothesis, only the mean revisions in Cells B and D were expected to be significant. Defining the proportion of successful predictions as  $Success_{risk}$ , and the proportion of the more frequent sign of revision as  $\theta_{risk}$ <sup>13</sup>, the test of H1 is formally stated as:

Conclude  $H1_o$ , if  $Success_{risk} \leq \theta_{risk}$       Conclude  $H1_a$ , if  $Success_{risk} > \theta_{risk}$

The second and main hypothesis, stated in the alternative form, predicts the effect of evidence consistency on evidence sufficiency judgments.

H2:  $Bel(S)$  will be increased by credible conditioning evidence unless such evidence is inconsistent with prior evidence indicating fair presentation.

It was therefore expected that  $Bel(S)$  would be revised positively in Cells A, C, and D, but negatively in Cell B. By defining the proportion of successful predictions based on

---

<sup>13</sup>Note that using the maximum likelihood outcome as the null provides a more rigorous test than specifying  $\theta=.5$ .

Hypothesis 2 as  $\text{Success}_{\text{suff}}$  and the proportion of the more frequent sign of sufficiency revision as  $\theta_{\text{suff}}$ , the test of Hypothesis 2 takes the form of:

Conclude  $H2_o$ , if  $\text{Success}_{\text{suff}} \leq \theta_{\text{suff}}$       Conclude  $H2_a$ , if  $\text{Success}_{\text{suff}} > \theta_{\text{suff}}$

Two simpler explanations may better account for the sign of sufficiency revision. First,  $\text{Bel}(S)$  may be influenced by the amount and credibility of evidence, but not by its consistency. In this case,  $\text{Bel}(S)$  would tend to increase whenever additional credible evidence was obtained and positive sufficiency revision would occur in all four cells. Second, if the conjecture that unique characteristics of the audit domain lead to asymmetrical effects for inconsistent evidence is false,  $\text{Bel}(S)$  would increase in the consistent cells (A and D) and decrease in both inconsistent cells (B and C). The degree to which the experimental evidence supports these alternative explanations is considered in Section 5.

The discussion of evidence sufficiency assessments and their relationship to audit risk judgments has led to hypotheses concerning the effects of consistent and inconsistent conditioning evidence. Results that corroborate the hypotheses will (1) Provide insight into the manner in which evidence drives audit risk and sufficiency assessments "usually jointly, but at times severally" (Waller and Felix 1984); and (2) support cognitive models which incorporate consistency of evidence as a determinant of second-order uncertainty.

## CHAPTER 4 RESEARCH DESIGN

The hypotheses presented in Chapter 3 were tested in a between subjects 2 X 2 factorial experiment with repeated measures of audit risk and evidence sufficiency (Figure 2, below).<sup>14</sup> Subjects read and responded to a fictional case describing a test of transactions in the revenue cycle. The first treatment (PRIOR RISK) was a blocking variable intended to suggest that sales revenue was either fairly presented or materially misstated and induce lower or higher audit risk assessments. Treatment levels are accordingly referred to as LOW PRIOR and HIGH PRIOR. The second treatment (ERROR SIZE) provided a stimulus for belief revision consisting of trivial or potentially significant audit differences which are referred to as SMALL ERROR and LARGE ERROR. O1 and O2 denote the initial and posterior assessments of Bel(R) and Bel(S). Bel(R) was defined as "the risk that sales revenue is overstated by an amount greater than tolerable error." Bel(S) was described as "your belief as to whether you already have sufficient evidence (as contemplated by the third standard of fieldwork) to justify [the audit risk assessment]."

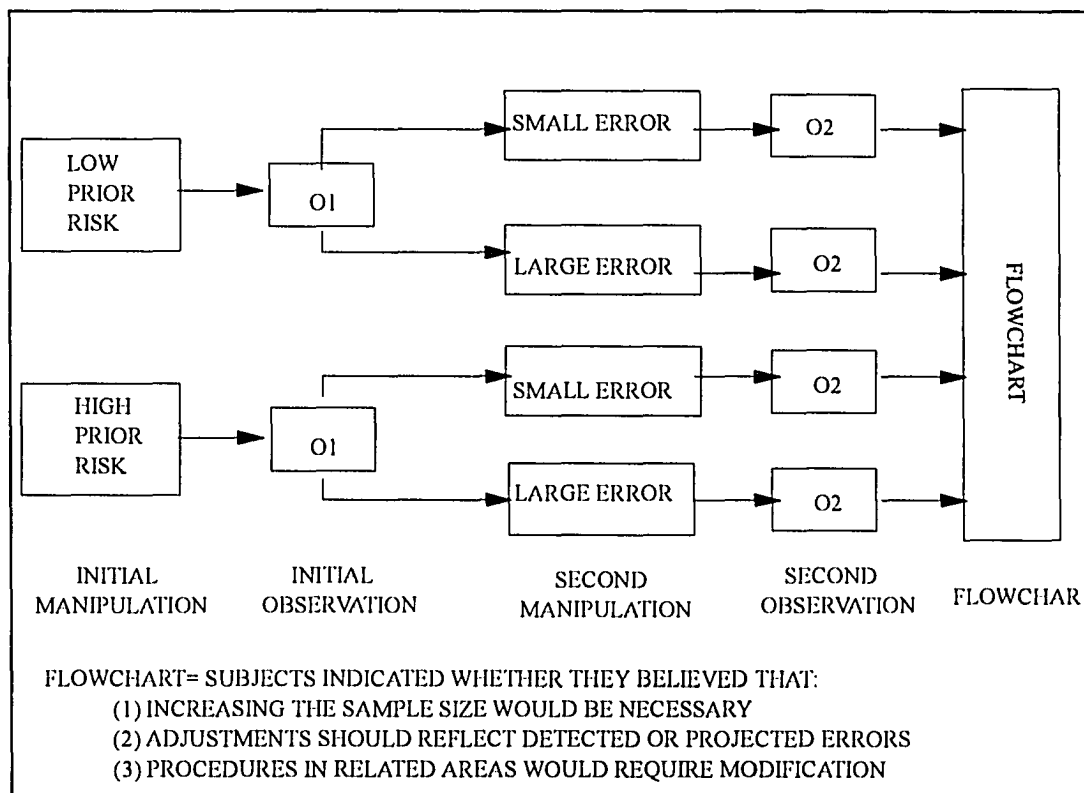
Prior risk assessments were obtained on both ordinal and continuous scales. As noted by Waller (1990), auditors typically make ordinal rather than continuous risk judgments and may be more comfortable using a format which allows for such responses. However, for research purposes the imprecision of an ordinal scale may obscure

---

<sup>14</sup>The assessments were elicited in terms of audit risk instead of detection risk to avoid the need to specify acceptable risk (R\*). Also, once substantive procedures have been performed PDR≠DR.

relationships between treatment variables and belief revision. This might be especially true for responses to consistent conditioning evidence. For example, if  $Bel(R)$  was initially assessed as low on a 3-point scale, it could not be negatively revised after the evaluation of consistent conditioning evidence. The continuous scale may have less external validity, but be better suited for the measurement of belief revision.

Figure 2  
Research Design



Following the second series of assessments, subjects reviewed a flow chart depicting possible sequences of follow-up actions that an auditor might undertake in response to the

test results. They then indicated the sequence considered most likely to occur. The purpose was to determine whether an association exists between the consistency of conditioning evidence and the decision to extend diagnosis.

#### **4.1 Subjects**

Fifty-nine subjects, of whom 55 returned usable instruments, were recruited from two medium-sized firms located in New York City. Fifty-eight percent of the subjects were CPA's, who ranged in age from 22 to 69 years and varied in experience from a few months to 41 years. Six subjects, all of whom were senior staff, indicated that they had not audited within the past 24 months. The seniors and managers participating in the study were more experienced than would be expected from a similar sample of Big Six auditors. Subject characteristics are presented in more detail in section 5.4.

The clients audited by the two firms are more homogeneous and smaller than those of the Big Six or other large firms. Staff frequently returns to the same engagements facilitating the development of skill through repetition. The subjects are less accustomed to dealing with complicated reporting issues than their Big Six counterparts, but appear to be good choices for performing a moderately complex task such as is required by the experimental materials. Firm X uses an audit technology similar to that employed by some large structured firms (eg., PMM and Grant Thornton). Subjects from Firm X use a work paper that requires the rating of risk on a 3 point scale. Firm Y is unstructured; risks are assessed and sample sizes selected without use of external decision aids. This difference

was a potential source for firm effects. Neither firm makes regular use of control testing, both because of efficiency considerations (Waller 1993) and because of the potential for override of controls.

#### **4.2 Procedure**

The instrument was first pretested using undergraduate auditing students who completed it at home. Approximately 60 instruments were distributed and 26 responses received. The results indicated that the subjects understood the instructions and case materials. Effects were weak but in the predicted directions. A revised instrument was completed by 12 Phd students and accounting faculty and final changes made.

Data was collected during three experimental sessions, two for Firm X and one for Firm Y. Each session, which lasted approximately 40 minutes, began with the distribution of case booklets. The order of distribution, after a random start, was A-B-C-D-B-A-B-C-D which ensured that a sufficient number of B instruments were collected to permit tests of Hypothesis 2.

Next, the researcher read the instructions and a sample of the response scales were passed out for examination. Subjects were told to complete the 3-point audit risk scale first and then the 100-point scale. They were informed that the purpose of the 100-point audit risk scale was to allow for a more precise location of their beliefs than could be done using a 3-point scale. The researcher stated that he understood that the subjects did not typically use a 100 point scale, but that they should do their best to express their beliefs. The

sufficiency scale was described as measuring the proportion of evidence already collected as compared to the total amount needed for sign-off. Subjects were requested not to consult with each other, not to change answers, thanked for their participation, and asked to start. Almost all subjects took at least 15 minutes to complete the case, but only a few continued beyond 25 minutes. No consultations were observed and most appeared attentive to the task, possibly influenced by the presence of senior staff.

#### **4.3 Task and Materials**

The case describes (See Appendix) the audit of a small manufacturing firm. Information was presented to familiarize subjects with the engagement including: (1) factors relevant to assessing inherent risk; (2) a brief summary of the study of internal control; (3) comparative income statements and; (4) a list of audit procedures completed and the results obtained.

Testing the hypotheses required that subjects be blocked into low and high planning risk groups. The PRIOR RISK treatment consisted of a series of binary phrases or values intended to suggest uniformly higher or lower audit risk. The description of the client's operations included several statements that the professional standards suggest are indicative of higher or lower inherent risk (AU-316). The comparative income statements showed a percentage increase in sales that was either consistent with variation in other accounts or significantly higher. Confirmation response rates and exceptions were stated to have either

improved or worsened relative to the prior year and the sales cut-off test had either been satisfactorily completed or was in progress.

The manipulation was intended to create differing, but not extreme initial audit risk assessments across treatment conditions so that subjects responses would not be unduly limited by range restrictions. Anchors near 0 or 100 could lead to the type of end-of-scale effects noted by Ashton and Ashton (1988) and McMillan and White (1993) that skew response functions and possibly create spurious treatment effects.<sup>15</sup>

Information concerning controls was held constant across PRIOR RISK conditions; it indicated an apparently adequate system that had not been tested due to efficiency considerations. Several reasons exist for not varying control risk. (1) As noted, the design was intended to induce lower or higher audit risk assessments, but was not intended to suggest extreme values. Major control weaknesses would be likely to induce near maximum assessments of audit risk. (2) Tests of internal control in the revenue cycle are usually performed simultaneously with substantive tests of transactions, a fact known to experienced auditors. The results of control tests would not be a typical type of evidence available to form prior belief as to the existence or valuation of sales. (3) Reporting test results supporting effective controls and then "discovering" large errors might cause subjects to perceive too clearly the hand of the researcher, thus detracting from experimental realism.

---

<sup>15</sup>See McMillan and White (1993) for a more detailed discussion.

After reading the materials, subjects provided initial assessments of audit risk and evidence sufficiency.

Next a vouching procedure designed to test the occurrence and valuation of sales transactions was described. The procedure is commonly used when control tests are not performed. Subjects then turned to the results of the procedure and read that a small (large) difference had been detected. The work paper contained the sample errors and a projection to the account balance. The small error was 2% of tolerable misstatement before projection to the account balance and about 15% when projected. The large error was 21% before projection and 144% afterwards. After the results of the procedure were revealed subjects reevaluated audit risk and evidence sufficiency.

The following page contained a flow chart which presented several options concerning likely follow-up procedures. Its purpose was to explore several conjectures related to the effect of evidence consistency on the decision to extend diagnosis of audit findings. Subjects were able to directly indicate their beliefs concerning the eventual disposition of the difference. In considering the flowchart, subjects were asked to assume that the correct application of the procedure had been checked, but that no other investigation had been conducted. The client had agreed to adjust for actual differences, but was unwilling to book the full projected amount. Subjects indicated the route through the flowchart that would be most likely to occur. The flow chart incorporated three dichotomous choices: (1) select/do not select an additional sample of invoices; (2) continue

with the audit as planned or modify other procedure; (3) accept the adjustment based on actual errors or propose adjustment based on projected error.

The first variable was of primary interest because it provides a means of testing whether inconsistent conditioning evidence prompts extended diagnosis. Subjects in the Cell B (LOW PRIOR/LARGE ERROR) were predicted to believe that expanding the sample size would be necessary. Subjects in the other cells were expected to be satisfied with the results of the procedure.

The second choice tested whether subjects believed that the outcome of the diagnosis was likely to affect procedures performed in other audit areas. SAS-53 requires that the impact of unexpected errors on other aspects of the audit be considered and, therefore, it was of interest to determine whether subjects in Cell B were more likely to consider revising related procedures. For example, the overstating of sales suggests that receivables might be over valued. In order to obtain sufficient evidence to justify  $Bel(R) \leq R^*$  for accounts receivable, additional testing of subsequent cash collections might be performed.

The last variable serves as a manipulation check, providing information as to whether subjects viewed the difference as the detected or projected amount. Burgstahler and Jiambalvo (1986) report that auditors often fail to project detected errors to account balances as required by AU-350. The detected difference in the HIGH ERROR condition was only 21% of tolerable misstatement (projected=144%). If subjects considered only the

actual difference when revising their beliefs, it would be unlikely that the treatment, as perceived, would be of sufficient strength to constitute a valid test of the hypotheses.

Lastly, a short exit questionnaire checking the effectiveness of the experimental manipulations and eliciting demographic data was completed. Subjects were asked to record the impressions they had developed concerning the levels of inherent risk and control risk, as well as the perceived effectiveness of the vouching procedure. They were queried as to which accounts they had initially considered to contain possible errors and asked to indicate whether the audit difference detected exceeded tolerable error.

#### 4.4

#### Conclusion

The purpose of the experiment was to create a valid artificial environment for testing the substantive hypotheses. Because this research seeks to provide knowledge useful to understanding auditor cognition it was necessary to use experienced subjects and provide materials which abstracted the features from their domain essential to testing the hypotheses. Valid testing of the hypotheses required subjects to: (1) form differing prior audit risk assessments; (2) perceive the described procedure as effective; and (3) consider the potential materiality of the projected difference. The comparisons of initial audit risk assessments (Section 5.1) and the tests for manipulation effectiveness (section 5.4) suggest that subjects generally perceived the cases as intended and that a valid environment for inference was established.



Descriptive statistics for the initial assessments of audit risk are presented in Table 3 below. LMH1 represents initial risk assessment on an ordinal scale (low, medium, high) and R1 denotes initial risk assessment on a continuous 100-point scale. Mean ranks and the Mann-Whitney U statistic are reported for LMH1 and arithmetic means and the T-statistic for R1.

The Mann-Whitney U test (Gibbons 1976) allows for inference as to whether two samples of ordinal data share the same location. To perform the test, responses were first ranked in ascending order (L,M,H).<sup>16</sup> Next, observations were classified by condition (LOW PRIOR, HIGH PRIOR) and separate rank sums computed. The actual rank sum for the smaller group (HIGH PRIOR) was compared with the rank sum that would be expected if responses were not affected by the treatment.<sup>17</sup> The Mann-Whitney U statistic of 191.5 indicates the absolute difference between the obtained and expected rank sums for the HIGH PRIOR condition.

The U statistic for LMH1 (191.5) and the T-statistic for R1 (3.02) indicate that risk assessments significantly differed across treatment conditions. The rank means (arithmetic means) of 22.18 (38.3) for LOW PRIOR and 35.52 (58.88) for HIGH PRIOR indicate that,

---

<sup>16</sup>Because the response variable was limited to 3 values, a procedure for dealing with ties must be employed. The method recommended by Gibbons (1976) and utilized by SPSS is to assign an average rank to all tied observations. In the present case the 18 low responses were all ranked 9.5, the 27 medium responses were each ranked 32, and the 10 high responses ranked 50.5. SPSS corrects significance levels for the effect of ties.

<sup>17</sup>The expected rank sum is  $n_1(n+1)/2$  or  $24(56)/2=672$ .

as intended, subjects in the LOW PRIOR condition formed lower risk assessments than those receiving the HIGH PRIOR treatment. As would be expected from alternate measures of the same construct (See Table 3), LMH1 and R1 are highly correlated ( $r=.83$ ).

Table 3  
Summary Statistics for Initial Assessments of Risk and Sufficiency

	LMH1	R1	S1
<u>Condition</u>	<u>rank mean</u>	<u>mean</u>	<u>mean</u>
LOW PRIOR, n=31	22	38.3	56
HIGH PRIOR, n=24	35	58.88	26.95
grand mean	28	46.87	43.24
standard dev	na	25.53	26.54
M-W U (T-Stat)	191.5	(3.02)	(5.01)
P-value	$\leq .0008$	$\leq .004$	$\leq .0001$
Rank means are obtained by dividing rank sums by the number of subjects in the grouping. LOW PRIOR= $687.5/31=22.18$ , HIGH PRIOR = $852.5/24=35.52$ , grand rank mean = $1,540/55=28$			

S1 represents initial assessments of evidence sufficiency measured on a continuous 100-point scale. Differing assessments for S1 were not necessary for valid inference concerning Hypothesis 2, but analysis provides additional insight into the patterns of sufficiency judgments. Mean response was 56 in the LOW condition and 26 in the HIGH condition ( $p \leq .0001$ ). As a result, S1 and the two risk measures show strong negative correlations (See Table 4). The difference in S1 across conditions may have been due to differing information concerning the results of related procedures (See Section 4.3).

Table 4  
Correlations Between Initial Assessments of  
Risk and Sufficiency

	R1	S1
LMH1	83#	-.63#
R1		-.68##
#Spearman correlation #Pearson correlation		

#### Posterior Assessments of Bel(R) and Bel(S)

Posterior assessments were made after subjects reviewed the results of the vouching procedure. Mean ranks for LMH2 and means for R2 and S2 are reported by cell in Table 5, Panel A. Inspection of the condition means in Panel B reveals that assessments of R2 were driven almost entirely by the size of the audit difference. The LOW PRIOR mean was 6 points greater than HIGH PRIOR, but the LARGE ERROR mean exceeded LOW ERROR by 40.47. Similar results were obtained for LMH2, when differences in mean ranks were computed. LOW PRIOR was greater than HIGH PRIOR by 2.7 ranks and SMALL ERROR exceeded LARGE ERROR by 23.25. Note that due to the imbalanced design, treatment effects are not additive. Summary statistics for ANOVA and Kruskal-Wallis models are presented in Panel C. These tests confirm the impact of ERROR SIZE on R2. The multiple  $R^2$  of .54 obtained for the ANOVA model indicates a higher than expected level of intersubject consensus on posterior risk.

Table 5  
Summary Statistics for Posterior Assessments of Risk and Sufficiency

		LMH2	R2	S2
Panel A Cell means		n	Rank	Mean
A	Low Prior/Small Error	13	18.35	33.31
B	Low Prior/Large Error	18	33.56	67.72
C	High Prior/Small Error	12	17.54	23.50
D	High Prior/Large Error	12	40.58	71.08
Panel B Condition Means		n	Rank	Mean
	Low Prior	31	29.18	53.29
	High Prior	24	26.48	47.29
	Small Error	25	15.32	28.60
	Large Error	30	38.57	69.07
Grand Mean (mean rank)		55	(28)	50.67
Standard deviation			na	27.72
Panel C ANOVA (K-W) Summary		LMH2	R2	S2
PRIOR	F-stat (Chi-square)	(.38)	.278	4.12
	p ≤ x	.5356	.600	.048
ERROR	F-stat (Chi-square)	(28.7)	59.82	3.484
	p ≤ x	.0005	.005	.068
Full Model	R <sup>2</sup>	na	.540	.137
	F-stat (Chi-square)	(22.3)	30.57	4.12
	p ≤ x	.0001	.005	.022

Table 5, panel A reveals that the posterior assessment of Bel(S) was lowest in cell B (48.78) and highest in cell C (73.33), the two cells which received inconsistent conditioning information. The high level of S2 recorded in response to the absence of expected differences supports the argument that auditors do not view evidence inconsistent with a high prior risk assessment as increasing second-order uncertainty. The subjects were

most satisfied with the sufficiency of evidence after vouching did not disclose expected errors and least satisfied when an unexpected error was detected. S2 was approximately the same for Cell A (61.62) and Cell D (62.75) the cells representing consistent prior and conditioning evidence. The summary statistics reported for the S2 ANOVA in Panel C indicates only modest fit ( $R^2=.133$ ). PRIOR RISK ( $F=4.133$ ,  $p\leq.048$ ) and ERROR SIZE ( $F=3.484$ ,  $p\leq.068$ ) are only marginally significant.

As reported in Table 6, the correlation between the alternate measures of risk (R2 and LMH2) remained high for the posterior assessments (.75), but the strong inverse relationship between risk and sufficiency that obtained for the first series of judgments was substantially weakened. R2 and S2 were correlated -.33, and LMH2 and S2 -.24 as compared with the correlations of -.68 and -.63 shown in Table 4 for the initial assessments.

Table 6  
Correlations Between Posterior Assessments of  
Risk and Sufficiency

	R2	S2
LMH2	.75#	-.24#
R2		-.33##
#Spearman correlation ##Pearson correlation		

This completes the discussion of the initial and posterior measures of risk and sufficiency. The analysis indicated that the subjects, as intended, perceived audit risk to be less in the LOW PRIOR than in the HIGH PRIOR condition. The alternate measures of risk

(ordinal and continuous) produced highly similar results. The results reported in Sections 5.2 and 5.3 are confined to those obtained from the continuous scale.

## 5.2 Experimental Results

This section reports the primary tests of the substantive hypotheses. The hypotheses presented in Section 3 predict the direction of audit risk and evidence sufficiency revision and the results are analyzed using binomial tests based on the proportion of correct predictions. Two reasons exist for analyzing the data categorically. First, the hypotheses predict the sign of individual belief revision, whereas ANOVA provides no clear indication of the proportion of subjects whose revisions were in the predicted directions. Second, ANOVA does not permit a means of determining the extent to which subjects whose responses are well-fit by the risk revision model are also well-fit by the sufficiency model. A joint binomial test of H1 and H2 can be performed based on the proportion of subjects who revised both belief functions as predicted. This section reports the primary tests of the substantive hypotheses.

The sign of belief revision can be viewed as a binomial event (+,-) in which revisions predicted by a test hypothesis are classified as "successes" and revisions in the opposite direction (or no revision) as "misses." For the separate tests of H1 and H2, responses were categorized by the direction of belief revision, with the proportion of the more frequent sign serving as the null value ( $\theta_{risk}$  or  $\theta_{suff}$ ) of parameter  $p$ . The observations were next classified as "successes" or "misses" depending on whether the sign of change was correctly predicted

by the test hypothesis. The proportion of "successes" were then compared with the appropriate value of  $\theta$ . If subjects tended to revise belief as predicted, then the proportion of "successes" would exceed that obtained by classifying all observations by the more frequent sign of revision. A similar test was performed for the joint hypotheses, differing only in that the null was the most frequently obtained pair of signs and a success required that the revision of both  $Bel(R)$  and  $Bel(S)$  be correctly predicted.

### **Test of Risk Revision**

The sign of audit risk revision used to test Hypothesis 1 was determined by subtracting the initial risk judgment ( $R1$ ) from the revised judgment ( $R2$ ).  $H1$  states that:

$Bel(R)$  will be increased (decreased) by conditioning evidence that implies material misstatement (fair presentation).

Accordingly, revisions in cells A and C were predicted to be negative (conditioning evidence suggested fair presentation) and those in cells B and D to be positive (conditioning evidence suggested material misstatement). Thirty-four risk revisions were positive, 19 were negative and 2 were unchanged. Positive revisions were coded as successes in Cells B and D and misses in Cells A and C. Negative revisions were coded as misses in Cells B and D and successes in Cells A and C. As the 0 revisions occurred in cells A and C, both were coded as positive to provide a conservative test and therefore Table 7 indicates 36 positive and 19 negative revisions. The value for the null hypothesis ( $\theta_{risk}$ ) specified in section 3.4 is the proportion of the more frequent sign of revision. Therefore the test is:

$$H_{1_0}: \text{Success}_{\text{risk}} \leq 65.4\% \quad H_{1_a}: \text{Success}_{\text{risk}} > 65.4\%$$

Table 7  
Sign of Risk Revision by Cell

Sign/Cell	A	B	C	D	n	%
-	<u>6</u>	1	<u>9</u>	3	19	34.6
+	7	<u>17</u>	3	<u>9</u>	36	65.4
n	13	18	12	12	55	
(%)	23.6	32.7	21.8	21.8		100

Successes=41 (74.6%), misses=14 (25.4%)  $p \leq .0996$ .  
Two 0 revisions, one each in cells A and C, were coded as misses (+)

Forty-one of 55 revisions (74.6%) were correctly predicted by H1. The likelihood that the superior explanatory power of H1 was due to chance is  $\leq .0996$ . Of the 30 subjects who received inconsistent evidence (Cells B and C), 26 revised risk as predicted. The results with respect to the effect of consistent evidence on risk revision were somewhat different. In Cell D, 9 of 12 subjects receiving information consistent with a prior expectation of error increased their risk assessments; in Cell A only 7 of 13 decreased Bel(R). The SMALL ERROR treatment may have had the unintended effect of increasing Bel(R) for some subjects.

### Test of Sufficiency Revision

The sign of evidence sufficiency revision was determined by subtracting the initial sufficiency judgment (S1) from the revised judgment (S2). Hypothesis 2 states that:

Bel(S) will be increased by credible conditioning evidence unless such evidence is inconsistent with prior evidence indicating fair presentation.

Accordingly, sufficiency revisions were expected to be positive in all cells, except Cell B (conditioning evidence inconsistently suggested material misstatement). Thirty-three revisions were positive, 21 were negative and 1 was unchanged. Positive revisions were coded as successes in Cells A, C, and D and misses in Cell B. Negative revisions were coded as misses in Cells A, C, and D and successes in Cell B. As the 0 revision occurred in cell D, it was coded as negative to provide a conservative test and therefore Table 8 indicates 60% positive and 40% negative. The value for the null hypothesis specified in section 3.4 is the proportion of the more frequent sign of revision. Therefore the test is:

$$H_{2o}: \text{Success}_{\text{suff}} \leq 60\% \quad H_{2a}: \text{Success}_{\text{suff}} > 60\%$$

Forty-three revisions (78%) were correctly predicted by H2 with Bel(S) increasing for 29 of 37 subjects in Cells A, C, and D, and decreasing for 14 of 18 subjects in cell B. The likelihood that the superior explanatory power of H2 was due to chance is  $\leq .0045$ .

Table 8  
Sign of Sufficiency Revision by Cell

Sign/cell	A	B	C	D	n	%
-	5	<u>14</u>	1	2	22	40.0
+	<u>8</u>	4	<u>11</u>	<u>10</u>	33	60.0
n	13	18	12	12	55	
%	23.6	32.7	21.8	21.8		100
<p>Successes=43 (78.2%), misses=12 (21.8%) <math>p \leq .0045</math>            A 0 revision in cell D was coded as negative.</p>						

The two alternative explanations for sufficiency revision discussed in Section 3 are not supported by the evidence. The first alternative held that sufficiency assessments monotonically increase as credible evidence is collected and are not affected by consistency. Because the more frequent sign of revision was positive, this explanation has already been tested and rejected as the null form of H2. Restricting analysis to the cells receiving inconsistent results (Cell B and Cell C), 25 of 30 subjects (83%) revised as predicted ( $p \leq .0003$ ). H2<sub>a</sub> clearly provides a better explanation of sufficiency revision than an account based solely on evidence quantity and credibility. The second alternative argued that inconsistent conditioning evidence would lead to sufficiency decreases regardless of whether prior evidence implied M or  $\sim$ M. This explanation also results in only 33 successes (60%), since 10 fewer successful predictions are achieved in Cell C (the successes and misses are reversed). The success rate achieved by H2 is again significantly greater ( $p \leq .0045$ ).

### **Joint Test of Audit Risk and Evidence Sufficiency Revision**

A crosstabulation of the signs of revision of Bel(R) and Bel(S) was computed to jointly test H1 and H2. As shown in Table 9, panel A (below), the most common pairing was a decrease in Bel(R) and an increase in Bel(S) which occurred 19 times (34.5%). Of the 41 subjects who revised audit risk in accordance with H1, 33 revised evidence sufficiency as predicted by H2. Therefore, as reported in Table 9, panel B, thirty-three of 55 (60%) joint revisions were correctly predicted by the hypotheses. Using .345 as the expected proportion of successes and .60 as the observed proportion, the likelihood that the

superior explanatory power of the model is due to chance is  $\leq .0001$ . It should be noted that only 3 subjects decreased risk assessments, while also decreasing sufficiency assessments, a combination of beliefs that is irreconcilable with the interpretation of evidence sufficiency developed in Chapter 3.

Table 9  
Joint Test of Hypotheses 1 and 2

Panel A Revision of Bel(R) and Bel (S)						
Audit Risk	Sufficiency		n			
	+	-				
+	15	19	34			
-	18	3	21			
n	33	22	55			
proportion of modal pairing = $19/55 = .345$						
Panel B Successes by Cell						
Cell	A	B	C	D	n	%
Successes	6	13	7	7	33	60.0
Misses	7	5	5	5	22	40.0
n	13	18	12	12	55	
%	23.6	32.7	21.8	21.8		100
Successes = 33 (60%), misses= 22 (40%) $p \leq .0001$						

The binomial tests provide support for H1 and H2. The sign of risk revision was, as expected, a function of whether audit procedures indicated the absence or presence of error. Sufficiency assessments increased unless results unexpectedly suggested a large error when initial evidence was inconsistent with fair presentation. This result supports the

argument that conditioning evidence increases the second-order uncertainty of auditors only when inconsistent with prior evidence implying fair presentation.

### 5.3 **Other Results**

This section reports additional analysis based on mean revisions, as well as the results obtained from the flowchart. The usefulness of the present research depends, on whether the hypotheses might plausibly explain important aspects of the opinion formulation process. Although sufficiency revision occurred as predicted, if revision magnitude was slight, it would be less likely that evidence consistency plays an important role in second-order uncertainty reduction. Similarly, the flowchart was intended to link inconsistent evidence with the decision to extend or alter audit procedures.

#### **Additional Tests of Audit Risk and Evidence Sufficiency Revision**

Tests of the magnitude of mean risk revision (Table 10) provide additional support for H1.<sup>18</sup> Inconsistent conditioning evidence (Cells B and C) led to large revisions in the predicted directions, whereas consistent evidence (Cells A and D) failed to induce significant change. The pattern is similar to those obtained by Ashton and Ashton (1988) and McMillan and White (1993).

---

<sup>18</sup>Analysis of the 3-point ordinal risk measure resulted in similar, but somewhat weaker conclusions.

Table 10  
Mean Risk Revision by Cell

		SMALL ERROR	LARGE ERROR
		A	B
LOW PRIOR RISK	expected sign	±	+
	mean revision	-0.77	+ 26.28
	std	20.36	24.18
	n	13	18
	t-stat	-0.14	4.61***
		C	D
HIGH PRIOR RISK	expected sign	-	±
	mean revision	-25.08	+3.92
	std	38.44	24.57
	n	12	12
	t-stat	-2.26**	0.55
**=p ≤ .05, ***=p ≤ .005.			

Hypothesis 2 predicts that sufficiency assessments will increase in all cells except for Cell B where LOW PRIOR is crossed with LARGE ERROR. As predicted, Cell B shows a negative revision, whereas all other cells means were positively revised. However, as reported in Table 11 (below), only 2 of 4 revisions were statistically significant.

The prediction of a negative revision for Bel(S) in Cell B was conditioned on the effect of inconsistent implications being sufficient to negate the added quantity and credibility of evidence provided by the vouching procedure. The inconsistency between prior and conditioning evidence may have been too weak to overcome the increment in quantity and credibility. In that case the Cell B revision should have still been the lowest of the four cells. The mean of Cell B appears significantly less than that of Cell A based on a

single comparison ( $p \leq .05$ ). However, since the comparison between Cells A and B was suggested by the data, rather than planned in advance, a Bonferroni adjustment (Neter, Wasserman, Kutner 1985) based on a family of three comparisons (B/A, B/C, B/D) is necessary to avoid overstating significance. After adjustment the difference is no longer significant.<sup>19</sup>

Table 11  
Mean Sufficiency Revision by Cell

		SMALL ERROR	LARGE ERROR
		A	B
LOW	expected sign	+	-
PRIOR	mean revision	+ 7.31	-9.06
RISK	std	16.49	33.69
	n	13	18
	t-stat	1.60*	1.14
		C	D
HIGH	expected sign	+	+
PRIOR	mean revision	+45.50	+38.00
RISK	std	36.36	29.85
	n	12	12
	t-stat	4.34***	4.41***
*= $p \leq .10$ , **= $p \leq .05$ , ***= $p \leq .005$ .			

In examining the cell means on a number-line (Figure 3a below), it is apparent that the Cell A revision is least compatible with the expected results. The mean sufficiency revision in Cell A was only +7 points, as compared to increases of +38 and +46 for Cells C

<sup>19</sup>Assuming homogeneity of variance. Tests based on unequal cell variances and those based on median revision yield only slightly stronger results and are not reported.

and D. As a result, the Cell A mean is closest to that of Cell B. Had the Cell A revision had been similar to that of Cells C and D (Figure 3b), the pattern of revision would have supported the conclusion that sufficiency increases are least when conditioning evidence is inconsistent with prior evidence suggesting fair presentation. Had Cell B shown the expected negative revision (Figure 3c), the lack of impact of the vouching procedure for Cell A, would still be puzzling. One possibility is that the SMALL ERROR difference was viewed as significant by some subjects and therefore perceived as inconsistent with the implications of prior evidence.

Attributing the weak Cell A revision to a design flaw is more appealing than the alternative hypothesis that auditors did not consider results consistent with initial evidence indicating low audit risk to provide competent evidence. However, neither explanation adequately explains the results. The analysis of mean sufficiency revision provides only weak support for H2, suggesting the need for further research focusing on factors affecting sufficiency revision magnitude.

Figure 3  
Comparison of Actual and Hypothetically Adjusted Means for Sufficiency Revision

(a) Results actually obtained					
	B	A		D	C
	-9	0 +7		+38	+46
(b) Cell A adjusted to conform to Hypothesis 2					
	B			D	A C
	-9	0		+38	+46
(c) Cell B adjusted to conform to Hypothesis 2					
B		A		D	C
		0 +7		+38	+46

### Analysis of Flowchart

The flow chart required the subjects to indicate likely actions that might result from evaluating conditioning evidence by indicating whether they expected to increase sample size, book the actual or projected difference, or revise the audit plan in related areas. The main purpose was to provide insight into whether the subjects in Cell B were more likely to believe that it would be necessary to extend diagnosis by sampling additional invoices. Subjects in Cells A, C and D were predicted to believe that a conclusion could be reached without sampling additional invoices. As reported in Table 12, 35 of the 49 subjects who properly completed the flowchart (71%), expected to extend sampling. Fifteen of 16 subjects in cell B expected to extend sampling, whereas 20 of the remaining 33 subjects so

indicated. That a majority of respondents in all cells believed that sampling would be extended suggests that a demand effect was created.

Since the responses correctly predicted (57%) were lower than the MLE (71%), the conjecture that only subjects in Cell B would expect to extend sampling can be rejected. Yet, inspection of cell frequencies suggests that subjects in the Cell B were more likely than those in Cells A, C, or D to expect to increase sample size. A Chi-square test rejects the equal frequency null at a marginal level of significance (CHISQ=6.99, df=3,  $P \leq .07$ ) with the result largely attributable to Cell B. Nonetheless, the link between unexpected differences and extended diagnosis is weak.

Table 12  
Expectation of Increased Sample Size by Cell

	A	B	C	D	n %
Yes	6	15	6	8	35 71.4
No	5	1	5	3	14 28.6
n	11	16	11	11	49
%	22.44	32.65	22.44	22.44	100
Chi-square = 6.99, df = 3, $p \leq .07$					

Two additional issues concerning the audit implications of the vouching procedure were also examined. SAS-53 suggests that the detection of errors and irregularities may lead to revision of the audit program in related audit areas. The correlation of risk factors across audit areas is not well understood, although Waller (1993) reports that same assertion/different area assessments of inherent and control risk display low correlations

during planning. Because subjects in Cells C and D were provided with information suggesting possible problems in other areas as part of the PRIOR RISK, manipulation analysis is restricted to Cells A and B. If subjects revise planned procedures in related areas or assertions due to unexpected errors, those in Cell B would be more likely than those in Cell A to select a node indicating such a modification. The results weakly support such a conclusion; 2 of 11 Cell A subjects wished to modify the audit program in related areas as did 7 of 16 in Cell C (CHISQ=1.9,  $p < .17$ ).

Auditors are required to project sample differences to account balances when considering the significance of detected errors, but determining when an adjustment is necessary is a matter of professional judgment (AU-350). Since the flowchart did not indicate either the additional procedures performed or results obtained, it is impossible to determine whether conformity with GAAS would require adjustments based on the projected differences calculated in the case. However, 18 of 27 subjects in the LARGE ERROR condition considered it most likely that it would be necessary to propose an adjustment for the projected difference, whereas 8 of 22 in the SMALL ERROR condition so concluded (CHISQ=4.7,  $p \leq .04$ ). This suggests that the subjects employed a strategy of disregarding clearly immaterial differences (even when projected), focusing attention on those which might be of audit significance. This finding might affect the conclusions reached by Burgstahler and Jiambalvo (1986) who report that auditors tend to isolate rather than project "atypical" errors. In the absence of guidelines concerning tolerable misstatement, the subjects in that study may simply have assumed that the detected errors

were insignificant.

#### **5.4 Manipulation Effectiveness**

Four questions included in the exit questionnaire (See Appendix) were intended to examine the efficacy of the experimental manipulations. Two responses were elicited to determine whether perceptions of control risk and evidence credibility were similar across conditions. These features of the case were not varied and should not have been viewed differently across conditions. Because responses may have been affected by hindsight bias or referrals back to the case materials, the manipulation checks may reveal threats to validity, but do not provide assurance supporting the validity of the experiment. Exit questions 1-4 pertain to the first treatment variable, PRIOR RISK. Questions 1, 3 and 4 were intended to check the effectiveness of the PRIOR RISK treatment and Question 2 provided evidence as to whether subjects viewed control risk as a constant across conditions.<sup>20</sup> Questions were coded so that positive correlations denote agreement between the intended and perceived treatment. Stated probabilities are one-tailed since a priori expectations existed concerning the sign of association. Table 13 displays the Spearman correlation coefficients between PRIOR RISK and subject responses. The Spearman

---

<sup>20</sup>To test for hindsight bias Spearman correlations between ERROR SIZE and responses to Questions 1-4 were computed. Only for question 1 did responses seem to have been significantly colored by subsequent information.

correlation is an appropriate statistic for analysis, as responses were measured on ordinal scales.

Table 13  
Association Between PRIOR RISK and Exit Questions

		Inherent Risk	Control Risk	Sales Change	Other Problems
Prior Risk	Spear-r	.1939	.0473	.4336	.3655
	N	55	54	54	54
	$p \leq x$	.078	.367	.001	.003

**Question 1** checks whether the word pairs intended to induce higher or lower inherent risk assessments had the desired effect. The manipulation was moderately effective (Spearman  $r=.1939$ ,  $p \leq .078$ ). Cross-tabulation shows that only 9 subjects responded in a manner clearly inconsistent with the research intent, recording low inherent risk in the HIGH PRIOR cell or high inherent risk in the LOW PRIOR cell.

**Question 2** required subjects to assess control risk as low, medium, or high. Information useful in assessing control risk was not varied and therefore subject judgments should have been uncorrelated with treatment condition. Also, the information was intended to avoid maximum assessments even though controls were not being tested. The differences between PRIOR conditions were not statistically significant (Spearman  $r=.0473$ ,  $p \leq .367$ ). Cross-tabulation shows that only 8 subjects assessed control risk as high. Incidental to the research purpose, but bearing on recent work by Waller (1993) was an observed correlation between control risk and inherent risk assessments (Spearman  $r=.49$ ,  $p \leq .001$ ). The

relationship was not weakened by first accounting for the effects of the PRIOR treatment. This supports Waller's hypothesis that inherent and control risk assessments are not independent.

**Question 3** required subjects to report whether the comparative income statement had alerted them to a possible sales error. 39 of 54 subjects responded as intended (Spearman  $r=.4336$ ,  $p\leq.001$ ), indicating an effective manipulation.

**Question 4** asked subjects to indicate if the case materials had alerted them to possible errors in other audit areas. Although up to three possible areas could be checked-off, the variable was analyzed dichotomously (0=no problem areas, 1=1 or more problem areas). Thirty-five of 54 subjects responded as intended by identifying possible problem areas in the HIGH PRIOR condition or by indicating that they were not alerted to other problem areas in the LOW PRIOR condition (Spearman  $r=.3655$ ,  $p\leq.003$ ).

**Question 5** required subjects to assess the strength of the vouching procedure. If subjects viewed vouching as a weak procedure, the experiment would not be a valid test of the hypotheses since conditioning evidence with low credibility would be unlikely to affect either first-order or second-order uncertainty. Forty-nine of 55 subjects believed the procedure to be moderately (22) or highly effective (27).

**Question 6** tested the effectiveness of the second manipulation (ERROR SIZE) by asking whether the audit difference was less than or greater than tolerable error. Forty-three of 49 subjects responded correctly (Chi-Square=28,  $p\leq.001$ ). It is possible that some of the incorrect or nonresponses were due to uncertainty as to whether detected or projected

differences should be compared to tolerable error. Failing to project or incorrectly recalling tolerable error should weaken the power of the tests, but should not produce inconsistent risk and sufficiency assessments.

The exit questions provide no reason to doubt that the subjects perceived the manipulations as intended. Subjects indicated that they were sensitive to sales fluctuations and problems in other audit areas, but less attentive to variation in inherent risk factors. Control risk assessments, as planned, did not vary across conditions. The vouching procedure was rated as moderately or highly effective by 88% of responding subjects and 89% correctly recalled whether the reported error was above/below tolerable error. Thus, the case materials appear to have provided a valid context for testing the experimental hypotheses.

## 5.5

### **Demographic Variables**

This section examines the distribution of potentially significant demographic characteristics across conditions and then reports their influence on the dependent variables. The characteristics examined include age, gender, firm affiliation, and several attributes which measure experience. The results indicate that: (1) the assumption of homogeneous distribution of subjects across treatment groups need not be rejected and; (2) only gender had an important effect on the revision of risk assessments.

### Homogeneity of Subject Attributes Across Treatment Conditions

To test the assumption that subjects were homogeneously distributed across treatment conditions, the appropriate statistics were computed for categorical, ordinal and continuous variables. As there were between 5 and 15 nonresponses to each question, analysis is based on a differing subset of the subjects. The equal distribution or no association null hypotheses are rejected at  $p \leq .10$  for any variable (Table 14), and severe confounding between demographic and experimental variables can be ruled out.

Table 14  
Distribution of Demographic Variables Across Treatments

Panel A Chi-Square Statistics				
	Prior		Error	
	Chi-sq	$p \leq x$	Chi-sq	$p \leq x$
Gender	.08	.77	.50	.48
Firm (A or B)	.04	.85	.20	.66
CPA	.07	.80	1.30	.25

Panel B Correlations				
	Prior		Error	
	r	$p \leq x$	r	$p \leq x$
Age	.02	.87	-.01	.99
Experience	-.01	.97	.03	.82
Recent Experience	.21	.13	.02	.89
Firm tenure	.11	.43	-.21	.13
Firm level	.04	.39	.19	.10

### Association Between Demographic Attributes and Dependent Variables

Pairwise correlations (See Table 15) were computed to test whether demographic characteristics influenced audit risk or evidence sufficiency revision. Any observed relationships must be cautiously interpreted unless supported by theory. Only the correlation between risk revision and gender ( $r=.35$ ,  $p\leq.013$ ) was sufficiently large to warrant extended analysis.

Table 15  
Correlations between Demographic and Dependent Variables

Variables	Risk Revision		Sufficiency Revision	
	r	$p\leq x$	r	$p\leq x$
Firm (A or B)	-.01	.93	.01	.94
Age	-.15	.32	-.20	.18
Gender	.35	.01	-.18	.21
Experience	-.05	.70	-.22	.12
Firm tenure	-.29	.04	-.02	.90
CPA	.11	.45	-.16	.26
Firm level	.01	.97	-.01	.95
Recent Experience	.01	.94	.03	.86

**Firm affiliation** was a plausible source of between subject variability since the audit technology and hiring practices of the two firms might produce significant differences in risk

and sufficiency assessments. Twenty six subjects were recruited from firm A and 29 from firm B.<sup>21</sup> Firm membership did not correlate with either dependent variable.

Except for gender, the remaining demographic variables measure experience and possibly expertise. Age and experience showed slight negative correlations with sufficiency revision (experience=-.20,  $p \leq .18$ ; age=-.22,  $p \leq .12$ ). Omitting subjects with less than 1 year of experience from the analysis does not affect the results reported in Sections 5.2 and 5.3. Subjects had a mean **firm tenure** of 6 years with their current employer and a range from 0 to 32. A significant association exists between firm tenure and risk revision ( $r = -.29$ ,  $p < .04$ ). Firm tenure may be a good measure of task related experience because it proxies for experience while eliminating extreme values. Subjects with 1, 5, or 10 years of experience may well form differing risk assessments, but, there may be no further change as experience further increases.

**Certification** and **firm level** are formal acknowledgements of achievement and may represent expertise as of the date bestowed. Recency of audit experience may indicate current level of expertise for experienced auditors. Fifty-eight percent of the subjects reported being **CPA's** and a low correlation exists between certification and sufficiency revision ( $r = -.16$ ,  $p < .26$ ). Subjects reported their **firm level** as: juniors 8, semi-seniors 5, seniors 17, managers 13, partners 5, no response 7. Firm level was uncorrelated with the dependent variables. Juniors displayed greater risk revision and less sufficiency revision than

---

<sup>21</sup>There was no question eliciting firm affiliation. Firm membership was coded by numbering completed instruments from Firm A 1-26 and Firm B 27-55.

other subjects, but the differences were not statistically significant. Results were unchanged when juniors were omitted from the model. **Recent Experience** was defined as months assigned to auditing in the last 2 years. Nine experienced subjects had no recent audit experience, but data fit was again not improved by their omission from analysis.

Seventeen responding subjects were female and 33 male. The correlation between **gender** and risk revision was .35 ( $p \leq .01$ ). Decomposition (Table 16) shows that women recorded lower prior, but similar posterior assessments, as compared to men. The mean prior risk assessment was 20 points higher for men, due largely to differences in the HIGH PRIOR condition. Mean revision was -5 for men and +20 for women, resulting in similar posterior assessments. As age and experience were correlated with gender in the subject population, it was thought that they might explain the difference in initial risk assessments. However, their inclusion with gender and PRIOR RISK in an ANCOVA model did not improve fit or decrease the gender effect. The manipulation checks, when reperformed by gender, did not reveal any significant differences in treatment effectiveness between men and women.

Table 16  
Decomposition of Risk Assessments by Gender

	R1	R2	R2-R1
Female (n=17)	33	53	20
Male (n=33)	53	48	-5
Mean Differences	-20	5	25

### Summary of Analysis of Demographic Variables

The study of the demographic variables indicates that the subjects did not significantly differ across treatment groups with regard to the attributes measured and that therefore no important confounding is likely to exist. Of the demographic variables analyzed, only gender and firm tenure were significantly correlated with either risk or sufficiency revision. No basis existed for expecting a gender effect and no persuasive explanation for its existence is apparent. Several variables that might proxy for expertise, including age, experience, certification and firm level, show only low correlations with the dependent variables.

## 5.6

### Discussion

The main purpose of the present study was to develop and test hypotheses based on Waller and Felix's (1984) assertion that evidence drives audit risk and evidence sufficiency judgments both jointly and severally. The research was also expected to contribute to the second-order uncertainty literature by identifying characteristics of conditioning evidence

which might cause decreases in the confidence of task-experienced cognizers. The present section considers the extent to which these objectives have been achieved, discusses the implications for future research, and concludes the monograph.

### **Summary of Results**

Chapter 2 reviewed cognitive research in the domains of psychology and auditing which suggested that the consistency of evidence is likely to be a common determinant of both first-order and second-order uncertainty. Lessened consistency, due to conflicting conditioning information, was identified as a potential cause of decreases in confidence across time. Chapter 3 examined the possible effects of evidence consistency on first-order and second-order judgments in the audit domain. The hypotheses stated that: (1) audit risk judgments are revised in accordance with the implications of conditioning evidence; and (2) evidence sufficiency judgments are increased by credible conditioning evidence unless such evidence is inconsistent with prior evidence indicating fair presentation. The second hypothesis differs from that suggested by the cognitive literature, predicting that the effect of inconsistent conditioning information on evidence sufficiency judgments is asymmetrical due to unique aspects of the audit task.

The binomial tests provided evidence in support of the joint hypotheses. Seventy-five percent of subjects revised  $Bel(R)$  in the direction specified by Hypothesis 1 ( $p \leq .1$ ) and 78% in the direction specified by Hypothesis 2 ( $p \leq .005$ ). Sixty-percent of the subjects jointly revised both assessments as predicted ( $p \leq .0001$ ). Of the subjects receiving

inconsistent conditioning evidence, 83% revised sufficiency assessments as predicted ( $p \leq .0003$ ). Evidence sufficiency assessments did not either monotonically increase with the execution of audit procedures, nor necessarily decrease when inconsistent conditioning evidence was evaluated. The proposition that evidence inconsistency increases second-order uncertainty (Ellsberg 1961, Smith, Benson, and Curley 1991) appears to apply in the audit environment only when initial evidence implies fair presentation.

The results obtained for mean audit risk revision were similar to those reported by Ashton and Ashton (1988) and McMillan and White (1993). They do not support the existence of a confirmation bias, or conservatism in the downward revision of audit risk judgments. The tests of means provided mixed results with respect to sufficiency revision; the signs were as predicted, but the magnitudes were not significant in all cells. The weak positive revision recorded by subjects who received consistent prior and conditioning evidence implying fair presentation of sales is puzzling and requires further explanation.

The flowchart yielded weak results linking decreased evidence consistency to the decision to diagnose cause. Potentially material audit differences that had not been signaled by prior evidence were the most likely to induce an expectation of expanded sample sizes. Evidence consistency had no effect on expectations concerning revision of the audit program in related areas. Auditors may not be evaluating the impact of unexpected misstatements on other areas in accordance with AU-316.

In sum, the empirical results generally support the experimental hypotheses. The unique characteristics of the audit environment may cause unexpected differences to be

perceived as increasing second-order uncertainty, triggering diagnosis and the possible extension of audit procedures, whereas the absence of expected differences appears less likely to induce further investigation.

### **Limitations and Implications**

The conjecture that second-order uncertainty is principally determined by the amount, credibility, and consistency of evidence has been repeated in varying forms since Ellsberg (1961). Although Einhorn and Hogarth (1985a) and Bedard (1992) conducted experiments involving all three attributes of evidence, their research did not directly measure second-order judgments. Because the present study was aimed at elucidating the joint pattern of audit risk and evidence sufficiency assessment, only the effect of consistency on second-order judgment was investigated. Future research should examine the combined effect of evidence quantity, credibility, and consistency on audit risk and evidence sufficiency assessments to provide a more complete understanding of the nature of the interaction.

Such an extension would address a limitation of the present research. The asymmetrical effect of inconsistent conditioning evidence on evidence sufficiency revision (Section 3.3) was conjectured to exist due to an inverse relationship between  $Bel(R_p)$  and the amount and credibility of evidence that the auditor expects to obtain. Although not apparent to subjects, due to the between-subject design, the amount and credibility of

evidence was constant in the experiment. Thus, the asymmetrical revisions obtained cannot be conclusively attributed to the postulated cause.

Further research will also be needed to determine if negative, or lower than expected sufficiency revisions, do indeed prompt diagnosis. Such evidence would provide enhanced understanding of what is perhaps the most important part of the diagnostic process, determining which phenomena require causal explanation. Increases in sufficiency assessments may explain the decision to accept an hypothesis as the cause of an unexpected finding, thereby terminating diagnosis.

The research design did not permit inference as to which aspects of the PRIOR treatment induced the differing assessments of R1 and S1. The case was designed to include a variety of cues consistently suggesting within a condition either fair presentation or material misstatement, so that if subjects discounted a particular cue the entire treatment would not be negated. Since initial sufficiency assessments differed by 30 points, determining the relative contribution of the cues would contribute to understanding the domain-specific factors that affect auditors second-order uncertainty judgments.

An unexpected result was the effect of gender on mean audit risk revision. Lower initial assessments of Bel(R) for women, in both the HIGH PRIOR and LOW PRIOR treatments, led to much greater positive revisions for those subsequently assigned to the LARGE ERROR condition. A review of the gender literature did not reveal any studies of differences in risk assessments between men and women. Whether the effect is specific to the group studied or generalizable to other practice units is indeterminable. If gender based

differences in planning risk assessments do exist, their possible implications for audit practice should be investigated

Differing viewpoints exist as to why cognizers are averse to second-order uncertainty. Smith, Benson, and Curley (1991) and Frisch and Baron (1988) view second-order judgments as improving decision quality by avoiding choices based on inadequate information. But Brown (1990) claims that without accountability, second-order uncertainty does not affect choice. However, little empirical evidence exists to support either position. Messier and Quilliam's (1992) propositions concerning the impact of the locus of accountability and sanction avoidance on the complexity of auditor's cognitive strategies may also help explain differences in aversion to second-order uncertainty. Research which investigates whether auditor attention to second-order uncertainty is affected by the likelihood and source of sanctions would provide further insight into the causes of ambiguity avoidance.

Recent investigation has focused on several factors besides the effect of initial anchors and the consistency of conditioning evidence on audit risk assessments. For example, Koonce (1992) investigated the effects of explanation and counter-explanation on belief and McMillan and White (1993) have studied whether professional skepticism affects the magnitude of revision depending on whether subjects began with an error or fair

presentation frame. Future research involving evidence sufficiency judgments should also incorporate greater contextual and task complexity.

### **Conclusion**

Because auditors must attend to the adequacy of the examination as well as the correctness of the opinion, it is likely that evidence sufficiency judgments are an important part of the opinion formulation process. This research was undertaken to provide insight into the role that evidence consistency plays in the revision of second-order uncertainty judgments. The results support the argument that the unique characteristics of the audit environment cause unexpected differences to be perceived as increasing second-order uncertainty, perhaps provoking diagnosis and the extension of audit procedures, whereas the absence of expected differences does not prevent increases in evidence sufficiency assessments. The consideration of evidence consistency has shed insight into the joint and several relationship between audit evidence and risk and sufficiency assessments posited by Waller and Felix (1984).

## APPENDIX A: RESEARCH INSTRUMENT

### INSTRUCTIONS

Thank-you for your participation in this research project. All responses are anonymous, so do not write your name in this booklet. The purpose of the research is to learn about auditors' evaluation of audit evidence. The experiment requires that you read through a brief case describing a hypothetical audit. As on an actual audit some data may support an opinion that the financial statements are fairly presented and some data may suggest that material errors exist. Different participants will receive different information. As you read the case you will be asked to make two types of judgments concerning an account balance. The following paragraphs describe these judgments.

The auditor's task is to form an opinion concerning the financial statements under examination. In doing so the auditor considers the possibility that errors exceed "tolerable error" the maximum amount that can exist for individual accounts or classes of transactions without requiring a modification of the opinion. **You will be asked to assess the risk that an account balance contains errors in excess of tolerable error.** This assessment will be made in two ways. First, you will assess risk as low, medium or high. Second, you will assess risk on a 100 point scale.

The third standard of fieldwork states that, "sufficient competent evidential matter is to be obtained... to afford a reasonable basis for an opinion regarding the financial statements under examination." Issuing a correct opinion based on insufficient evidence is a violation of GAAS. Auditors consider the sufficiency of evidence when planning and evaluating the results of audit procedures and before issuing an opinion. **You will be asked to judge whether sufficient evidence has been obtained to justify your assessment as to whether errors in excess of tolerable error exist.** In sum, you will make both an audit risk judgment and a judgment regarding the adequacy of the evidence used for the audit risk judgment.

The case begins on the next page. The information presented is much less than would be available for an actual audit. However, you are asked to imagine that it describes an actual audit and to make the best responses given the available data. The schedule below lists the case materials provided.

1. Background information on client.
2. Comparative income statements.
3. Results of significant audit procedures completed.
4. Assessments 1 and 2.
5. Description of an additional procedure to be performed.
6. Result of the additional procedure.
7. Assessments 3, 4 and 5.

The materials should be read in order of presentation. You may refer back to earlier sections of the case to refresh your memory, but do not change responses once they have been made. There is a brief questionnaire following the case. Once again, thank-you for your participation.

PLEASE TURN TO THE NEXT PAGE

## CASE

As in-charge accountant for a public accounting firm you have been assigned to the annual audit of ABC Inc. for the year ended 12/31/X1. ABC is a small privately held company which produces subcomponents for refrigerator manufacturers at a single plant in New York City. ABC's profits have been average for the industry and cash flows have been adequate to sustain debt-servicing, dividends and new investment. The industry has been stable and moderately profitable with few new firms entering the market and few bankruptcies. The top managers of ABC are all experienced in the industry. The president has held office for 12 years, the treasurer for 3 years and the controller, who is a CPA, for 7 years. Management has a conservative attitude towards financial reporting and does not appear to place undue emphasis on meeting earnings projections. ABC has on occasion requested assistance in accounting for complex transactions and when implementing new accounting standards. No material adjustments were made during the prior year's audit. Your firm has audited ABC for 8 years during which time all opinions have been unqualified.

Fieldwork began on December 27 and is expected to last through January 31. The tentative date for delivering the opinion to the client is February 7. Evidence is being obtained almost entirely from substantive procedures with little or no reliance being placed on internal controls. This decision was based primarily on audit efficiency considerations. The study of internal control indicates that management is aware of the importance of control and that the philosophy and operating style of the organization encourages adherence to prescribed procedures. The accounting system is properly designed to record the firm's transactions and control procedures exist to detect errors should they occur. In planning the audit, tolerable error for individual account balances has been set at \$300,000. A comparative condensed income statement appears below.

INCOME STATEMENT	AUDITED	UNAUDITED	
000's omitted	19X0	19X1	% CHANGE
SALES	75361	83598	10.9
COST OF GOODS SOLD	45094	50185	11.3
GROSS PROFIT	30267	33413	10.4
EXPENSES:			
SELLING	11766	13183	12.0
GENERAL & ADMINISTRATIVE	5640	6027	6.9
OPERATING INCOME	12861	14203	10.4
GROSS PROFIT %	40.2	40	

PLEASE TURN TO THE NEXT PAGE

The current date is now January 11, 19X2. Audit procedures have been performed with the following results:

1. Cash balances have been confirmed. Adjustments were made for routine items such as unrecorded bank charges.

2. Confirmations for receivables have been sent and exceptions are being investigated. The confirmation response rate has improved compared to last year and exceptions have decreased.

3. Cutoff for sales and purchases was reviewed and appeared proper.

4. The inventory count was observed and appeared accurate with respect to quantities on hand. Pricing tests are in progress.

The next area to be tested is sales revenue. The study of internal control in the revenue cycle indicates that the accounting system is adequately designed to identify and record on a timely basis all valid transactions at their proper value. Procedures exist to ensure that transactions are properly authorized and documented and that segregation of key duties occurs. As mentioned above, tests of controls were not performed so their effectiveness is unknown. Please make the judgments indicated below. You may review the previous pages before responding.

---

ASSESSMENT 1

19X1 sales revenue may be fairly presented or contain errors in excess of tolerable error (\$300,000).

Circle the risk level that best represents your assessment of the risk that sales revenue is **overstated** by an amount greater than tolerable error.

LOW

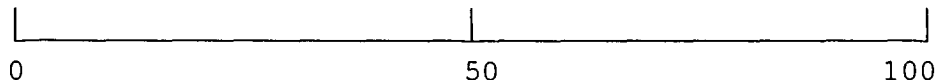
MODERATE

HIGH

Mark an "X" at the point on the 100 point scale that best represents your assessment of the risk that sales revenue is **overstated** by an amount greater than tolerable error.

(no risk)

(maximum risk)



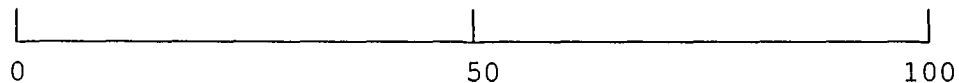

---

ASSESSMENT 2

Mark an "X" at the point on the 100 point scale that best represents your belief as to whether you already have sufficient evidence (As contemplated by the third standard of fieldwork) to justify Assessment 1.

the evidence  
is definitely  
not sufficient

the evidence  
is definitely  
sufficient



PLEASE TURN TO THE NEXT PAGE

The audit program requires that the following procedure be performed to test that recorded sales actually occurred and were accurately recorded:

Vouch sales from sales journal to the matched file containing shipping reports, customer orders and duplicate invoices. Compare invoice details with price lists and verify extensions and totals.

Review of the prior year's work papers reveals that there were no errors noted due to the recording of nonexistent sales. Accuracy errors were insignificant and adjustments were waived. A judgmental sample of 200 items has been selected for this year's testing. Individually significant invoices (over \$100,000) are being audited on a 100% basis.

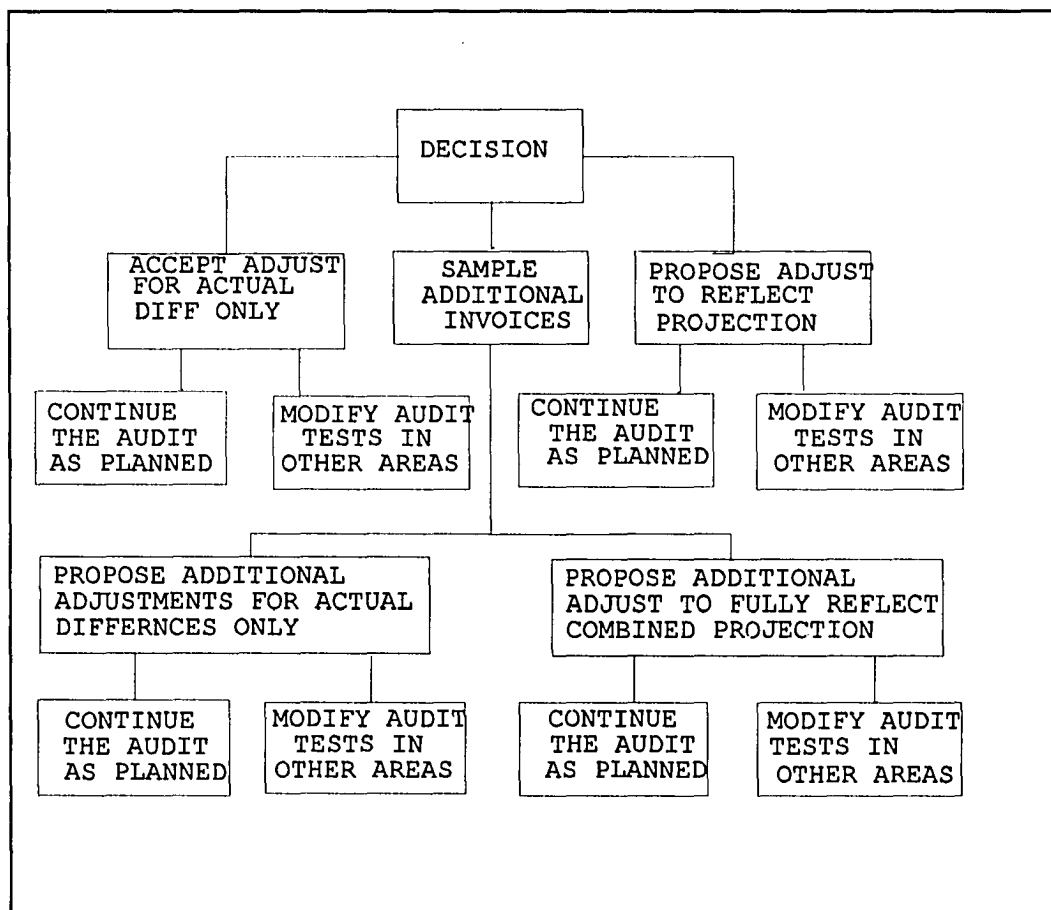
Invoices selected for testing	# items	Sample Tested	Sales Revenue
Below \$100,000	173	\$10,867,497	\$76,072,514
All invoices above \$100,000	<u>27</u>	<u>7,525,266</u>	<u>7,525,266</u>
Totals	200	\$18,392,763	\$83,597,780
% of sales revenue to be tested		22%	

PLEASE TURN TO THE NEXT PAGE



## ASSESSMENT 5

After completing an audit procedure the auditor must consider his or her next action. Assume that the client has been advised of the differences detected. The controller agrees to adjust sales revenue for the actual differences, but believes that the sample was not representative and that therefore an adjustment to reflect the projected difference is unwarranted. A flow chart representing some possible action sequences is presented below. The terminal nodes have been numbered 1-8. Mark an "X" through the node at the end of the sequence which best represents the steps that you would consider necessary to perform in order to comply with GAAS.



PLEASE TURN TO THE NEXT PAGE

**QUESTIONNAIRE**

In answering these questions please do not refer back to the case materials.

Circle the best answer for questions 1-6.

(1) The information presented suggested that inherent risk for the client was:

HIGH                      MODERATE                      LOW

(2) The information presented suggested that control risk, including elements relating to the control environment, accounting system, and control procedures was:

HIGH                      MODERATE                      LOW

(3) The comparative income statement directed your attention to a possible overstatement of sales revenue.

TRUE                      FALSE

(4) The results of other audit procedures performed alerted you to possible problems in:

(a) cash      (b) accounts receivable      (c) inventory      (d) none of the proceeding.

(Circle as many of the above as necessary.)

(5) Vouching sales is a      STRONG      MODERATE      WEAK      test for detecting overstatements.

(6) The procedure resulted in an audit difference that was  
LESS THAN                      GREATER THAN                      tolerable error for the account balance.

(7) What is your age? \_\_\_\_\_                      (8) What is your gender?      M      F

(9) How long have you worked in public accounting? \_\_\_\_\_ years and \_\_\_\_\_ months.

(10) How long have you worked with your present firm? \_\_\_\_\_ years and \_\_\_\_\_ months.

(11) Are you a CPA?      YES                      NO                      (circle one).

(12) What is your level within the firm? \_\_\_\_\_

THANK-YOU VERY MUCH FOR YOUR PARTICIPATION. THE RESULTS WILL BE MADE AVAILABLE IN A FEW WEEKS TO ANYONE WHO IS INTERESTED

**REFERENCES**

- AICPA. 1988. **The Auditor's Responsibility to Detect and Report Material Errors and Irregularities**. New York: AICPA.
- \_\_\_\_\_. 1993. **Codification of Auditing Statements on Auditing Standards**. New York: AICPA.
- Ashton, A. H. 1991. Experience and error frequency knowledge as potential determinants of audit expertise. *The Accounting Review* 66 (April): 218-39.
- \_\_\_\_\_, R. Ashton. 1988. Sequential belief revision in auditing. *The Accounting Review* 63 (October): 623-41.
- Ashton, R. 1992. Effects of justification and a mechanical aid on judgment performance. *Organizational Behavior and Human Decision processes* 52: 293-306.
- Baron, J. and J. C. Hershey. 1988. Outcome Bias in Decision Evaluation. *Journal of Personality and Social Psychology*. 569-579.
- Beach, L. R., Smith, B, Lundell, J. and T. R. Mitchell. 1988. Image theory: descriptive sufficiency of a simple rule for the compatibility test. *Journal of Behavioral Decision Making*. 17-28.
- Bedard, J. 1992. Unreliable evidence and auditor's judgments. Working Paper, University of Laval, Quebec.
- Brown, R. 1990. Assessment uncertainty technology for making and defending risky decisions. *Journal of Behavioral Decision Making*. 3: 213-228.
- Burgstahler, D., and J. Jiambalvo. 1986. Sample error characteristics and projection of error to audit populations. *The Accounting Review* (April): 233-48.
- Church, B. K. 1991. An examination of the effect that commitment to an hypothesis has on auditors' evaluations of confirming and disconfirming evidence. *Contemporary Accounting Research* 7 (Spring): 513-24.
- Curley, S., Yates, J. and R. Abrams. 1986. Psychological sources of ambiguity avoidance. *Organizational behavior and human decision processes* (October): 230-256.
- Einhorn, H. 1976. A synthesis: accounting and behavioral science. *Journal of Accounting Research* 14. (Supplement): 196-206.

- \_\_\_\_, and R. Hogarth. 1985a. Ambiguity and uncertainty in probabilistic inference. *Psychological Review* (October): 433- 461.
- \_\_\_\_, and \_\_\_\_\_. 1985b. A contrast/surprise model for updating beliefs. Working Paper, University of Chicago, Chicago.
- \_\_\_\_, and \_\_\_\_\_. 1986. Judging probable cause. *Psychological Bulletin* (No. 1): 3-19.
- Ellsberg, D. 1961. Risk, ambiguity and the Savage Axioms. *Quarterly Journal of Economics* (November): 643-669.
- Frisch, D. and J. Baron. 1988. Ambiguity and Rationality. *Journal of Behavioral Decision making*. 149-157
- Grant, Thornton. 1992. **Accounting and Auditing Manual**.
- Graessner, A. and C. McMahan. 1993. Anomalous information triggers questions when adults solve quantitative problems and comprehend stories. *Journal of Educational Psychology* 85 (No.1): 136-151.
- Koonce, L. 1992 Explanation and counter explanation in audit analytical review. *The Accounting Review* 67 (January): 59-76.
- Libby, R. 1985. Availability and the generation of hypotheses in analytical review. *Journal of Accounting Research* 23(Autumn): 648-67.
- McMillan, J and R. White. 1993. Auditor's belief revisions and evidence search. *The Accounting Review* 68 (July): 443-466
- Messier, W. and Quilliam. 1992. The effect of accountability on judgement: development of hypothesis for auditing. *Auditing a Journal of Practice and Theory*: 123-149.
- Morton, J. and W. L. Felix, Jr. 1990. Assessing control risk: effects of procedural differences on auditor consensus. In **Proceedings of the 1990 University of Kansas Auditing Symposium**. Lawrence: University of Kansas.
- Neter, J., Wasserman, W., and Michael Kutner. 1985. **Applied Linear Statistical Models**. Homewood, Illinois: Richard D. Irwin.
- Nisbett, R. and L. Ross. 1980. **Human Inference: Strategies and Shortcomings of Social Judgment**. Englewood, NJ: Prentice-Hall.

- Piaget, J. 1952. **The Origins of Intelligence in Children**. New York: Internationales Universities Press.
- Pyszczynski and Greenberg. 1981. Role of disconfirmed expectancies in the instigation of attributional processing. *Journal of Personality and Social Psychology* 40 (No.1): 31-38.
- Sanbonmatsu, D., F.R. Kardes and P. M. Herr. 1992 The Role of Prior Knowledge and Missing Information in Multiattribute Evaluation. *Organizational Behavior and Human Decision Process*, 51,76-91
- Savage, L. 1972. **Foundations of Statistics**. New York: Dover Press.
- SEC. 1977. **Accounting Series Releases and Staff Accounting Bulletins as of September 1, 1977**. Chicago: Commerce Clearing House.
- Smith, G., P. G. Benson and S. Curley. 1991. Belief, knowledge and uncertainty: A cognitive perspective on subjective probability. *Organizational Behavior and Human Decision processes* 48: 291-321.
- Waller, W. 1990. Auditors' risk assessments: effects of experience and conditioning information on second-order uncertainty. Working Paper, University of Arizona, Tucson, Az.
- \_\_\_\_\_. 1993. Auditors' inherent/control risk assessments in field settings. Working Paper, University of Arizona, Tucson, Az.
- \_\_\_\_\_, and W. Felix, Jr. 1984. Cognition and the auditor's opinion formulation process: a schematic model of interactions between memory and current audit evidence. in S. Moriarity and E. Joyce (eds.), **Decision Making and Accounting: Current Research**, University of Oklahoma Press: Norman, Oklahoma.
- \_\_\_\_\_, and \_\_\_\_\_. 1984. The effects of incomplete outcome feedback on auditors' self-perception of judgment ability. *The Accounting Review* (October): 637-646.
- \_\_\_\_\_, and \_\_\_\_\_. 1984. The auditor and learning form experience: some conjectures. *Accounting Organizations and Society* (No. 3/4): 383-406.
- \_\_\_\_\_, and \_\_\_\_\_. 1987. Auditors' covariation judgments. *The Accounting Review* (April): 275-92.

- \_\_\_\_, and \_\_\_\_\_. 1989. Auditors' causal judgments: effects of forward vs. Backward inference on information processing. *Accounting Organizations and Society* (No. 1): 179-200.
- \_\_\_\_\_. and T. Mitchell. 1991. Conditional probability judgments: effects of imagining vs experiencing the conditioning event. *Organizational Behavior and Human Decision processes* 49: 302-325.
- Yates, J.F. and L. G. Zukowski. 1976, Characterization of Ambiguity in Decision Making. *Behavioral Science* vol. 21 12-25.