

Cue Combination in Auditor's Internal Control Judgments

Jonas Gaudernack

DISSERTATION SUBMITTED IN PARTIAL FULFILLMENT OF THE
REQUIREMENTS FOR THE PHILOSOPHIAE DOCTOR (PhD) DEGREE AT THE
NORWEGIAN SCHOOL OF ECONOMICS AND BUSINESS ADMINISTRATION,
DEPARTMENT OF ACCOUNTING, AUDITING AND LAW

Abstract

A key issue in audit judgment research has been how auditors combine information cues in order to make judgments. Prior research has shown that appropriate forms of cue combination can be identified by analyzing task characteristics, and that many auditors combine cues in a manner appropriate to the interrelationship of cues in the judgment task (Brown and Solomon 1990, 1991; Bonner 2007, 155). Furthermore, appropriate cue combination has been found to be important for judgment quality (Brown and Solomon 1990, 1991; Hooper and Trotman 1996). However, evidence on cue combination in auditor's internal control judgments is mixed and from the pre-1990's. Prior findings may therefore be outdated and incomplete compared to the task characteristics auditors' face, and the behavior they exhibit, when making internal control judgments in the current audit environment.

This dissertation uses task analysis to identify appropriate forms of cue combination in internal control judgments as a function of two task characteristics; the interrelationship of controls and the judgment response scale. A policy capturing experiment is then conducted in order to test whether auditors combine cues in the appropriate and predicted manner. Findings indicate that auditors generally make internal control judgments by combining cues in a predictable and appropriate manner, given the task characteristics of cue interrelationships and the judgment response scale.

The main contributions of this dissertation are to (1) define and clarify important concepts in internal control judgment research, and (2) to extend normative theory and descriptive evidence on how auditors should, and do, respond to changes in judgment task characteristics by changing the functional form of their judgment policy (i.e., how they combine cues). This knowledge is important for audit practice, since task analysis can help auditors make better judgments and thereby perform better audits.

Acknowledgements

In finalizing this dissertation I realize that many people have provided help and support along the way. First and foremost I am grateful to my advisor at NHH, Aasmund Eilifsen, who was patient and encouraging in the three years I spent on deciding to do a PhD in auditing. After the decision was made he has provided excellent advice, both regarding the content of the dissertation and on who should be on the doctoral committee.

In addition to Aasmund, Bill Messier completes the committee. Bill has been of invaluable help. First of all he opened the doors to the U.S. by facilitating a study year at Georgia State University's doctoral program in accounting. Second, he has in depth knowledge of literature, methodology and practice relevant to this dissertation. He has therefore been the ideal guide in navigating the field. Finally, on the personal level Bill has been a good motivator. The provision of encouragement, demands and perhaps a good glass of Sassacaia has been well balanced and timed.

As the dissertation has developed over the last year I have received helpful comments from the following people: Ken Trotman, Gorm Grønnevet, Lisa Lambert, Larry Brown and participants in the Critical Thinking in Accounting Seminar at Georgia State University, Kathryn Kadous and the PhD auditing class at Emory University, Robert Knechel (discussant) and participants at the EARNet 2007 conference in Aarhus, Denmark.

I am also thankful for financial aid from the Norwegian Institute of Public Accountants and PricewaterhouseCoopers. Without this aid I would not have been able to go through the program, and it would certainly have been impossible to spend a year in the U.S.

I final thanks goes to my wife and three small daughters. Completing a PhD while living in a foreign country with a two-year old and, in addition, one year old twins was an eccentric experience. However, it was fun and things got done.

Table of contents

ABSTRACT	3
ACKNOWLEDGEMENTS	4
TABLE OF CONTENTS	5
1. INTRODUCTION.....	13
1.1 BACKGROUND.....	13
1.2 PRIOR RESEARCH FINDINGS.....	15
1.3 DEVELOPMENT OF AUDIT REGULATION AND PRACTICE.....	16
1.4 RESEARCH OBJECTIVE AND CONTRIBUTION.....	18
1.4.1 <i>Judgment Setting and Conceptual Model</i>	18
1.4.2 <i>Research Questions</i>	20
1.4.3 <i>Contribution</i>	22
1.5 OUTLINE OF THE DISSERTATION.....	24
2. INTERNAL CONTROL: DEVELOPMENT OF REGULATION AND PRACTICE	25
2.1 WHAT IS INTERNAL CONTROL?.....	26
2.1.1 <i>Internal Control Defined</i>	26
2.1.2 <i>Internal Control Defined in Audit Regulation</i>	28
2.2 THE ROLE OF INTERNAL CONTROL IN AUDITING.....	32
2.2.1 <i>Overview of the Audit Process</i>	32
2.2.2 <i>Internal Control in Auditing</i>	34
2.3 CHANGES IN AUDIT REGULATION AND PRACTICE.....	36
3. LITERATURE REVIEW AND THEORY DEVELOPMENT	50
3.1 INTRODUCTION	50
3.2 POLICY CAPTURING: APPROACH AND METHODOLOGY	56

3.3	FINDINGS IN PSYCHOLOGY RESEARCH	60
3.3.1	<i>General Findings Indicate Linear Judgment Policies</i>	60
3.3.2	<i>Extent of Configurality may be Underestimated</i>	61
3.3.3	<i>Development of Judgment Models</i>	63
3.3.4	<i>Summary of Psychology Research</i>	63
3.4	FINDINGS IN AUDIT RESEARCH	64
3.4.1	<i>Initial Descriptive Findings; General Use of Linear Judgment Models</i>	64
3.4.2	<i>Development of Normative and Causal Theory</i>	65
3.4.3	<i>Subsequent Audit Research with Relevance for Cue Processing</i>	67
3.4.4	<i>Summary of Prior Audit Research Findings</i>	69
3.5	FURTHER DISCUSSION AND THEORY DEVELOPMENT	70
3.5.1	<i>A Closer Look at the Findings of Configurality in Internal Control Judgments</i>	70
3.5.2	<i>Control Interrelationships vs Cue Interrelationships</i>	72
3.5.3	<i>Incomplete Range of Control- and Cue Interrelationships</i>	75
3.5.4	<i>Cue Interrelationships May Not Be Well Understood by Participants</i>	82
3.5.5	<i>The Importance of the Judgment Response Scale</i>	84
3.5.6	<i>Judgment Policies and Judgment Models</i>	93
3.5.7	<i>Changes in Audit Practice and Regulation</i>	99
3.6	SUMMARY: CONTRIBUTION FROM LITERATURE REVIEW	100
4.	CONCEPTUAL MODEL AND RESEARCH HYPOTHESES	101
4.1	CONCEPTUAL MODEL AND THEORY	101
4.2	HYPOTHESES	103
4.2.1	<i>H1: Multi-Step (i.e., Completely-Dependent) Controls</i>	103
4.2.2	<i>H2: Substitutable Controls</i>	105

4.2.3	<i>H3: Independent Controls</i>	108
4.2.4	<i>H4: Compensating Controls</i>	111
4.2.5	<i>H5: Amplifying Controls</i>	114
4.2.6	<i>Comparison of Hypotheses with Prior Research</i>	117
4.2.7	<i>Predictions for Three-Cue Judgment Tasks</i>	118
5.	METHODOLOGY AND RESEARCH DESIGN	121
5.1	PARTICIPANTS	121
5.2	EXPERIMENTAL DESIGN AND PROCEDURES	122
5.2.1	<i>Research Design</i>	122
5.2.2	<i>Independent and Dependent Variables</i>	122
5.2.3	<i>Materials and Procedures</i>	126
5.2.4	<i>Pilot Testing</i>	132
6.	HYPOTHESES TESTING	133
6.1	INTRODUCTORY DISCUSSION	133
6.1.1	<i>Statistical Analysis</i>	133
6.1.2	<i>Control Interrelationship Treatment</i>	136
6.2	POST EXPERIMENTAL QUESTIONS	138
6.3	TEST OF HYPOTHESES AND DISCUSSION.....	141
6.3.1	<i>H1: Multi-Step (i.e., Completely Dependent) Controls</i>	142
6.3.2	<i>H2: Substitutable Controls</i>	144
6.3.3	<i>H3: Independent Controls</i>	147
6.3.4	<i>H4: Compensating Controls</i>	150
6.3.5	<i>H5: Amplifying Controls</i>	155
6.4	SUMMARY OF FINDINGS	160

7. DISCUSSION AND IMPLICATIONS	161
7.1 ANSWERS TO RESEARCH QUESTIONS	161
7.2 CONTRIBUTION TO RESEARCH.....	174
7.3 CONTRIBUTION TO AUDIT PRACTICE.....	176
7.4 LIMITATIONS	177
7.5 SUGGESTIONS FOR FUTURE RESEARCH	178
REFERENCES.....	180
APPENDIX 1 – COVER LETTER AND INTRODUCTION.....	186
APPENDIX 2 – EXAMPLES OF CASES	191
APPENDIX 3 – POST EXPERIMENTAL SURVEY	196

List of Tables

Table 1: Mathematical Representation of Cue Interrelationship Functions

Table 2: Mathematical Representation of Judgment Models

Table 3: Variation in Control Interrelationships in Three-Cue Judgment Tasks

Table 4: Overview of Hypotheses and Predicted Coefficients in Three-Cue Judgment Tasks

Table 5: Example of Compensating Control Case

Table 6: Order of Envelopes in Experiment

Table 7: Order of Cases within an Envelope

Table 8: Order of Cases

Table 9: Crossing of Envelope Order with Case Order

Table 10: Summary of Post Experimental Question Responses

Table 11: Overview of Hypotheses and Predicted Coefficients in Three-Cue Judgment Tasks

Table 12: Results from Linear Regression H1a

Table 13: Results from Linear Regression H2a

Table 14: Results from Linear Regression H3a

Table 15: Results from Linear Regression H4a

Table 16: Results from linear regression H5a

Table 17: Summary of Results from Test of Hypotheses

Table 18: Mathematical Representation of Control Interrelationship Functions

List of Figures

Figure 1: Lens Model for Control Risk Judgment with Three Control Cues

Figure 2: Conceptual Model

Figure 3: Lens Model for Control Risk Judgment

Figure 4: Cue Interrelationships (Lens Model Environment)

Figure 5: Cue Interrelationships (Lens Model Environment)

Figure 6: Cue Interrelationship Continuum

Figure 7: Conceptual Model

Figure 8: Judgment Response Scale: Number of Response Options

Figure 9: Judgment Response Scale and Criterion Scale

Figure 10: Belief Formation and Judgment Rules

Figure 11: Model of Determinants of Cue Interrelationships

Figure 12: Independent Cue Interrelationships

Figure 13: Completely-Dependent Cue Interrelationships

Figure 14: Conceptual Model in Prior Studies

Figure 15: Conceptual Model in This Dissertation

Figure 16: Conceptual Model H1a

Figure 17: Conceptual Model H1b

Figure 18: Conceptual Model H2a

Figure 19: Conceptual Model H2b

Figure 20: Conceptual Model H3a

Figure 21: Conceptual Model H3b

Figure 22: Conceptual Model H4a

Figure 23: Conceptual Model H4b

Figure 24: Conceptual Model H5a

Figure 25: Conceptual Model H5b

Figure 26: Cue Interrelationships (Lens Model Environment)

Figure 27: Model of Determinants of Cue Interrelationships

Figure 28: Control Interrelationships (Lens Model Environment)

Figure 29: Cue Interrelationship Continuum

Figure 30: Cue Interrelationships (Lens Model Environment)

Figure 31: Judgment Response Scale: Number of Response Options

Figure 32: Lens Model for Control Risk Judgment (Lens Model Judgment Side)

Figure 33: Conceptual Model in This Dissertation

1. Introduction

1.1 Background

“Auditing is a systematic process of objectively obtaining and evaluating evidence regarding assertions about economic actions and events to ascertain the degree of correspondence between those assertions and established criteria and communicating the results to interested users” (American Accounting Association, 1973)

In performing auditing, auditors need to master the analytical and logical skills necessary to evaluate (i.e., judge) both the systems and processes that produce information as well as the information itself (Eilifsen et al., 2006, 6). For example, in a financial statement audit the auditor makes judgments about the accounting systems and processes as well as about the financial statement. At the general level, auditors can therefore be viewed as experts in gathering and assessing evidence in order to evaluate assertions vis-à-vis criteria, and reporting the findings to interested parties (Solomon and Shields, 1995, 138).

Worded in judgment and decision making research terms, auditing consists of selecting cues for evaluation and weighting and combining them in order to make a judgment. The quality of judgments therefore depends on (1) the cues selected for evaluation, and (2) how those cues are weighted and combined. The aim of this dissertation is to contribute to improving judgments by providing knowledge about how cues should be combined (i.e., normative focus) and compare this to how judges actually combine cues (i.e., descriptive focus).

The focus of this dissertation is limited to auditor judgments, and not subsequent decisions that may be based on these judgments. The term “judgment” thus refers to subjective assessments made as a prelude to taking action while the term “decision” means actions that people take to perform some task or solve some problem (definitions from Solomon and Trotman 2003).

Cue combination has been an area of focus for judgment analysis research in the basic field of psychology and in many applied fields, including auditing (Libby 1981; Cooksey 1996). In audit judgment research, a focus has been on whether auditor’s judgment policies involve

simple weighted averaging of cues (i.e., linear cue processing) versus more complex forms of cue processing (i.e., configural cue processing) (Brown and Solomon 1990, 1991; Trotman 1998, 2005).¹ Less focus has been put on developing normative theory for how cues should be combined in audit judgments (with the exception of Brown and Solomon 1990 and 1991).

Audit research on cue combination has primarily used internal control judgment tasks (Libby and Lewis 1982; Brown and Solomon 1990; Solomon and Shields 1995; Trotman 1998, 2005). Internal control judgment tasks are important because inappropriate cue combination may cause judgment errors with serious consequences for the audit. Although few studies directly examine consequences of judgment errors, serious consequences cannot be ruled out (Brown and Solomon 1990, 1991; Bonner 2007, 155). Examples include: (1) judgment differences may impact audit planning and thereby audit efficiency and effectiveness (Brown and Solomon 1990 and 1991; Hooper and Trotman 1996; ISA 315, IFAC 2008; AS5, PCAOB 2007), (2) due to erroneous audit judgments about presence/absence of internal control deficiencies, audit clients might initiate unnecessary remediation efforts (i.e., reorganization), or not initiate necessary remediation efforts, and (3) judgment differences may lead to different audit reports under the Sarbanes-Oxley regime, and these audit reports may impact cost of debt and equity (e.g., Doss and Jonas 2004; Ashbaugh-Skaife et al. 2008; Ogneva et al. 2007).

Studies of cue combination have typically used a policy capturing approach (i.e., Lens Model), where ANOVA models are constructed of participants' judgments, and judgment variance attributable to interaction terms is used to examine the extent and form of configularity in judgment policies (Libby and Lewis 1982; Brown and Solomon 1990, 1991; Solomon and Shields 1995; Trotman 1996, 1998, 2005). This dissertation uses a similar approach.

¹ Linear cue processing means that the effect of a cue on the judgment does not depend on the level of other cues (i.e., only main effects of cues). Configural cue processing means that the effect of a cue on the judgment may depend on the level of other cues (i.e., cue pattern or interaction effects) (Brown and Solomon 1990). Different kinds of configural processing are discussed in detail later in this dissertation.

1.2 Prior Research Findings

The study of cue combination is performed through policy-capturing, which is a method to assess how judges use available information when making judgments (i.e., how they weight, combine or integrate information) (Karren et al. 2002). The policy capturing approach evolved from Egon Brunswik's probabilistic "lens" model, developed in the 1950's. The general finding from psychology research using this approach is that configularity is not beyond human judges, but that it is not very typical of human judgment (Brehmer 1994; Cooksey 1996). Furthermore it was concluded that in order to answer questions about whether there will be configural components in judgment models, consideration of the characteristics of the specific task is needed (Brehmer 1994; Cooksey 1996; Stewart et al. 1997).

Audit judgment research "imported" the Lens Model approach from psychology research through Ashton's (1974) study of auditors' internal control judgments. Many similar studies followed in the 1970's and 1980's. These studies generally used internal control judgment tasks and focused on judgment consensus, although descriptive evidence on e.g., cue weighting, cue combination, judgment insight and judgment consistency over time was also provided (Solomon and Trotman 2003; Trotman 1998, Trotman 2005). However, none of these studies included apriori predictions or normative theory development about the effect of task characteristics on cue combination (Brown and Solomon 1990). Results showed that very little judgment variance was explained even by aggregating all possible interaction terms in auditor judgment models (Trotman 1996, 105). Furthermore there were large differences between individuals as to which interaction terms were significant (Trotman 1996, 105). Overall, no consistent evidence of configural judgment policies was found (Libby and Lewis 1982; Solomon and Shields 1995; Trotman 1998, 2005). The general understanding was therefore that auditors did not process cues configurally (Brown and Solomon 1990; Trotman 2005). No theory was developed as to whether the lack of configularity was appropriate or not.

Brown and Solomon (1990) argued that the lack of configularity in prior studies could be due to experimental task characteristics (i.e., independent cues) that made linear judgment policies appropriate. They therefore designed an internal control judgment task where cue interrelationships were such that configural judgment models were appropriate and expected.

This was achieved by manipulating cue interrelationships according to normative theory about how cue interrelationships should affect cue combination. Findings showed that a substantial portion (40.5%) of the auditors made judgments that could be described by configural models. Furthermore, reported results state that for some of the auditors at least one of the identified interactions was of a form and nature consistent with those predicted given the manipulation of cue interrelationships. Although no subsequent studies of auditor's cue processing in internal control judgments have been identified by the author, similar results as Brown and Solomon's (1990) have been reported when using misstatement risk judgment tasks (i.e., the risk of material misstatement after the auditor has performed substantive testing) (Brown and Solomon 1991; Hooper and Trotman 1996; Leung and Trotman 2005).

Thus, the general finding from policy capturing studies of auditor's judgments is that judgment policies can be mathematically represented by linear models unless experimental tasks are purposefully constructed so that configural judgment policies are appropriate (Brown and Solomon 1990, 1991; Trotman 1998, 2005; Bonner 2007, 155). Furthermore, when tasks are constructed so that configural cue processing is appropriate, many, but not all, auditors are able to process cues configurally (Brown and Solomon 1990, 1991; Trotman 1998, 2005; Bonner 2007, 155). This finding is consistent with results from psychology research using other kinds of expert judges and judgment tasks (Einhorn 1979; Brown and Solomon 1990, 1991; Brehmer 1994; Hooper and Trotman 1996; Stewart et al. 1997; Elrod et al. 2004; Bonner 2007, 155).

1.3 Development of Audit Regulation and Practice

Since Brown and Solomon's study in 1990, regulatory changes in the internal control and audit landscape have been significant. Examples of those changes include:² (1) COSO (1992), which provided an internationally accepted framework for designing and evaluating internal controls, focusing on defining internal controls as a process; (2) the Sarbanes-Oxley

² COSO is the Committee of Sponsoring Organizations of the Treadway Commission. IAASB refers to the International Auditing and Assurance Standards Board. ISA refers to International Standards on Auditing, which are issued by the IAASB. AICPA refers to the American Institute of Certified Public Accountants.

Act (2002), which introduced a separate audit opinion on internal controls over financial reporting for public companies and emphasized a top-down audit approach starting with entity-level controls; (3) ISA 315 (IFAC 2008, issued late 2003, effective as of 2004) and similar standards imposed in the U.S. by the AICPA, which introduced more detailed benchmarks for what the auditor should consider when judging internal controls (e.g., see appendix two in ISA 315, IFAC 2008); and (4) European Union (EU) 8th directive (2006) that requires the auditor of public-interest entities to report to the audit committee on material findings, including material weaknesses in internal controls over financial reporting (effective as of June 2008). It is possible that these regulatory changes have changed the characteristics of the internal control judgment task and that they may contribute to more consistent approaches to internal control judgments. The external validity of research on internal control judgments conducted prior to these regulatory changes may therefore be reduced.

Other important changes, with a potential affect auditor's internal control judgments, have been initiated in audit practice. The development of the business risk audit (or strategic systems audit) in the 1990's increased auditor focus on risk and control (Knechel 2007; Peecher et al. 2007). Furthermore, mergers resulted in four dominating audit firms (i.e., the Big-4)³ where anecdotal evidence suggests generally increased use of firm-wide, computerized audit training, guidance, documentation and review in order to achieve consistent, high quality judgments.⁴ Finally, increased regulatory pressure may have increased quality control procedures within accounting firms, thus contributing to more appropriate and consistent control judgments.

Changes in audit regulation and audit practice may therefore have changed both the judgment task (i.e., the task characteristics and the accompanying appropriate judgment policies) and the behavior of the judge (e.g., increased use of appropriate judgment policies compared to findings in prior research). The relevance of prior research findings for

³ The term "The Big-4" refers to the four large international accounting firms: Ernst&Young, PricewaterhouseCoopers, Deloitte and KPMG.

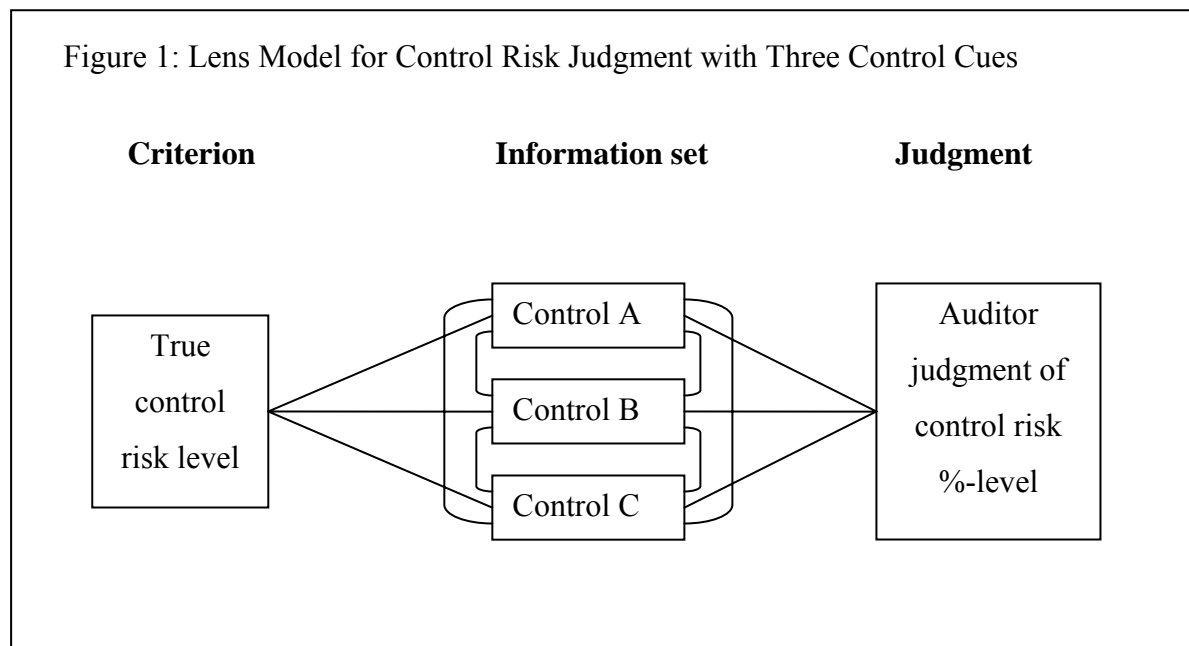
⁴ Anecdotal evidence was given by the technical partner of the firm providing audit participants for this study, and is consistent with audit practice descriptions in prior research (e.g., Brazel et al. 2004; Knechel 2007).

describing and/or improving auditor judgment behavior in today's audit environment may therefore be questionable.

1.4 Research Objective and Contribution

1.4.1 Judgment Setting and Conceptual Model

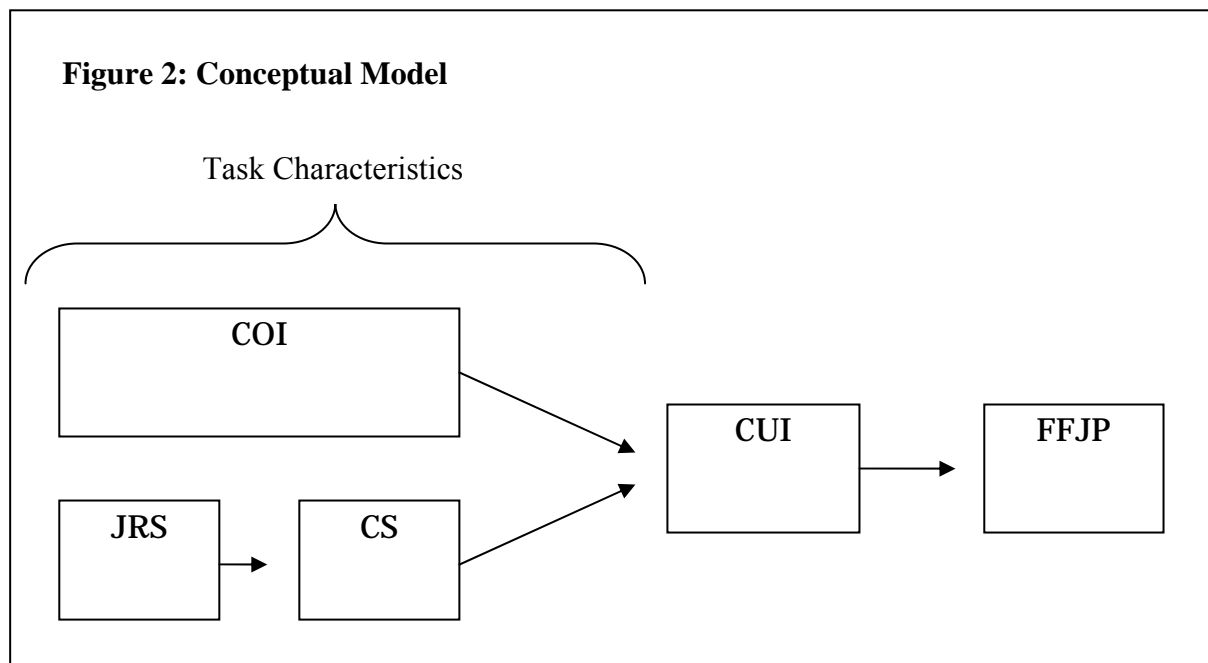
This dissertation assumes a judgment setting where the judge (i.e., the auditor) is provided with an information set consisting of cues (i.e., internal control cues) and asked to provide a judgment about the true state of a criterion (e.g., control risk) on a given response scale (e.g., control risk on a percentage scale). This judgment setting can be modeled in a Lens Model framework (see figure 1 below), and a policy capturing approach can be used to study the functional form of the auditor's judgment policy.



The objective the dissertation is to develop normative theory and provide descriptive evidence on how task characteristics affect the functional form of the auditor's judgment policy in internal control judgments in today's audit environment (see conceptual model in figure 2 below). The developed theory proposes that:

1. The criterion scale (“CS”) is a function of the judgment response scale (“JRS”)
2. Cue interrelationships (“CUI”) are a function of control interrelationships (“COI”) and the criterion scale (“CS”)
3. The functional form of the judgment policy (“FFJP”) is a function of cue interrelationships (“CUI”)

This means that the auditor combines cues in a similar manner as to how cues relate to the criterion in the environment, and that audit task characteristics (i.e., COI and JRS) determine cue interrelationships.



The three propositions above can be expressed mathematically in terms of functions and inner functions:⁵

1. $CS = h(JRS)$
2. $CUI = g[COI, CS] = g[COI, h(JRS)]$
3. $FFJP = f\{CUI\} = f\{g[COI, CS]\} = f\{g[COI, h(JRS)]\}$ ⁶

⁵ Variables in the functions can be categorical, discrete or continuous. Furthermore, no functional form is defined since the functions express generic relationships.

Where:

JRS = Judgment Response Scale

CS = Criterion Scale

COI = Control Interrelationships

CUI = Cue Interrelationships

FFJP = Functional Form of Judgment Policy

The conceptual model will be discussed further in the theory development section of the dissertation. Normative evidence is provided in the form of theory development about how task characteristics should affect the functional form of judgment policies. Descriptive evidence is provided in the form of experimental evidence. In order to achieve the objective of providing normative and descriptive evidence, it is necessary to first define the nature and range of variation of the constructs in the conceptual model.

1.4.2 Research Questions

Defining the nature and range of constructs

Prior audit research has not been clear on the difference between control- and cue interrelationships. Furthermore prior research has used an incomplete range of control- or cue interrelationships compared to the control interrelationships that may be relevant in today's audit environment. Finally, no attempt to define a full range of variation in control- or cue interrelationships has been identified. It therefore seems that a definition of potentially important control- and cue interrelationships may be a relevant contribution to the literature. The following research questions are therefore formulated (see below). RQ1 – RQ5 regard construct development. RQ6 regards normative relationships between constructs and RQ7 regards empirical relationships between constructs.

⁶ The function “ $FFJP = f\{CUI\} = f\{g[COI, CS]\} = f\{g[COI, h(JRS)]\}$ ” can be read as follows: FFJP is a function of CUI, which is a function of COI and CS, where CS is a function of JRS.

RQ1: What is the difference between control interrelationships and cue interrelationships?

RQ2: What is the nature and range of variation in control interrelationships?

RQ3: What is the nature and range of variation in cue interrelationships?

No studies have been identified where another potentially important task characteristic, the judgment response scale, has been taken explicitly into consideration. This may be an important task characteristic since auditors make judgments on many kinds of response scales (e.g., binary scales such as yes/no and accept/reject, or continuous scales such as percentage scale risk judgments). A definition of the nature and range of the judgment response scale may therefore be a potentially important contribution to the literature:

RQ4a: What is the nature and range of variation in the judgment response scale?

The judgment response is the judge's estimate of the criterion (see figure 2 above). The criterion can therefore also be represented on a scale, and this scale may also be an important task characteristic.

RQ4b: What is the nature and range of variation in the criterion scale?

In the literature, the functional form of the studied judgment policies hitherto has been limited to compensatory models, while important judgment tasks in today's audit environment may call for other judgment policies. A potentially important contribution to the literature is therefore to identify other judgment policies and models that may be relevant in internal control judgments:

RQ5: What forms of judgment policies and models are relevant in auditors internal control judgments?

Normative relationships between constructs

If research is to contribute to improving judgments, it is useful to have a normative benchmark for how judgments should be made. This dissertation therefore poses the following research question about how task characteristics should affect judgment policies:

RQ6: How do control interrelationships and the judgment response scale affect cue interrelationships, and how should judgment policies be affected?

Descriptive/empirical relationships between key constructs

Descriptive/empirical evidence on how control interrelationships affect judgment policies in auditors internal control judgments is limited to one study in 1990 (i.e., Brown and Solomon 1990). The effect of the judgment response scale has not been studied. Furthermore, the relevance of findings in Brown and Solomon (1990) for internal control judgments in today's audit environment may be limited. An extension and update is therefore warranted, and the following research question is posed:⁷

RQ7: How do control interrelationships and the judgment response scale affect judgment policies?

1.4.3 Contribution

This dissertation contributes to audit judgment research through construct development, normative theory development and by providing empirical evidence on cue combination in control judgments. The construct development and normative theory development of the dissertation should also be relevant for judgment research in other fields (i.e., other judges and/or other tasks) where similar task characteristics are relevant.

Audit research is extended by: (RQ1) clarifying the difference between control interrelationships and cue interrelationships, (RQ2) developing a framework defining the range of variation in control interrelationships, (RQ3) developing a framework defining the range of variation in cue interrelationships, (RQ4a) introducing and defining the judgment response scale and (RQ4b) the criterion scale as task characteristics in internal control judgments, including defining the range of variation, (RQ5) defining relevant functional forms of judgment policies in internal control judgments, (RQ6) developing normative theory about how control interrelationships and the judgment response scale should affect cue interrelationships and the functional form of judgment policies, and (RQ7) providing

⁷ This dissertation argues that cue interrelationships are logically determined by other task characteristics (i.e., control interrelationships and the criterion scale which is a function of the judgment response scale). The descriptive/empirical research question does therefore not include statements about cue interrelationships.

empirical evidence on how control interrelationships and the judgment response scale affect the functional form of judgment policies.⁸ Section 7.2 summarizes the contribution of this dissertation by comparing it to the state of knowledge after Brown and Solomon (1990), which is the first and only study of auditor's control judgments finding evidence of configularity.

The theoretical contribution of the dissertation should also benefit audit practice. First of all it is unclear whether today's auditors apply appropriate judgment policies in internal control judgments. Prior evidence revealed a relatively low extent of configural judgment policies even when the internal control task required it (Brown and Solomon 1990). Therefore, it should be of interest for practitioners to receive updated evidence on whether this is still a problem.

Second, if inappropriate judgment policies are applied, judgment quality, and thus audit quality may suffer (Brown and Solomon 1990, 1991; Hooper and Trotman 1996; Leung and Trotman 2008). Audit practice should therefore benefit from the development of normative benchmarks for evaluation of actual judgment policies. Such normative benchmarks may help in identifying differences between appropriate and actual judgment policies and thus shed light on where improvement is needed. Deviations from normative benchmarks may also provide an explanation for poor performance and disagreements between judges (Libby 1981, 31-32).

Third, knowledge about the relationship between task characteristics and the appropriate form of the judgment policy can help in training decision makers and in developing decision aids.

Finally, the construct development in this dissertation may provide useful frameworks and definitions for analyzing task characteristics both in real life audit settings and in the classroom.

⁸ The focus of this study is auditor judgments, and not subsequent decisions that may be based on these judgments. The term "judgment" thus refers to subjective assessments made as a prelude to taking action while the term "decision" means actions that people take to perform some task or solve some problem (definitions from Solomon and Trotman 2003)

1.5 Outline of the Dissertation

The rest of this dissertation is organized as follows: Section two reviews the development of relevant audit regulation and practice. Section three reviews and discusses relevant theory and findings in psychology and audit research. The review of psychology research focuses on methodology and general findings from the policy capturing paradigm regarding configularity. The review of audit research primarily concentrates on policy capturing studies of internal control judgments, but also discusses relevant findings from research on other judgment tasks where configural cue usage has been found. Section three furthermore develops the theoretical constructs of interest for this study and proposes normative theory about how control interrelationships and the judgment response scale should affect the functional form of the judgment policy. Section four develops the hypotheses. Section five presents the experimental design and procedures. Section six presents and discusses the experimental results. Section seven summarizes the responses to the research questions posed in the introductory section of this dissertation. Section seven ends with a discussion of limitations of this study, implications for audit practice and suggestions for future research.

2. Internal Control: Development of Regulation and Practice

Over the 35 years that have elapsed since the first study of auditor's internal control judgments (Ashton 1974) changes in regulation and practice relevant to internal control and auditing have been significant. These changes may have had an impact on what kind of judgments the auditor performs, and on how, and how well, the judgments are performed. The following chapter therefore provides an overview of the development of regulation and practice relevant to internal control judgments in auditing.

The historical review focuses primarily on regulation and practice in the United States. International developments are commented on where relevant. Such an approach is reasonable for several reasons: First, because internal control judgment research has mainly been conducted in the U.S., U.S. regulation is the relevant context for this research. Second, because of the assumed historical lead role of the U.S. branches in the development of audit practice in the dominating audit firms, U.S. practice is of primary interest. Third, the international auditing standards have historically been less comprehensive, although not fundamentally different, from the U.S. standards.

It is however, important to recognize that international standards on auditing (ISAs) have been playing an increasingly important role over time, especially the last decade. Since the International Federation of Accountants (IFAC) was founded in 1977, beginning with 63 members, IFAC has grown, and as of 2007 includes 158 members and associates in 123 countries and jurisdictions worldwide (Humphrey and Loft 2007). By 2007, more than 100 countries had adopted ISAs or were using them as a basis for their national standards (ibid). Furthermore, the European Commission is considering how to endorse ISAs as the auditing standard for all 27 European Union member states (ibid).⁹ In the U.S., the Audit Standards

⁹ Norway is not member of the European Union (EU). The European Economic Area (EEA) unites, however, the 27 EU member states and the three European Free Trade Association (EFTA) states, Iceland, Lichtenstein, and Norway. This implies among other things that EU regulatory measures of auditing apply to all EEA states, including Norway.

EU 8th directive, article 26 (EU 2006) reads as follows "*Member States shall require statutory auditors and audit firms to carry out statutory audits in compliance with international auditing standards adopted by the Commission in accordance with the procedure referred to in Article 48 (2)*". The Commission's procedure for adopting ISAs is, however, dependent on many things such as the IAASB clarity project. It is therefore not clear what parts of the ISAs will be adopted. It is therefore fair to state that the EU is aiming to adopt ISAs, but they are considering how to do it.

Board is working towards increased harmonization with the IAASB. Finally, the largest accounting firms have committed to auditing in accordance with ISAs.¹⁰ Today, it is therefore fair to state that both ISAs and U.S. regulation are important. The former because they are widely used internationally, and the latter because they apply to the audit of all companies listed on U.S. stock exchanges, including foreign listed companies. Furthermore, the so-called “risk standards” discussed in this chapter are similar at the international level and in the U.S. The overview of audit regulation and practice in today’s audit environment is therefore balanced between the U.S. and internationally.

2.1 What is Internal Control?

2.1.1 Internal Control Defined

The definition, meaning and use of internal controls in business practice and auditing has developed and transformed continuously since early in the 20th century (Heier et al. 2005). In 1992, the highly influential COSO acknowledged that internal control meant different things to different people and that this caused confusion among business people, legislators, regulators and others. The resulting miscommunication and different expectations caused problems within enterprises, and problems were compounded when the term, if not clearly defined, was written into law, regulation or rule (COSO 1992). As a response, COSO issued *Internal Control – an Integrated Framework* (COSO 1992). The report was a milestone for the understanding of the meaning and content of internal control. The purpose of the report was to define internal control, describe its components and provide criteria and supporting materials for evaluating internal control systems. The report provided an internationally accepted framework for evaluating internal controls, focusing on defining internal controls as a process:

Internal control is broadly defined as a process, effected by an entity's board of directors, management and other personnel, designed to provide reasonable assurance regarding the achievement of objectives in the following categories:

¹⁰ The large accounting firms that participate in IFAC through the Forum of Firms, established in 2001.

1. *Effectiveness and efficiency of operations.*
2. *Reliability of financial reporting.*
3. *Compliance with applicable laws and regulations.*

Internal control consists of five interrelated components. These are derived from the way management runs a business, and are integrated with the management process:

- *Control Environment*
- *Risk Assessment*
- *Monitoring*
- *Control Activities*
- *Information and Communication*

The COSO dimensions and components of internal control became a benchmark that all businesses were expected to pursue as part of their operations (Knechel 2007). Today there is little doubt that the COSO definition of internal control has won widespread acceptance and that it is the foundation for internal control definitions in today's regulation. Examples include international and U.S. audit standards (ISA 315, IFAC 2008; AS5, PCAOB 2007) and corporate legislation such as the U.S. Sarbanes-Oxley Act (2002).

As an expansion and/or clarification of the framework, COSO has issued subsequent reports on enterprise risk management (2004), small entities (2006) and monitoring (2008):

Enterprise Risk Management - Integrated Framework (ERM) (COSO 2004): The framework defines essential enterprise risk management components, discusses key ERM principles and concepts, suggests a common ERM language, and provides clear direction and guidance for enterprise risk management.

Internal Control over Financial Reporting - Guidance for Smaller Public Companies (COSO 2006): The small business guidance takes the concepts of the 1992 Internal Control – Integrated Framework and demonstrates their applicability for achieving financial reporting objectives of smaller publicly traded companies.

Guidance on Monitoring Internal Control Systems (Exposure Draft, COSO 2008): Based on COSO's Internal Control - Integrated Framework (1992), this guidance is designed to help organizations monitor the quality of their internal control systems.

These newer frameworks/guidance are not yet referred to in audit regulation, such as the COSO (1992) integrated framework (e.g., see AS5.5, footnote 7, PCAOB 2007), but it is not unlikely that they will in the future. The main argument for this view is that they to a large degree clarify the content of the integrated framework. In doing so, they provide more detailed, but not new, benchmarks for internal control in companies already applying the integrated framework as their existing control framework. The new frameworks should therefore be of interest for board members, management, auditors, regulators and others.

For this dissertation, the frameworks are important for a number of reasons. First, because they provide evidence of internal controls including more than the transaction level control activities studied in prior audit research. That is, internal control includes many processes that are performed at higher organizational levels, such as risk management and monitoring, and such control processes may require other judgment policies than what has studied in prior research. Second, the frameworks provide acknowledged benchmarks for internal controls. The mere existence of such benchmarks should contribute to increased consensus in judgments and judgment policies about controls, and to increased use of configural judgment policies where appropriate.

2.1.2 Internal Control Defined in Audit Regulation

The glossary of terms in the international standards on auditing (IFAC Handbook, IFAC 2008) defines internal control in a similar manner as COSO (1992). The definition is therefore not repeated here. Under the clarity standards project,¹¹ the definition has been shortened and the importance of implementation and maintenance of internal control has been emphasized, but the substance is unchanged. The components of internal control that the definition refers to are the five components of the COSO Framework (COSO 1992):

¹¹ The clarity standards project is explained by the IAASB as follows: In seeking continually to improve its standards, the IAASB undertook in 2003 to review the drafting conventions used in its International Standards. The objective of the review was to identify ways to improve the clarity, and thereby the consistent application, of standards issued by the

“Internal control is the process designed, implemented and maintained by those charged with governance, management and other personnel to provide reasonable assurance about the achievement of an entity’s objectives with regard to reliability of financial reporting, effectiveness and efficiency of operations, and compliance with applicable laws and regulations. The term “controls” refers to any aspects of one or more of the components of internal control.” (ISA 315.4c redrafted, IFAC 2008)

It can be noted that the ISAs apply a broad definition of internal controls, which includes the effect of internal control on operations and compliance. However, although the definition is broad, the auditor is to focus on controls relevant to the audit (ISA 315.20 redrafted, IFAC 2008). This implies that some operational and compliance controls may be relevant to the audit, and that not all financial reporting controls are automatically relevant to the audit (ISA 315.12 redrafted, IFAC 2008).

The use of this broad control definition in auditing may be due to two considerations. First, (business) risks, and therefore controls, may have a simultaneous impact on reporting, operations and compliance, and it is therefore not feasible to define them solely as reporting controls. An example of such controls could be safeguarding of assets required by law (e.g., for medication) (ISA 315.A59 redrafted, IFAC 2008). Second, controls aimed at operations and compliance objectives may also have an impact on financial reporting and therefore be of interest for the audit (ISA 315.A58 redrafted, IFAC 2008). A broad definition of internal control, covering compliance and operations, is therefore appropriate in auditing.

Furthermore, the reporting requirements in international auditing standards seem to imply a broad approach by the auditor. ISA 315.32 redrafted (IFAC 2008) mandates communication about material weaknesses in internal control identified during the audit. Since the auditor looks at internal controls relevant to the audit, and since these may include controls over operations and compliance, the auditor must report material weaknesses in these if they come to the auditor’s attention. ISA 265 Communicating Deficiencies in Internal Control - Exposure Draft (IFAC 2008) goes even further in that it states *“Nothing in this ISA precludes the auditor from communicating control matters that the auditor has identified*

during the audit that are not relevant to the audit but that the auditor considers important.”. Together, this shows that the auditor must communicate all *material weaknesses* identified during the audit, regardless of the controls being operational, reporting controls or compliance controls. Furthermore, all control matters, even if they are not relevant to the audit, may be communicated by the auditor. With this view, it is reasonable that the IAASB has adopted a broader definition of internal controls than a definition just focusing on reporting controls.

U.S. audit regulation (i.e. AS5, PCAOB 2007) also builds on the definitions of the COSO framework. However, AS5 (PCAOB 2007) includes a specific definition of internal control over financial reporting:

“Internal control over financial reporting is a process designed by, or under the supervision of, the company's principal executive and principal financial officers, or persons performing similar functions, and effected by the company's board of directors, management, and other personnel, to provide reasonable assurance regarding the reliability of financial reporting and the preparation of financial statements for external purposes in accordance with GAAP and includes those policies and procedures that;

- 1. Pertain to the maintenance of records that, in reasonable detail, accurately and fairly reflect the transactions and dispositions of the assets of the company;*
- 2. Provide reasonable assurance that transactions are recorded as necessary to permit preparation of financial statements in accordance with generally accepted accounting principles, and that receipts and expenditures of the company are being made only in accordance with authorizations of management and directors of the company; and*
- 3. Provide reasonable assurance regarding prevention or timely detection of unauthorized acquisition, use, or disposition of the company's assets that could have a material effect on the financial statements.*

Although this definition has a transaction and book-keeping focus, the general picture is that both the U.S. PCAOB regime and the IAASB ISA regime may require the auditor to assess a wide range of operational and compliance controls since these may be relevant to the audit. This is evidenced by the emphasis on a top-down approach permeating AS5 (PCAOB 2007)

and ISA315 (IFAC 2008) – a top down approach implies focusing on risk management and entity level controls before continuing with transaction level controls (AS5.5, PCAOB 2007).

In addition to the definition of internal control in auditing, it is important to understand key terms related to control design versus control effectiveness, and to the degree of deficiencies in these. The definitions are presented below, and their application in auditing is discussed in the subsequent presentation of the audit process. In general, judgments about the existence of deficiencies in control can be decomposed into judgments about control design and judgments about operational effectiveness of controls (PCAOB AS5.A3 2007):

A deficiency in internal control over financial reporting exists when the design or operation of a control does not allow management or employees, in the normal course of performing their assigned functions, to prevent or detect misstatements on a timely basis.

- *A deficiency in design exists when (a) a control necessary to meet the control objective is missing or (b) an existing control is not properly designed so that, even if the control operates as designed, the control objective would not be met.*
- *A deficiency in operation exists when a properly designed control does not operate as designed, or when the person performing the control does not possess the necessary authority or competence to perform the control effectively.*

Judgments about the importance of deficiencies are performed by classifying deficiencies. Under the PCAOB regime, deficiencies are classified as material weakness, significant deficiency or deficiency. Under the ISA regime, deficiencies are classified as material weakness or not. The deficiency categories are defined as follows:

“A material weakness is a deficiency, or a combination of deficiencies, in internal control over financial reporting, such that there is a reasonable possibility that a material misstatement of the company's annual or interim financial statements will not be prevented or detected on a timely basis” (PCAOB AS5.A7, 2007).

“A material weakness is a weakness in internal control that could have a material effect on the financial statements” (ISA glossary of terms, IFAC 2008)

“A significant deficiency is a deficiency, or a combination of deficiencies, in internal control over financial reporting that is less severe than a material weakness, yet important enough to merit attention by those responsible for oversight of the company's financial reporting” (PCAOB AS5.A7, 2007).

It can be noted that the IAASB has an ongoing project regarding a separate ISA for classifying and communicating control deficiencies. The current status of the project is the issuance of an exposure draft named ISA 265 *Communicating Deficiencies in Internal Control*, expected to be approved in 2009. If approved as existing, the new ISA will have a significant impact on judgments about the classification of deficiencies. The term “material weakness” will not be used, and it will be removed from all other ISA’s. Instead, the term significant deficiencies will be implemented. This is defined similar to the definition from the PCAOB above. Apart from the changes in definitions, the issuance of the new standard is in itself evidence of an increased emphasis on internal controls and internal control judgments and reporting in auditing.

2.2 The Role of Internal Control in Auditing

2.2.1 Overview of the Audit Process

In order to provide a background for the role of internal controls in auditing, a brief overview of the audit process is presented.

“Auditing is a systematic process of objectively obtaining and evaluating evidence regarding assertions about economic actions and events to ascertain the degree of correspondence between those assertions and established criteria and communicating the results to interested users” (American Accounting Association, 1973)

“The objective of an audit of financial statements is to enable the auditor to express an opinion whether the financial statements are prepared, in all material respects, in accordance with an applicable financial reporting framework.” (ISA glossary of terms, IFAC 2008).

The audit process can be conceptualized through the audit risk model and its elements (see definitions in paragraph below): After defining the materiality thresholds in an audit, the auditor assesses the likelihood of the occurrence of material misstatements (i.e., risk of material misstatements). This likelihood is a function of the inherent risk of misstatements and the level of control (i.e., the product of inherent risk and control risk). The auditor performs control testing in order to provide evidence that the risk of material misstatements is at a certain level (i.e., by reducing control risk). Substantive testing is then performed in order to identify misstatements. The extent of substantive testing depends on the risk of material misstatement. When any detected material misstatements are corrected, and when the risk of further undetected material misstatements (i.e., detection risk) is judged to be sufficiently low, the auditor issues an unmodified audit opinion.¹²

In assessing risk, the audit risk model is a key tool. The model is presented since it provides a clear understanding of the role of control judgments in the overall process of reducing audit risk. Audit risk is defined as “the risk that the auditor expresses an inappropriate audit opinion when the financial statements are materially misstated (ISA glossary of terms, IFAC 2008). Audit standards require a high, but not absolute, level of assurance, but do not provide specific guidance on acceptable levels of audit risk; this is up to the auditor to judge (Eilifsen et al. 2006, 63).

Audit risk (“AR”) is a function of “the risk of material misstatement” (“RMM”) and “detection risk” (“DR”). The “risk of material misstatement” (“RMM”) furthermore has two components: “inherent risk” (“IR”) and “control risk” (“CR”). The audit risk model can therefore be specified as follows (Eilifsen et al. 2006, 63):

$$AR = RMM \times DR = IR \times CR \times DR$$

The elements of the model are: (ISA glossary of terms, IFAC 2008):

- AR is the risk that the auditor expresses an inappropriate audit opinion when the financial statements are materially misstated.

¹² This is the normal outcome of an audit. For other outcomes, the auditor is referred to read AS5 (PCAOB 2007) or ISA 700 (IFAC 2008).

- RMM is the risk that the financial statements are materially misstated prior to audit. RMM can be decomposed into IR and CR, where;
- IR is the susceptibility of an assertion to a misstatement that could be material, individually or when aggregated with other misstatements, assuming that there were no related internal controls.
- CR is the risk that a misstatement that could occur in an assertion and that could be material, individually or when aggregated with other misstatements, will not be prevented or detected and corrected on a timely basis by the entity's internal control.
- DR is the risk that the auditor will not detect such misstatement; the risk that the auditor's procedures will not detect a misstatement that exists in an assertion that could be material, individually or when aggregated with other misstatements.

In the audit risk model, internal control judgments are relevant when assessing control risk. In assessing control risk, the auditor first obtains an understanding of internal controls and evaluates their design and whether they have been implemented. If the control design is deficient and/or if the control is not implemented, no control risk reduction is achieved and the auditor performs substantive procedures in order to reduce detection risk and thereby audit risk. If the control design is effective, the control is implemented, and the auditor considers testing of controls to be better (e.g., more efficient) than substantive procedures in reducing audit risk, then the auditor tests operational effectiveness of controls in order to reduce control risk and thereby audit risk.

2.2.2 Internal Control in Auditing

Internal controls may be tested as part of an audit of financial statements. The PCAOB AS5 (2007, B4) defines this role of internal controls as follows:

“Tests of Controls in an Audit of Financial Statements: To express an opinion on the financial statements, the auditor ordinarily performs tests of controls and substantive procedures. The objective of the tests of controls the auditor performs for this purpose is to assess control risk. To assess control risk for specific financial statement assertions at less than the maximum, the auditor is required to obtain evidence that the relevant controls operated effectively during the entire period upon which the auditor plans to place reliance

on those controls. However, the auditor is not required to assess control risk at less than the maximum for all relevant assertions and, for a variety of reasons, the auditor may choose not to do so.”

However, the auditor may also perform tests of controls in an audit of internal control such as under the U.S. Sarbanes-Oxley regime. The PCAOB AS5 (2007, B1) defines this role of internal controls as follows:

“Tests of Controls in an Audit of Internal Control: The objective of the tests of controls in an audit of internal control over financial reporting is to obtain evidence about the effectiveness of controls to support the auditor's opinion on the company's internal control over financial reporting. The auditor's opinion relates to the effectiveness of the company's internal control over financial reporting as of a point in time and taken as a whole”

The role of internal controls in a financial statement audit versus an audit of internal controls thus differs along two important dimensions:

- Time period: In a financial statement audit, the auditor tests controls for the entire period on which reliance is placed on controls. In an internal control audit the auditor's opinion relates to controls at a point in time. In order to issue such an opinion, the auditor obtains evidence that internal control over financial reporting have operated effectively for a sufficient period of time, which may be less than the entire period (ordinarily one year) covered by the company's financial statements PCAOB AS5 (2007, B2).
- Extent of controls tested: In a financial statement audit, the auditor may choose to test controls for less than all relevant assertions (i.e., and instead perform tests of detail). In an internal control audit the auditor obtains evidence about the effectiveness of selected controls over all relevant assertions. This requires that the auditor test the design and operating effectiveness of controls that ordinarily would not be tested if expressing an opinion only on the financial statements PCAOB AS5 (2007, B2).

Under both the PCAOB and the ISA regime, auditors need to obtain an understanding of internal controls and evaluate their design and implementation. The extent of testing of operational effectiveness of controls, however, may vary. Under both regimes,

understanding, evaluating and testing internal controls is a key part of the audit process. The control judgments have consequences for the following audit phases:

- **Audit planning:** Control judgments have consequences for the planning of audit tests (e.g., the extent of tests of detail) and thus for audit efficiency and effectiveness (ISA 315.41, IFAC 2008).
- **Audit reporting to management and those charged with governance:** If controls are judged to be deficient, the auditor may be required to report those deficiencies to management and to those charged with governance, depending on the degree of the deficiency (ISA 315.120 and ISA 260.11, IFAC 2008; AS5.78-84, PCAOB 2007).¹³ Material weaknesses and significant deficiencies are required reported under the PCAOB regime. Under the current ISA regime, material weaknesses are required to be reported (note that significant deficiencies is not a term under the current ISA regime). If ISA 265 (Exposure Draft) is approved as existing, the ISA regime will remove the term material weakness, define significant deficiencies, and mandate reporting of significant deficiencies.
- **Audit opinion:** Under the PCAOB regime, material weaknesses in internal control are reported in the audit opinion (AS5.78-84, PCAOB 2007). ISAs do not include a requirement for reporting on internal controls deficiencies in the audit opinion.

Although the purpose of the assessment of controls is different depending on whether the scope of the audit is the financial statements or the internal controls, the kind of tests and the accompanying judgments the auditor performs are similar. In both cases the auditor assesses existence of deficiencies, and consequences and degrees of deficiencies/weaknesses.

2.3 Changes in Audit Regulation and Practice

The following section presents an overview of the development of audit regulation and practice. Such an overview is important for understanding the context in which prior internal

¹³ Similar requirements exist in the US for the audit of non-public companies through requirements in Statements on Auditing Standards (SAS) issued by the Audit Standard Boards (ASB).

control judgment research was conducted, and it may be interesting in its own right. The overview therefore starts with the time of the first internal control judgment studies in the 1970's and comments on the most important subsequent changes.

The audit risk model (ARM) was originated in the 1950's (Bell and Wright 1995), and it appeared in audit regulation for the first time in equation form in 1972 in Appendix B of *AICPA Statement on Auditing Procedure No. 54 Precision and Reliability for Statistical Sampling in Auditing*. The overview therefore assumes that the model was known to auditors and audit researchers at the time of Ashton's study of internal control judgments in 1974. In the overview, focus is therefore on regulatory and practice development relevant to control risk judgments. The overview will seek to provide insight into the following questions: First, did auditors in the early 1970's test internal control effectiveness or just evaluate design? If they only evaluated design, when did auditors start testing controls? Second, how has the extent of testing changed over time? Third, what changes have occurred in terms of the kinds of controls assessed; transaction level controls versus higher level controls such as the control environment, management controls and company level controls? Fourth, when did auditors start adopting a top-down audit approach (i.e., focus on risk management and company level controls before continuing with transaction level controls)?

Early 1970's – Limited control focus

Although the audit risk model was known to auditors (at least) since 1972 (through SAP 54, AICPA 1972), the model first became mandatory in 1984 with the issuance of SAS 47 (AICPA 1983). When did auditors start using the model and perform control risk judgments?

It is reasonable to assume that the model was generally used, at least, from 1984. However, there is no clear record of exactly what practice was regarding the use of the audit risk model's concepts prior to 1984 (POB 2000, appendix A para 13-14). Generally, although the inverse relationship between control work and substantive work had been around since before the 1970's, it was believed that audits tended to be conducted using a variety of substantive testing approaches with less reliance on judgments about control risk (ibid). This may be due to the audit risk model permitting defaulting to an assumption that control risk is maximum (ibid). Such a default assumption permits ignoring internal controls

and jumping directly to substantive audit procedures, which may be a more efficient audit approach.¹⁴ This general picture is confirmed through the descriptions of the role of internal controls in auditing found in early audit research and current audit research describing the period (e.g., Ashton 1974, Joyce 1976, Heier et al. 2005).

At the time of Ashton's study of internal control judgment in 1974, the characteristics of sound internal control were, according to Joyce (1976) well defined, and presumably widely known. According to Heier et al. (2005) there was, however, no uncontroversial definition of internal control, and disagreement existed about the differentiation between accounting and administrative controls. Early research applied tasks with transaction level controls (Trotman and Wood 1991), and not management level controls. It is therefore not unreasonable to assume that agreement existed about characteristics of sound transaction level internal controls, and that the potential disagreement regarded the more complex management level controls.

In the early 1970's, the auditor was required to review and evaluate internal control for audit planning purposes (Ashton 1974):

"The primary purpose of the internal control review is to enable the auditor to determine the particular auditing procedures to be applied, the timing of those procedures, and the extent of their application." (...) "There is to be a proper study and evaluation of the existing internal control as a basis for reliance thereon and for the determination of the resultant extent of the tests to which auditing procedures are to be restricted." (Ashton 1974, citing the second standard of fieldwork at the time)

The auditor thus documented and evaluated internal controls. However, although the auditor may have conducted control tests in order to perform a "proper study", such testing did, according to Ashton, not seem to have been performed for control risk reduction:

"The audit is conditioned by the auditor's judgment of the strength of the internal control system, regardless of the controls actually employed or the evidence gathered to evaluate them." (Ashton 1974)

¹⁴ Such an approach may have been difficult for large companies such as multinational clients, even prior to the 1970's. Some audits, especially for large companies, may therefore have included control testing.

Ashton (1974) furthermore provides an insight into the focus regarding internal controls:

“In the evaluation of a client's system of internal controls, auditors typically concentrate upon individual internal control "subsystems, for example, cash receipts, inventories, etc.”

Although the picture of audit practice in the early 1970's is somewhat unclear (POB 2000, appendix A para 13-14), the following overall conclusion is drawn: The auditor had a very narrow interpretation of risk and control, focusing on accounting error (Knechel 2007). Controls were assessed for planning purposes (i.e., to determine the extent of substantive procedures) (Joyce 1976; Knechel 2007). Focus was on transaction level subsystems and very detailed process controls (Joyce 1976; Knechel 2007). Although it might have been good practice to test controls in order to reduce the extent of substantive procedures, there is no clear evidence that this was the case. It can therefore not be ruled out that it was sufficient for the auditor to assess control design in order to reduce substantive testing (Ashton 1974). There is no indication of auditors using a top-down audit approach.

The 1970's and 1980's

The late 1970's and 1980's saw significant changes in audit and internal control regulation. These changes affected the definition of internal control, the client's responsibilities, and the auditor's responsibilities.

The first major change for companies came through the Foreign Corrupt Practices Act in 1977, which required management to develop and implement systems of internal control to reduce various risks, including a system of internal control over financial reporting. Although the act did not change the basic AICPA definitions of internal control, it put internal control on the corporate agenda (Heier et al. 2005). Further guidance was developed in 1977 when the AICPA formed a committee to provide guidance on internal control that would benefit management, boards of directors and other parties: *Report of the Special Advisory Committee on Internal Accounting Control* (AICPA 1979). Based on these two developments it is reasonable to assume that internal control systems became more structured in client companies, and that some form of holistic management level control appeared.

Soon after, the audit risk model became more prominent in audit regulation: The audit risk model (ARM) first appeared in equation form in 1972 in Appendix B of *AICPA Statement*

on Auditing Procedure No. 54 Precision and Reliability for Statistical Sampling in Auditing (AICPA 1972). The Auditing Standards Board later included a similar equation in SAS No. 39 *Audit Sampling* (AICPA 1981). With SAS No. 47 *Audit Risk and Materiality in Conducting an Audit* (AICPA 1983), use of the model became mandatory (POB 2000, appendix A para 13). It is therefore reasonable to assume that auditors started assessing control risk more formally sometime before the mid 1980's. This implied that the auditor could assess control risk at a lower level, below the maximum, by: (1) identifying specific controls that are likely to prevent or detect material misstatements relative to specific aspects of the financial statements, and (2) performing tests of those controls to evaluate their effectiveness. When auditors "relied" on controls in an area, it therefore meant that they had assessed control risk below the maximum level and had tested the effectiveness of those controls. If one accepts that, at the time of Ashton's 1974 study, the extent of substantive testing could be reduced based on control design evaluation alone, the SAS 47 requirement for testing was a significant development. The default solution of setting control risk to maximum and not relying on internal control was, however, still permitted. It is therefore, due to efficiency considerations, not given that controls were always relied upon even though effective controls may have been in place in companies.

SAS 55 *Consideration of the Internal Control Structure in a Financial Statement Audit*. (AICPA 1988) changed the definition of internal control and identified three elements of control: (1) Control Environment (2) Accounting System and (3) Control Procedures. Furthermore SAS 55 required the auditor to obtain an understanding of internal control adequate for planning the audit. It is therefore reasonable to assume that at this point, auditors had started assessing the design of management level controls such as the control environment. The auditor's responsibility was, however, primarily limited to evaluating "accounting controls", and "management level controls" were to be considered only to the degree that they had importance for financial statements (Heier et al. 2005). Furthermore, obtaining an understanding of internal control does, however, not require the auditor to reach any conclusions about the effectiveness of internal control. SAS 55 did therefore not always require testing the effectiveness of internal control. However, if control risk was to be reduced, testing of effectiveness was required.

Approaching 1990 it is therefore reasonable to assume that the importance and extent of internal controls within companies had increased. Furthermore auditor's approach to internal

controls had changed in three ways since the early 1970's: First, since firms had more controls, the auditor was presumably required to understand more controls for planning purposes. Second, the control environment was assessed in addition to the usual transaction level controls. Third, testing of control effectiveness was clearly required if controls were relied upon.

1990's – accelerating pace of change

In the 1990's the development of internal control relevant regulation accelerated. SAS 55 had been criticized for being difficult to apply in practice (Heier et al. 2005). As a response to this, and to other criticisms, COSO, which was heavily influenced by auditors, released *Internal Control – an Integrated Framework* (COSO 1992).¹⁵ The purpose of the report was to define internal control, describe its components and provide criteria and materials for evaluating internal control systems. The report thus provided an internationally accepted framework for designing and evaluating internal controls, focusing on defining internal controls as a process and including management level components such as risk management and monitoring.

The COSO dimensions and components of internal control and risk management became something that all businesses were expected to pursue as part of their operations (Knechel 2007). It is therefore reasonable to assume that the extent of company level controls in client firms increased and that they became better structured.

With the issuance of the COSO report (1992) the auditor's concept of risk and control changed (Knechel 2007). Auditors were presented with a definition of risk and control that reflected much more than accounting errors (ibid). Auditors may therefore have adopted a broader view of their responsibilities for evaluating risk and control (ibid), focusing more on company level controls and operational controls.

¹⁵ The Committee of Sponsoring Organizations of the Treadway Commission (COSO), is a U.S. private-sector initiative, formed in 1985. COSO is sponsored and funded by 5 main professional accounting associations and institutes; American Institute of Certified Public Accountants (AICPA), American Accounting Association (AAA), Financial Executives Institute (FEI), The Institute of Internal Auditors (IIA) and The Institute of Management Accountants (IMA). Furthermore, the COSO "Internal Control – Integrated Framework" (1992) was authored by Coopers and Lybrand. COSO is, and was, therefore heavily influenced by auditors.

The COSO framework was embedded in U.S. audit standards in 1995 through SAS 78 *Consideration of the Internal Control Structure in a Financial Statement Audit: An Amendment to SAS No 55* (AICPA 1995). With this standard, the definitions, dimensions and components of internal controls in the COSO report (1992) were formally recognized in audit regulation. It is therefore reasonable to assume that company level controls such as the control environment, monitoring and risk management were now a general part of the audit. The POB report (POB 2000, para 2.67), provides evidence that, at least, the control environment was generally assessed in audits at the end of the 1990's.

However, SAS 78 still only required that the auditor understand internal control sufficient to plan the audit and to evaluate whether such controls were suitably designed and placed in operation. There was still no requirement to test operational effectiveness of controls, and the default solution of setting control risk to maximum was still possible. International auditing standards at that time had similar requirements as SAS 78. It is therefore reasonable to assume that international audit approaches were similar to those found in the U.S.

How were the auditors handling the changes in regulation, practice and the general business environment? The 1980's saw rapid growth of audit practices due to deregulation, expansion of the professional pool, improvements in technology and a perceived need to reduce costs in the audit process due to competition (Knechel 2007). This resulted in an implementation of highly structured and formalistic audit processes (i.e., audit structure) that were intended to reduce the risk of serious judgment errors, reduce costs and increase judgment consistency (ibid). This development may have contributed to improved, better documented and more consistent judgments in audit firms.

However, according to the Public Oversight Board (POB), substantive procedures were still the dominating audit approach (POB 2000, para 275):

“Anecdotal and other evidence indicates that many (but by no means all) audits continued to be performed using substantive testing approaches with little or no attention paid to the results of the risk assessments called for by the audit risk model.” (POB 2000, appendix A para 14)

Furthermore, the 1980's 1990's were a period of globalization and growth, and many companies now grew to a size where substantive audit procedures were neither effective nor

efficient (POB 2000, Knechel 2007). The traditional substantive audit approach therefore started running into problems (Knechel 2007):

“The sheer volume of transactions processed by client organizations, the fast pace of technological developments affecting client organizations and audit firms alike, and economic constraints on the ability of audit firms to recover rising costs (...) led some firms to conclude that many audits were being conducted without sufficient consideration being given to the risk assessment process and that they consequently lacked in both effectiveness and efficiency.” (POB 2000, appendix A para 15)

While audit regulation was open for extensive control reliance, both at the transaction level and regarding company level controls, such an approach was not generally adopted by audit practice (POB 2000, para 275). However, the increasing recognition of problems with audit effectiveness and efficiency led to a redesign of the audit process (Knechel 2007).

In the late 1980's audit firms started (1) recruiting better educated and more mature staff, (2) placing more focus on tests of controls and analytical procedures, and (3) developing audit programs for audit testing based on more comprehensive risk assessment procedures (ibid). This development continued in the 1990's due to the introduction of the COSO report (ibid). Into the 1990's, the development of the business risk audit (or strategic systems audit) increased focus on risk and control further (Knechel 2007; Peecher at al. 2007), resulting in a more cost efficient audits with less reliance on substantive procedures (Knechel 2007).

The traditional substantive audit approach had, however, far from disappeared. The 2000 report of the POB Panel on Audit Effectiveness (POB 2000) noted that auditor's evaluations of internal controls generally were quite limited.

“Assessing control risk below the maximum level and relying on controls to reduce detailed substantive audit tests were found to be somewhat uncommon, particularly for small and medium-sized entities. Testing and relying on specific application controls were more common on larger engagements. In high-risk key areas, controls usually were not relied on in lieu of detailed tests”. (POB 2000, para 2.71)

Furthermore several problems were identified regarding the sufficiency of the depth of the auditors understanding of internal control, the extent of reliance on the control environment and other management controls, the link between risk, controls and substantive testing and

several other issues (POB 2000, para 2.77). The POB panel report (POB 2000) therefore sparked a process of regulatory change around the turn of the millennium.

2000 – ISA risk standards, SOX 404 and further COSO development

Based on the POB report (POB 2000), the IAASB and the AICPA formed the Joint Risk Assessment Task Force, with the mandate of updating audit standards governing the use of the audit risk model. This resulted in the IAASB issuing the “risk standards” in 2003.¹⁶ The purpose of the risk standards was to *“increase audit quality as a result of better risk assessments through a more detailed understanding of the entity and its environment, including its internal control, and improved design and performance of audit procedures to respond to assessed risks of material misstatements”* (Project History: Audit Risk, IAASB 2008).

For internal control judgments the risk standards had the following effect: First, the requirement for understanding the business, its risks and its controls was expanded and specified in detail through extensive guidance based on the COSO Integrated Framework (COSO 1992). This guidance results in an increased focus on company level controls such as monitoring, risk assessment and the control environment. Second, it was required that the auditor assess both risk of material misstatements, control design and whether controls were implemented. Third, control design evaluations had to be performed for all significant risks (ISA 315.113 IFAC 2008) and for risks where substantive procedures were insufficient (ISA 315.115 IFAC 2008). Fourth, documentation requirements were increased. Similar requirements were introduced in the U.S. with AU 319 (AICPA 2002) and AU 314 (AICPA 2008), apart from the requirement regarding significant risks.

At the time of the issuance of the risk standards exposure drafts in 2002, the financial scandals at the turn of the millennium were taking place (e.g., Enron, Ahold). The U.S. responded to the scandals, and the content of the POB report (POB 2000), through the Sarbanes-Oxley Act (2002). This was a major change in regulation. Section 404 of the Act required that management report on the effectiveness of its internal control over financial reporting at year end, and that the auditor attests to the accuracy of its report. The Act thus

¹⁶ The “risk standards” refer to ISA 315, ISA 330 and ISA 500. These standards replaced ISA 310, ISA 400 and ISA 401 as of 2004.

introduced a requirement an audit of internal control over financial reporting. The act furthermore introduced a separate judgment about an overall effectiveness of internal controls (i.e., material weaknesses in internal control or not). Prior to the Act, no overall judgments had been made regarding internal controls in a financial statement audit. Detailed requirements for the audit process were issued in AS2 (PCAOB, 2004) and AS5 (PCAOB, 2007). The main difference between AS2 and AS5 is that the latter emphasizes a top-down audit approach starting with risk management and entity level controls.

The act also, presumably, had a large impact on the extent and quality of internal control over financial reporting within client firms. This, together with the extensive auditor testing of internal control, facilitated the possibility of more reliance on controls in an audit of financial statements. A major change in audit strategy was therefore presumably taking place, with more controls reliance.

The ISA risk standards, SOX regulation and general pressure on auditors due to the financial scandals, litigation risk and reputation risk thus caused a tremendous effort on the part of the auditor regarding internal control assessments (Heier et al. 2005, Knechel 2007). Both under the IAASB (i.e. ISA) and the PCAOB regime much more focus was put on entity level controls and control design evaluations, and control testing increased vastly - obviously most in the U.S. where a full audit of internal control over financial reporting was mandatory (i.e., controls over financial reporting as presumably relevant for financial statements and tests of many of these controls may therefore be used as audit evidence).

After SOX (2002), internal control regulation and guidance has continued to develop. In 2004, COSO issued *Enterprise Risk Management – an Integrated Framework* (COSO 2004). Guidance for internal control over financial reporting in “Smaller Public Companies” was issued in 2006 (COSO 2006). COSO is expected to issue a similar framework for the monitoring dimension of internal control in 2008.

On the regulatory side, the IAASB is updating its standards through the clarity project, which aims at implementing redrafted standards in 2009. The purpose of the clarity project is to facilitate more consistent standard application by auditors. The IAASB has, furthermore, issued an exposure draft (ISA 265 ED, IFAC 2008) on communication of control deficiencies to management and those charged with governance, thus increasing the emphasis on internal controls in the audit even more. The PCAOB issued AS5 in 2007,

which replaced AS2 (PCAOB 2004); AS5 has a clearer focus on risk and the importance of a top-down audit approach. The EU, through the 8th directive (EU 2006, effective June 2008), introduced new requirements for public-interest entities and their auditors. Such companies must have audit committees that monitor risk and control, and they must issue a yearly report on risk management and internal control over financial reporting.¹⁷ There is, however, no requirement for management testing of controls. Furthermore, their auditors are required to report to the audit committee on material findings, including material weaknesses in internal controls over financial reporting. However, there is no requirement for an audit report on controls. The scope of the EU regulation is therefore far less than the U.S. SOX regime.

The turn of the century has thus brought an increased focus on entity-level controls and a top-down audit approach (i.e., starting with risk management and entity level controls) (PCAOB AS5, 2007; ISA 315, IFAC 2008). Board members and management are now clearly responsible for risk and control monitoring, the meaning of risk and control monitoring is clearly defined, and auditors are required to report to those charged with governance on significant/material deficiencies/weaknesses in control. This has presumably introduced judgment task characteristics that are different from what existed in the earlier transaction level accounting control judgments.

Furthermore the audit firm mergers have resulted in four major audit firms (i.e., the Big-4),¹⁸ with generally increased use of firm-wide, computerized audit training, guidance, documentation and review in order to achieve consistent, high quality judgments.¹⁹ This has presumably led to more consistent and more appropriate auditor judgments and judgment policies. Both the kind of control judgments the auditor makes, as well as how, and how well, they are performed may therefore have changed during the years that have elapsed since the most recent control judgment study in 1990 (Brown and Solomon 1990).

¹⁷ Although the directive is effective June 2008, the implementation of the requirements in affected countries may come at a later stage.

¹⁸ The term “Big Four” refers to the four large international accounting firms that remain after the collapse of Arthur Andersen in 2002, i.e., PricewaterhouseCoopers, Deloitte, Ernst & Young, and KPMG.

¹⁹ Anecdotal evidence was given by the technical partner of the firm providing audit participants for this study, and is consistent with prior research (e.g., Brazel et al. 2004)

Norway

The Norwegian Institute of Public Accountants (DnR) is a member of IFAC. DnR and their members (i.e., Norwegian auditors) are thus required to comply with the ISAs as issued by the IAASB. The importance of ISAs is furthermore evident in the Norwegian Auditing Act (revisorloven) which requires auditors to follow good auditing practice (GAP). GAP is interpreted as following Norwegian audit standards, which are translations of ISAs (for practical reasons, DnR translates ISAs into Norwegian). In Norway, regulatory development has therefore generally followed the development of the ISAs. However, some minor differences exist.

In Norway, company law requires that the board and management implements adequate internal control (“Aksjeloven” §6.12-6.15, 1997). Furthermore, requirements for book keeping are codified in a separate law on book keeping (“Bokføringsloven”, 2004). Since internal control and book-keeping presumably have an impact on the quality of financial statements, compliance with these laws is relevant for the audit. Norwegian auditing standards (RS 250, DnR 2008) therefore requires that the auditor assesses the likelihood and impact of potential non-compliance with these laws. Finally, the Norwegian Auditing Act (Revisorloven) requires that the auditor reviews (“se etter”) the administration of assets (“formuesforvaltning”) to see if it is sound. These requirements seemingly come in addition to the requirements placed on the auditor through the Norwegian translation of ISA 315. It is, however, not clear whether they imply any additional work for the auditor, since ISA 315 is already fairly extensive. This will be discussed further below.

The Norwegian audit opinion (RS 700, 2008) includes a statement that the audit has included an examination of the administration of assets (“formuesforvaltning”) and the accounting- and internal control systems, to the extent required by Norwegian generally accepted auditing standards (“god revisjonsskikk”). Furthermore, the audit opinion states that management has fulfilled its duty to provide proper (“ordentlig og oversiktlig”) registration and documentation of accounting information as required by law (i.e., “Bokføringsloven”). Non-compliance with the book keeping law (“Bokføringsloven”) must be reported in the audit opinion (RS 700.40, 2008).

Norwegian audit standards are translations of ISA’s and thus include a requirement to judge design and implementation of internal control over financial reporting (i.e., similar to ISA 315) and to test operational effectiveness of controls that are relied upon to reduce control

risk (i.e., similar to ISA 330). Material weaknesses are required to be reported to the client at an appropriate level (similar to ISA 315.120) and to those charged with governance (similar to ISA 260.11). No direct requirement exists for reporting material deficiencies in internal control over financial reporting in the audit opinion. However, such deficiencies are not unlikely to impact book keeping and may therefore be required to be reported as book-keeping deficiencies.

It is an ongoing discussion whether the Norwegian requirements for the auditors work regarding administration of assets (“formuesforvaltning”) and book keeping extend beyond what is included in international audit standards (ISA’s). The original intent of having the auditor review the “administration of assets” (“formuesforvaltning”) is to have a form of stewardship verification (i.e., verification of management's managing and control of the firm's financial affairs) (Eilifsen, 1998). Such stewardship verification may include elements of compliance and operational auditing, thereby a responsibility to detect fraud (ibid). However, it is not evident that the Norwegian auditor's responsibility for stewardship verifications has resulted in professional practice significantly different from common international practice (ibid). This dissertation shares this point of view; that ISA 315 is already so extensive that the specific Norwegian requirements for the scope of the audit do not extend beyond it. However, the specific wording in the Norwegian audit opinion regarding book-keeping and administration of assets may result in an additional requirement for reporting material weaknesses in internal control over financial reporting in the audit opinion. These minor differences in regulation are not assumed to be relevant for the study conducted in this dissertation as they are not thought to affect the form of judgment policies used when assessing internal control.

Summary

Audit approaches evolve in response to changes in regulation (Knechel 2007). Changes in the regulatory landscape since the early control judgment studies have been significant (e.g., COSO 1992; ISA 315, IFAC 2008; Revised EU 8th directive 2006; AS5 PCAOB 2007). Internal control judgment task characteristics may therefore have changed fundamentally compared to the situation when relevant prior research was conducted (i.e., pre 1990’s). One example is the introduction of audit reports on internal controls under SOX 404. Another example is the increased importance of entity-level controls, which include multi-step control processes like risk assessment and monitoring (COSO 1992; ISA 315, IFAC 2008;

Revised EU 8th directive 2006; AS5 PCAOB 2007). Changes in audit regulation and audit practice may therefore have changed both the judgment task (i.e., the task characteristics and the accompanying appropriate judgment policies) and the behavior of the judge (i.e., increased use of appropriate judgment policies compared to findings in Brown and Solomon (1990)).

3. Literature Review and Theory Development

3.1 Introduction

Research aimed at understanding how judgments and decisions are made and how they can be improved can be classified as belonging to an area of psychological research called 'behavioral decision theory', which has its roots in cognitive psychology, economics and statistics (Libby 1981, 2; Trotman 1998). Behavioral decision theory uses many approaches (e.g., formal probability theory, Bayesian probability theory, utility theory), but any approach where data from human judgment is analyzed by using multiple regression equations belongs to the general methodology of 'judgment analysis', which is also commonly called the 'policy capturing paradigm' (Cooksey 1996, xi). The purpose of this paradigm is to develop algebraic models that describe the method by which individuals weight and combine information (Slovic et al. 1977). The approach has evolved over many years:

“As a system for the study of human judgment, policy capturing has roots traceable to Hoffman’s (1960) article on the paramorphic²⁰ representation of clinical judgment and, through Hoffman, has indirect links to Brunswick’s (1947) probabilistic functionalism and Hammond’s (1955) application of probabilistic functionalism to the problem of clinical judgment” (Cooksey 1996, 57).

Most policy capturing research has used ANOVA techniques to analyze data (Trotman 1998). This technique was introduced to the judgment literature by Hoffman, Slovic and Rorer (1968) and it was first introduced to the audit literature by Ashton (1974). Before this, no systematic research on audit judgment had been carried out (Trotman 1998); instead, answers to research questions had traditionally been supported by informal consensus of practitioners' experience (Libby 1981, 2).

²⁰ Paramorphic means that the judgment policies do not necessarily represent the actual cognitive processes of the judge, but only the surface statistical relation between input and output. This is however still useful as a starting point for improving judgments. For more on paramorphic representation see for example Hoffman (1960), Ashton (1982), Trotman (1996).

Ashton's (1974) study of auditor's internal control judgments had a significant impact on the audit literature, and the degree of interest in research on auditor judgment increased substantially (Trotman 1998). The policy capturing approach has been used to model such auditor judgments as audit planning (Joyce 1976); materiality (Messier 1983); evaluations of internal auditing (Abdel-khalik, Snowball and Wragge, 1983); uncertainty disclosure decisions (Libby, 1979); inherent risk (Colbert, 1988); and analytical review judgments (Brown and Solomon, 1991).²¹

In the literature subsequent to Ashton (1974), one of the key issues addressed has been whether auditors integrate information in a configural manner (Trotman 2005). Research on cue combination has primarily used internal control judgment tasks (Brown and Solomon 1990; Solomon and Shields 1995; Trotman 1998, 2005). Although there is some evidence of auditor's ability to process cues configurally (e.g., Brown and Solomon 1990, 1991), many questions still remain unanswered:

“Research is needed to specify the different attributes of auditors who process configurally and those that do not. Also, what factors increase the likelihood of auditors processing configurally? Can configural processing be increased by training and the other decision aids? Does learning to process configurally on one task lead to configural processing on other tasks?” (Trotman 1998)

Furthermore this dissertation questions the relevance of prior research in describing audit judgments in today's audit environment. This will be treated in the discussion section following the literature review.

The following literature review covers both psychology and audit research. The review of psychology research focuses on methodology and general findings from the policy capturing paradigm regarding configurality. The review of audit research primarily concentrates on policy capturing studies of internal control judgments, but also discusses relevant findings from research on other judgment tasks where configural cue usage has been found.²²

²¹ The overview of judgment settings is based on Trotman and Wood (1991) and Solomon and Shields (1995, p158).

²² Internal control judgments are the dominant context for judgment rule studies in auditing (e.g., see review in Shields and Solomon 1995).

Relevant changes in audit practice and regulation will be discussed in relation to their effect on the task characteristics of focus in this dissertation.

An introductory overview of the main departing points for this dissertation is provided in order to help readers see where this dissertation is “coming from”. The main departing points within the psychology literature are as follows: First, Brunswick’s Lens Model (1952) provides the conceptual framework for judgment analysis. The framework’s components consist of the criterion event (i.e., the true state of what the judge is making a judgment about), the information set on which the judge bases his judgment, and the judgment itself. Regression equations and correlations between these components can be used to study various aspects of judgment making (e.g., accuracy, consistency, and cue usage). The framework is thus suited for the main purpose of this study, which is to study configularity in internal control judgments.

Second, Einhorn (1970 and 1971) and Elrod et al. (2004) provide models for describing task characteristics and judgment policies. These models are useful for structuring and discussing task characteristics that may occur in audit practice, such as control interrelationships and the judgment response scale. Furthermore, mathematical models of judgment policies are a useful departing point for developing expectations for what specific judgment policies should look like in the form of Lens Model regression equations. The models are thus important for operationalizing the dependent variable of this study (i.e., the functional form of the judgment policy)

Third, Brehmer (1994) and Stewart et al. (1997) provide a psychological theory background for looking to the task characteristics when analyzing expert judgment. This provides general motivation for a more thorough study of task characteristics in internal control judgments.

Finally, Cooksey (1996) provides a comprehensive overview of the methodology and history of policy capturing research on judgments. This is useful for understanding the place of policy capturing research within the wider domain of behavioral decision theory. Furthermore Cooksey’s practical solutions to issues with policy capturing experiments have helped resolve design and methodology issues in this study.

From the audit literature the main departing points are: First, Brown and Solomon's (1990 and 1991) studies of the effect of compensating and amplifying cues on configural cue usage. These studies are important for several reasons. They introduced cue interrelationships as an important task characteristic in auditor judgments. They developed the first causal theory for determinants of the functional form of the judgment policy in auditor judgments. Furthermore, no studies of configurality in internal control judgments have been identified since Brown and Solomon (1990). Their 1990 study is therefore, presumably, the first and only evidence of configurality in internal control judgments.

Second, the literature review in this study is based on literature reviews of audit judgment research by Solomon and Shields (1995), Trotman (1991, 1996, 1998 and 2005), and Bonner (2007). These literature reviews are extensive, and there is little to add to their review of relevant results prior to 1990 (i.e., lack of configurality in auditor judgments). This dissertation therefore focuses on Brown and Solomon's 1990 study and subsequent development.

Third, the call for more focus on the task in audit judgment research (Gibbins and Jamal 1993, Bonner 1994, Trotman 2005) provides audit specific motivation for developing more extensive theory on task characteristics in internal control judgments.

Motivation: Improving auditor judgment and decision making:

The aim of behavioral audit research is to improve auditor's judgment and decision making (Trotman 1998; Bonner 1999, 2007). Libby (1981, 2) suggests three ways for improving judgments:

1. By changing the information (i.e., changing the cues)
2. By educating the decision maker to change the way information is processed
3. By replacing the decision maker with a model

If judgment and decision making is to be improved through improved information processing (item 2 above), then a necessary first step is to find out how information should be processed, how it is processed and compare the two. If it is found that judgment policies lack appropriate configural cue processing, then research can turn to developing theory about circumstances that lead to increased use of configural cue processing (e.g., by

providing feedback comparing the individual judge's judgment policy with the normatively appropriate judgment policy – i.e., cognitive feedback).

While behavioral audit research focuses on improving both judgments and decisions, this dissertation focuses only on judgments. Judgments can be discriminated from decisions in the following manner (Bonner, 1999; Peecher and Solomon, 2001):

The term *judgment* typically refers to forming an idea, opinion, or estimate about an object, an event, a state, or another type of phenomenon. Judgments tend to take the form of predictions about the future or an evaluation of a current state of affairs.

The term *decision* refers to making up one's mind about the issue at hand and taking a course of action. Decisions typically follow judgments and involve a choice among various alternatives based on judgments about those alternatives and preferences about factors such as risk and money. In other words, *judgments* reflect one's beliefs, and *decisions* may reflect both beliefs and preferences.

Motivation: The importance of audit task characteristics

Psychology research has a longstanding tradition of focusing on task characteristics. For example, Simon (1956) emphasized development of theory about how environmental characteristics affect judgment, and Brehmer (1994) stated that judgment analysis research could go no further without focusing more on the task. This is relevant for audit research since audit judgment settings have unique features for which theories in basic science disciplines are not always well developed (Bonner 1999). For example, the various kinds of internal control interrelationships and judgment response scales existing in auditing have not been studied in psychology research as generic task characteristics that may combine in affecting judgment policies.

Prior audit judgment and decision making research has been conducted within frameworks where task characteristics are viewed as important causal variables for judgment performance (see frameworks in Libby and Lewis 1977 and 1982; Libby 1981; Bonner 1994; Solomon and Shields 1995; Nelson and Tan 2005). However, early policy capturing research in auditing (i.e., before 1990) was mainly descriptive. Tasks were generally chosen because of convenience or because of a good fit with the psychological phenomena behind the theory development, rather than the researcher searching to learn something about a

particular audit judgment task (Trotman 2005). Some studies after 1990 have looked at the impact of task characteristics on the use of configularity in judgment models (e.g., Brown and Solomon 1990 and 1991), and found support for task characteristics affecting configularity. Another motivation for this dissertation is therefore to increase knowledge about audit task characteristics and extend theory of how audit specific task characteristics affect audit specific judgments. Such theory development is grounded in the audit task, rather than grounded in theory from psychology. However, audit task characteristics may possess generic features, so that a contribution may also be made at the generic theory level. Such a task-oriented approach is consistent with the call in psychology literature for more focus on the task (e.g., Einhorn 1971, Brehmer 1994, Stewart et al. 1997). Several audit researchers have also argued for the same:

Gibbins and Jamal (1993) argued for much more extensive task and content analysis in order to produce better theory in accounting judgment and decision making research:

“Understanding behavior follows from an analysis of the task. Theory development should therefore be grounded in the task, rather than grounded in theory from psychology and reaching out to accounting through accountants.”

Bonner (1994) developed a model of the impact of audit task complexity on judgment performance. It was argued that task variables deserved more attention as it could affect the interpretability of research results as well as auditors’ performance on audit judgment tasks. The purpose of her study was to develop a model that future audit research could apply in studying the effects of task characteristics on judgments. The contribution was therefore to structure and highlight the importance of task characteristics in audit judgment research.

Solomon and Shields (1995) suggest use of task analysis to identify the information processing demands for successful task completion, and advocate a generally increased attention to audit task content.

Nelson and Tan (2005) review audit judgment and decision making research from three perspectives, including the task perspective. Internal control judgment is not viewed as a separate task in the review, but their general emphasis of focusing on the audit task is clear.

In a discussion of the Nelson and Tan (2005) review Trotman (2005) considers the separate emphasis of the task perspective as a strength of the review, and calls for more development

of theory about audit tasks. Emphasis is put on the importance of considering how specific audit tasks are different from generic judgment and decision making tasks and how these differences impact the processing of information and the judgments made.

To summarize: frameworks for studying the impact of task characteristics on judgments have a long history in audit research. However, the task frameworks have not been applied to any significant extent in developing causal models for the effect of task characteristics on judgment models. A clear call for more task focus in audit judgment research therefore persists. This dissertation is partly motivated by contributing to theory about audit tasks.

3.2 Policy Capturing: Approach and Methodology

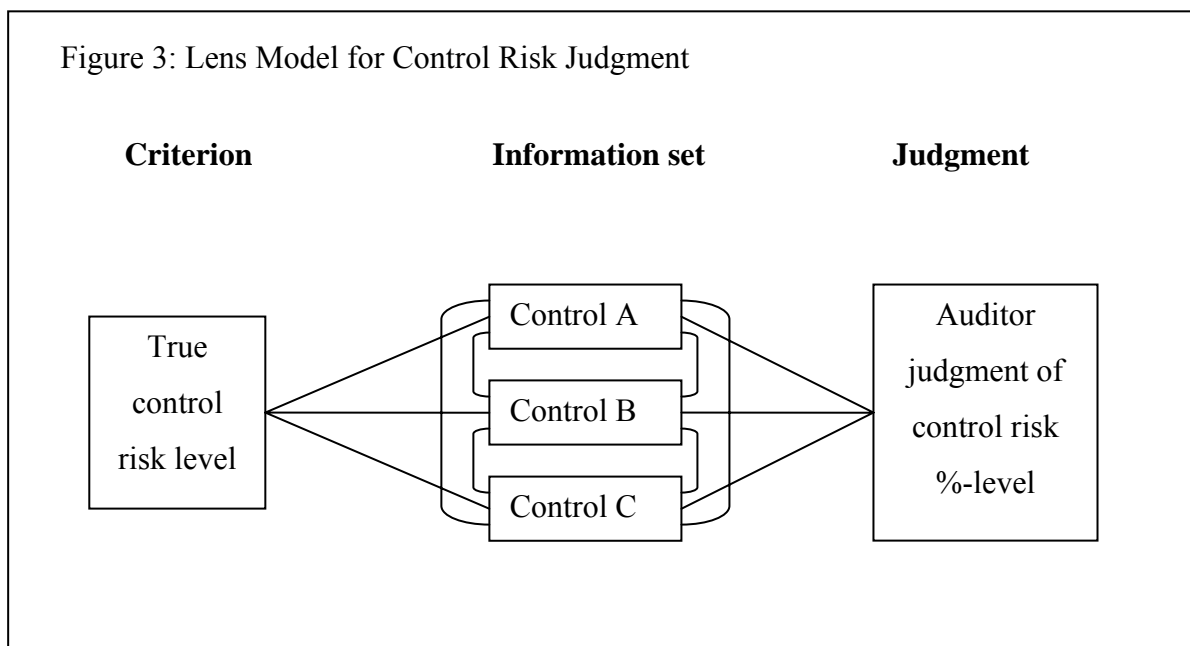
Policy capturing is a method to assess how judges use available information when making evaluative judgments (i.e., how they weight, combine or integrate information cues) (Karren and Barringer 2002). Judgment policies can be studied at the individual level by analyzing each judge's policy (i.e., ideographic level analysis) or at the aggregate group level by analyzing the group's judgment policy (i.e., nomothetic analysis).

Policy-capturing evolved from Egon Brunswik's probabilistic "lens" model (Brunswik 1952), which is a model of individual perception.²³ The Lens Model consists of three elements (Libby 1981, 19) (see figure 3 below):

1. The criterion; which is the true state of the reality that the judge is making a judgment about (typically an event or state). Although the level of the criterion is unknown to the judge, the nature and range (but not the level) of the criterion is known through knowledge of the judgment response scale (e.g., the judge knows that the nature of the criterion is control risk and that the range is a percentage scale since this is what he is asked to make a judgment about, but he does not know that the true level of control risk is e.g., 27%). The nature and range of the criterion are therefore important task characteristics.

²³ For an introduction to the Lens Model and policy capturing, see Libby (1981, 19) Trotman (1996, 33), Cooksey (1996), Aiman-Smith et al (2002).

2. The information set; which contains probabilistic information about the true state of the criterion. An example could be cues that contain information about the test results from auditor testing of individual controls, which the auditor is supposed to combine into an overall estimate of control risk.
3. The judgment; which is the judge's estimate of the true state of the criterion. The judge is required to provide a judgment on some form of judgment response scale. The judgment response scale as a task characteristic is the nature of the auditors' judgment response, and the range is the number of judgment response options the auditor has available for a given judgment task.. In the example in figure 3 below, the judgment response scale is a 100-point percentage scale indicating the level of control risk.



The Lens Model can thus be depicted within a linear framework that defines both judgments and the criterion as functions of cues in the environment (Hogarth and Karelaia 2007). On the environment side of the Lens Model individual cues may relate to the criterion through linear, quadratic or higher order relationships, and cues may also interact in their relationship to the criterion.²⁴ Similarly, the judge can process cues by using simple linear judgment

²⁴ It can be noted that when cues are binary (i.e., coded as "0" or "1") quadratic and higher order relationships are not relevant (i.e., squares and higher order powers of "0" and "1" remain equal to their first order value). Interactions are, however, relevant.

policies or by using more complex configural judgment policies. In general, judgmental accuracy therefore depends on matching characteristics of judgment policies and environments (Hogarth and Karelaia 2007).

Policy capturing is performed by having the subject(s) judge each of a series of cue profiles and then regress the judgments on the cues in order to derive a weighted linear composite which represents (captures) the subject's method (policy) for combining cue information into a judgment (Cooksey 1996, 57). Even though the functional form of a cue's relationship to the judgment is nonlinear (i.e., a power polynomial) a linear regression model is fitted. This is done by including squares, cubes or higher order polynomials of the cues in the linear regression. The same approach is used to handle cue interactions; product terms of cues are included in the regression. A linear regression model can therefore handle additive nonlinear judgment models (Cooksey 1996, 177).

Policy capturing is performed by using cases that are representative designs or systematic designs. Representative designs attempt to maintain the realistic statistical properties of cues and events (e.g., correlation between cues). This may be done by (1) abstracting cases from a representative group of actual cases or by (2) fabricating cases that represent realistic relationships. Systematic designs, however, do not try to maintain statistical properties of real life cases. Instead they often build on orthogonal designs (i.e., where cues are not correlated). If cue usage is the focal point of the study, a systematic design is necessary. This is due to the difficulty in cue weight interpretation created by the correlation between cues that is usually found in the environment (Libby 1981, 39). Representative designs are, however, preferable if judgment performance is the focus of the study such as in this study. This is due to the non-existence or fabrication of criterion values in fabricated cases (Libby 1981, 41).

Policy capturing can provide insight into the following research questions (Cooksey 1996, 56-73; Trotman 1997, 35-40):

1. Accuracy of judgments, when a criterion/reality exists and is known. By correlating the judgments for the series of cue profiles to the true criterion an achievement index can be calculated. The achievement index shows the relation between the judgment and the true criterion. In auditing, the true level of the criterion is often unknown. Studies of accuracy have therefore been rare in auditing (Trotman, 1998).

2. Consensus in judgments between judges. Consensus is an often used proxy for accuracy when a criterion is unknown (see Ashton 1985). Consensus in judgments is also named consensus in fact, which differs from consensus in principle in that the former regards similarity of judgments while the latter regards the similarity of the judgment policies (i.e., similarity in how cues are weighted and combined) (see nr 8 below). Consensus is calculated by correlating the judgment responses between judges. It is also possible to compare judgments to the judgments of a panel of experts. Such an approach uses the expert panel's judgments as a surrogate for accuracy. High consensus with the panel is thus interpreted as being consistent with high accuracy (Trotman 1996, 37).
3. Weighting of cues. Cue weights indicate if a cue is used to reach a judgment and how much weight is being placed on that cue. In multiple regression, standardized regression weights can be interpreted as cue weights.
4. The nature of the functional form relating each cue to the judgments made (i.e., linear or nonlinear square, cubic or higher order polynomial).
5. Organizing principle for cue integration (e.g., linear or configural) (this is the main focus of this dissertation). While the functional form discussed above relates to individual cues, the organizing principle for cue integration relates to how cues interact in their effect on the judgment. Standardized regression weights for interaction terms can be interpreted as configural policy elements in multiple regression.
6. Judgment insight; i.e., can judges describe or recognize their judgment policy? This is performed by asking judges about their judgment policy and comparing this to judgment models from analysis of their judgments.
7. Cognitive control in the form of judgment stability (i.e., the same cue profile is consistently judged by the same judge when repeated). This is performed by giving the judge repetitions of the same case and correlating judgments. This can be done within the same sitting, or after some time has elapsed (e.g., Ashton 1974 provided participant's with the same cases after approximately two months).

8. Similarity of judgment policy between judges (i.e., consensus in principle). A consensus about the judgment policy implies that cues are weighted and combined in the same manner by judges. This is assessed by comparing regression coefficients in judgment models.

Audit research using the policy capturing method has focused on tasks where there is no objective criterion (Ashton 1985). This is due to the nature of auditing where the true state of the criterion is generally unknown (e.g., it is not possible to know the true control risk of an audit client). The purpose of such a research design is to capture and compare a number of individual judgment policies (e.g., cue weighting, consistency, consensus, insight and cue combination) (Cooksey 1996 p.77).

3.3 Findings in Psychology Research

3.3.1 General Findings Indicate Linear Judgment Policies

Cooksey (1996, p.79) quotes Hammond et al. (1975) as succinctly summarizing the result of policy capturing studies comparing judges:

“Empirical regularities include the following general conclusions: (1) people do not describe accurately and completely their judgment policies, (2) people are often inconsistent in applying their judgment policies, (3) only a small number of cues are used, (4) it is difficult to learn another person’s policy simply by observing his judgments or by listening to his explanations of them (5) cognitive aids can reduce conflict and increase learning, and (6) linear additive organizational principles (i.e., nonconfigural information integration) are often adequate to describe judgments.”

Other reviews are consistent with this and state that findings generally indicate that linear judgment models can describe most judges and that experts are limited in the same manner as novices (Einhorn 1971; Slovic et al. 1977; Shanteau and Stewart 1992; Brehmer 1994).

3.3.2 Extent of Configurality may be Underestimated

The validity of the linear model in representing human judgment processes has, however, been questioned for many decision-making tasks (Hogarth and Karelaia 2007). This is due to linear judgment policies being cognitively demanding to execute (Elrod et al. 2004; Hogarth and Karelaia 2007). Judges may therefore resort to simplifying heuristics when the amount of information increases (e.g., more than three cues) or when they find the trade-offs involved in linear judgment policies to be too difficult (cognitively or emotionally) (Hogarth and Karelaia 2007). Such simplifying heuristics may imply configural cue processing (e.g., judgment policies where judgments are based on only one cue and remaining cues are ignored).

Furthermore, Cooksey (1996, 183) cautions that the lack of findings supporting configurality may be due to judgment analysis research designing out configurality by way of cue selection (i.e., configurality is not appropriate or expected in solving the experimental tasks since the relationships between the criterion and the cues is linear). This view is consistent with Brown and Solomon's (1990 and 1991) explanation for not finding configural auditor judgments; it was not appropriate for the given tasks (i.e., the tasks were not designed in a way that made configurality appropriate).

Finally, the accuracy of judgments depends on both the inherent predictability of the environment and the extent to which the weights humans attach to different cues (and their interactions) match those of the environment (Hogarth and Karelaia 2007). In judgment tasks where the relationship between cues and criterion is nonlinear, configural judgment policies should therefore be appropriate. It is not unreasonable to assume that experts over time learn to apply such configural judgment policies when appropriate to the environment. Since most research has used tasks with linear cue-criterion relationships (Cooksey 1996, 183), the potential use of configural judgment policies in tasks with nonlinear cue-criterion relationships may have been overlooked.

Empirical psychology research has found some evidence of configurality. Examples include studies of stockbroker's judgments (Slovic, 1969), of psychiatric medical professionals (Rorer et al. 1967), and of moral judgment (Leon et al. 1973) (Brown and Solomon 1990). Studies focusing specifically on the use of noncompensatory models (e.g., Einhorn 1970 and

1971) have also found some supportive evidence, but further evidence is needed before definite conclusions can be reached (Cooksey 1996 p.185).²⁵

Cooksey's (1996 p.183) refers to Slovic and Lichtenstein (1973) for an early and comprehensive review of configural cue usage and Stewart (1988) for further discussion. His own summary of findings and the state of research reads as follows (Cooksey 1996, 183-185):

“While one can frequently find support for the existence of configural cue usage (...) the contribution from such usage is typically quite small compared to the overall contribution of linear main effect cue usage (...) While current judgment research comparing linear judgment representations to configural or nonlinear representations is relatively rare, it is fair to say that the balance of the evidence favors the use of linear compensatory judgment models to capture judgment policies. Nonlinear and noncompensatory models may, for certain judges under certain specific conditions, yield marginally better predictive power for the sample of cases on which the model is constructed, yet these models will frequently not cross-validate as well as the more parsimonious linear model. Nevertheless, judgment analysis should routinely investigate such models rather than assuming apriori that they will not be appropriate.”

Cooksey therefore urges researchers to continue looking for configural judgment policies even though they historically have been found to be of little importance and sample specific.

Overall, psychology research gives an impression of unease about the general finding of linear judgment policies. This unease is reasonable given (1) experimental tasks generally including linear cue-criterion relationships, and (2) the extensive heuristics and biases literature, where explicit recognition is given to the limits of human information processing (i.e., use of simplifying heuristics that may involve configurality) (Hogarth and Karelaia 2007).

²⁵ In a compensatory model the judgment, based on any cue, may be offset by considering one or more of the other cues. In a noncompensatory model, the judgment may be determined by the level of only one of the cues, irrespective of the level of other cues.

3.3.3 Development of Judgment Models

Parallel with the search for configural judgment policies, a related research stream has focused on developing mathematical models of judgment policies (Einhorn 1971; Cooksey 1996; Elrod et al. 2004). Valid mathematical models of judgments are important because they provide precise specifications of theory (Elrod et al. 2004). The availability of mathematical models also allow researchers to infer the unobserved judgment policy from the observed judgments, eliminating reliance on self-reports, protocol data or multiple observations of intermediate steps which are often unavailable or unreliable (Elrod et al. 2004).²⁶ Even though the models do not measure the actual mental process which produces the judgment (i.e., they only provide surface relationships between inputs and outputs), the implications of different types of models may be important in terms of the mental processes they suggest (Libby 1981, 44). Libby (1981, 44) suggests that the most relevant models in accounting research may be additive compensatory models with positive or negative interactions, and conjunctive and disjunctive models (these models will be defined under the theory development section of this dissertation).

3.3.4 Summary of Psychology Research

The general finding from psychology research is that linear models describe human judgment (Hogarth and Karalaia 2007). Furthermore, configurality is not beyond human judges, it is just not very typical of human judgment (Brehmer 1994; Cooksey 1996). Overall, however, an impression of unease about this general picture exists. First, the heuristics and biases literature suggests widespread use of heuristics that may involve configural cue processing. Second, experimental tasks may have made linear judgment policies appropriate. Third, some evidence of configurality exists, especially for noncompensatory models.

In order to answer questions about whether there will be configural components in judgment models, consideration of the characteristics of the specific task is therefore needed (Brehmer 1994; Cooksey 1996; Stewart et al. 1997). Furthermore, nonlinear, noncompensatory models

²⁶ Self-reports may be unreliable because subjects may be unaware of their own judgment policies or unable to report them accurately. Methods for collecting protocol data may interfere with the decision process they measure (Elrod et al. 2004).

may be useful for describing potential judgment policies (Einhorn 1971; Libby 1981, 44; Cooksey 1996; Elrod et al. 2004).

3.4 Findings in Audit Research

Prior literature reviews of behavioral audit research identified a total of 28 policy capturing studies (i.e., studies empirically modeling auditor's judgment and decision policies) (Trotman and Wood 1991; Solomon and Shields 1995, 152). 14 of the 28 studies used an internal control judgment setting, while the remaining 14 studies used other settings. 21 of the 28 studies examined cue usage (i.e., 14 in an internal control setting and 7 in other settings). Apart from Brown and Solomon (1990 and 1991) no consistent evidence of configural cue processing was found. Since Brown and Solomon (1990) no studies focusing on judgment policies in internal control settings have been identified. Furthermore, no audit studies focusing on configurality, regardless of setting, have been identified by the author since Brown and Solomon (1991). However, some more recent studies include findings that may be relevant for this dissertation. The following literature review will therefore briefly summarize the findings prior to 1990. Brown and Solomon's study in (1990 and 1991) will thereafter be discussed in detail. Finally, other potentially relevant audit research conducted after 1990 will be discussed.

3.4.1 Initial Descriptive Findings; General Use of Linear Judgment Models

Audit judgment research using the policy capturing paradigm was initiated in 1974 with Ashton's descriptive study of auditor's internal control judgments. The main focus of the study was on consensus in auditor judgments. The form of the judgment policy was one of several other questions examined, and it was generally found to be linear.²⁷ Many subsequent policy capturing studies of auditor judgments followed in the 1970's and 80's. Trotman and Wood (1991) provide a compact overview:

²⁷ In Ashton (1974), the 15 interaction terms accounted for only 6% of the judgment variance, while the six main effects accounted for 80%. Furthermore, consistency between auditors was low with respect to which interaction terms that were important. The single most important interaction term was significant for only 15 out of 63 auditor models, and accounted for less than 4% of those participants' judgment variance (Brown and Solomon, 1990).

“Studies investigating the generality of Ashton's (1974) consensus results have considered the moderating effects of more complex cases (Reckers and Taylor 1979), student surrogates (Ashton and Kramer 1980), an increased number of cues (Ashton and Brown 1980), a wider range of audit experience (Hamilton and Wright 1982), personality variables (Hall, Yetton, and Zimmer 1982), time pressure (Choo and Eggleton 1982), interacting and composite groups (Trotman, Yetton, and Zimmer 1983), more realistic cases and the review process (Trotman and Yetton 1985), and multiple systems (Choo and Eggleton 1983).”

Cue usage was often studied as one of many issues in these studies, but no consistent evidence of configural judgment models was found (see reviews in Libby and Lewis 1982; Brown and Solomon 1990; Solomon and Shields 1995; Trotman 1998, 2005; Bonner 2007). Findings therefore seemed consistent with psychological research in that linear models were sufficient for describing auditor's judgment policies (Brown and Solomon 1990 and 1991).

The studies from the 1970's and 80's will not be discussed in further detail for three reasons: First, the studies focused primarily on judgment consensus. Although configural cue processing was discussed, the experiments did not include interrelated cues. Without variation in cue interrelationships the findings are of little relevance for this dissertation. Second, audit regulation and practice have changed to such a degree that the findings may be of little relevance for today's auditors. Third, audit judgment literature relevant to this dissertation has been extensively reviewed in dedicated literature reviews and in the literature review sections of conducted studies. These reviews give an extensive overview of the field and the place and state of policy capturing studies of auditor's internal control judgments. Providing yet another review of these older studies would therefore not add to knowledge: the older studies were based on experiments with linear cue interrelationships and configurality did not occur. Trotman (1998, 2005) and Bonner (2007) provide the most updated literature reviews that include findings from policy capturing studies of auditor's internal control judgments.

3.4.2 Development of Normative and Causal Theory

The development of theory about how cue interrelationships affect judgment policies in auditor's judgments started with Brown and Solomon's (1990) internal control judgment

study. They developed theory for how task characteristics (i.e., interrelationships of control cues) should affect the form of the judgment policy.

When analyzing prior policy capturing studies in auditing, they discovered that cue interrelationships were not systematically designed into the experiments, and that control cues were not interrelated in their effect on risk (i.e., the relationship was linear). Tasks were therefore of a kind where linear cue usage would be appropriate for making judgments (i.e., configularity would not be needed). The lack of findings of configularity in previous audit research could therefore be due to task characteristics (i.e., lack of control cue interrelationships). Since no counterfactual existed (i.e., no studies included interrelated cues), it was still an unanswered question whether an experiment with interrelated control cues would result in configural judgment models.

Brown and Solomon (1990) therefore designed judgment tasks where specific control cues were interrelated in the form of compensating and amplifying controls while remaining control cues were independent.²⁸ Configularity of a specific form and nature would thus be appropriate. Reported results from ANOVA analysis of each auditor's judgment model revealed that a high proportion of auditors (40.5%) exhibited at least one of the two expected forms of configural information processing.²⁹ Furthermore, predictive judgment models revealed significant differences in risk judgments depending on whether interaction terms were included in the prediction model. It was therefore concluded that auditor's configularity might be more prevalent than previously recognized. However, even though 40.5% processed cues configurally this does not mean that the observed *form* of configularity was appropriate given cue interrelationships. This observation will be discussed further in the discussion section of this literature review.

²⁸ Independent control cues imply that cues are independent (i.e., linear) in their effect on risk. Compensating control cues imply that cues have a nonlinear compensatory form ordinal effect on risk. Amplifying cues imply that cues have a nonlinear amplifying form ordinal effect on cues. These forms of control cue interrelationships will be discussed in the theory development section of this study.

²⁹ Configural cue processing was defined as presence of interaction terms that individually account for more than 4% of judgment variance. Since each auditor observation cell had only one observation, the models were fully determined and had no error estimate. Formal significance testing could therefore not be performed. The motivation for using 4% as a criterion is discussed in Brown and Solomon (1990 and 1991) and in Hooper and Trotman (1996); findings show that 4% judgment variance can have significant impact on judgments and therefore is a relevant criterion.

Brown and Solomon (1991) conducted two separate experiments using misstatement risk judgment tasks (i.e., the risk of material misstatement after the auditor has performed substantive/analytical testing). The setting was therefore not internal control judgments. The first experiment included five cues, where two cues were predicted to be processed configurally in a compensating manner. Results showed that 27 out of 49 auditors (55.1%) had the predicted interaction. Furthermore all 27 interactions were in the appropriate compensating manner. The second experiment included five cues, where two cues were predicted to be processed configurally in an amplifying manner. Results showed that 21 out of 22 auditors (95.5%) had the predicted interaction. Furthermore 20 out of the 21 auditors having the interaction had it in the appropriate amplifying manner. Overall, it was concluded that many auditors configurally process information in a predictable and appropriate manner, given the interrelationship of cues, and that lack of configularity in judgment models can lead to judgment differences with significant consequences for audit efficiency and effectiveness.

3.4.3 Subsequent Audit Research with Relevance for Cue Processing

After Brown and Solomon (1990 and 1991) it was accepted that many, but not all, auditors are able to process cues configurally (Trotman 1996, 106). A call was therefore made for directing research towards (1) factors that increase the likelihood of configularity, and (2) why some auditors process information configurally while others do not (Brown and Solomon 1990). In 1998, in a review of policy capturing research, Trotman continued to direct attention to circumstance that lead to configularity and specified three areas of future interest: (1) attributes of the auditor, (2) the effect of training and decision aids on configularity, and (3) whether the learning of configularity from one task transfers to other tasks. Below, relevant findings from audit research after 1990 are reviewed, including research on internal control tasks and other audit tasks.

Bedard and Biggs (1991) studied the auditors' process of cue pattern recognition and hypothesis generation in an analytical review task and found that auditors generally attempt to use all cues in forming a pattern diagnostic of an overhead application error, thus showing evidence of configularity.

Maletta and Kida (1993) studied whether inherent risk factors, control risk factors and internal audit quality components are combined configurally in audit planning judgments about whether to rely on internal audit functions to reduce planned audit work. Findings revealed that: (1) auditors show an interaction effect of inherent risk and control strength on the extent to which they rely on internal audit functions to reduce planned audit work, and (2) interactions between inherent/control risk and internal audit quality components (competence, objectivity and work performed) when making reliance decisions. Auditor's ability for configural cue processing was therefore found to be present in this judgment setting.

Hooper and Trotman (1996) replicated experiment one in Brown and Solomon (1991) (i.e., a misstatement risk judgment task with two compensating cues), but with modified and varying experimental procedures. The proportion of auditors showing above criterion judgment variance attributable to interaction terms was above chance (22% showed configularity in the treatment group that was not required to give explanations for judgments, while 37% showed configularity in the treatment group that was required to give explanations for judgments). Furthermore, analysis of participants' documented reasons for judgments showed that, overall, the main difference between configural versus non-configural judges was that the former considered cues to be interrelated while the latter did not (i.e., the judge's perception of cue interrelationships may be important for the occurrence of configularity). Finally, the level of judgment consensus was higher for auditors who processed information configurally compared to those who did not process information configurally, thus indicating improved judgment performance.³⁰

Leung and Trotman (2005) studied the effect of four different types of feedback on auditor judgment performance. The task was a modification of experiment one in Brown and Solomon (1991) (i.e., a misstatement risk judgment task where one treatment included compensating cues while the other treatment had only independent cues). Three of the feedback types provided information about the appropriateness of judgment policies (i.e., cue interrelationships) and/or the form of the individual auditor's judgment policy. The study found that auditors are capable of configularity. Furthermore findings support the

³⁰ See Ashton (1985) for a discussion of why consensus is a good surrogate for judgment performance.

importance of understanding cue interrelationships for (1) deriving apriori appropriate judgment policies, (2) communicating these through task property feedback, (3) understanding the meaning of cognitive feedback in configural tasks, and (4) improving performance in configural tasks.³¹

Hammersley (2006) studied how well auditors interpret incomplete cue patterns. The study examined experimentally whether industry-specialist auditors develop problem representations about a seeded misstatement to facilitate interpretation of incomplete patterns. Findings showed that an auditor who is a specialist in the clients industry is able to process cues configurally, even when he only knows part of the cue pattern, but when he is not an industry specialist he is not able to process cues configurally, even when he knows all of the cue pattern. The study is therefore consistent with auditors being able to process cues configurally.

3.4.4 Summary of Prior Audit Research Findings

The current status of audit research relevant for internal control judgments is therefore that (1) auditors use linear additive cue processing unless experimental tasks are constructed to allow the judge to exhibit configural cue processing (Bonner 2007, 155), (2) even though configural cue processing is appropriate in a task, many auditors do not process cues configurally (e.g., Brown and Solomon 1990, 1991; Hooper and Trotman 1996), and many auditors who do process cues configurally do not do it in the appropriate form (e.g., Brown and Solomon 1990), (3) configularity may lead to better judgments (Brown and Solomon 1990, 1991; Hooper and Trotman 1996), and (4) the judges understanding of cue interrelationships may be important for the occurrence of configularity (Hooper and Trotman 1996; Leung and Trotman 2005). This is consistent with findings from judgment analysis research in psychology (see Cooksey 1996, 183-185).

³¹ Task property feedback regards the functional form of the correct policy in the environment (i.e., cue interrelationships), cognitive feedback regards an individuals own judgment policy and combined feedback is about both (Leung and Trotman 2005).

3.5 Further Discussion and Theory Development

The relevance of prior research findings for describing and/or improving auditor judgment behavior in today's audit environment may be debatable. First, the results in prior research on configularity in internal control judgment tasks (i.e., Brown and Solomon 1990) may not be as strong as what seems to be the current understanding. Second, control interrelationships may not translate directly to cue interrelationships, and prior audit research may be incomplete in the studied range of both control and cue interrelationships. Third, prior research has not systematically taken into consideration the importance of the judge's understanding of cue interrelationships (i.e., this might deviate from intended/constructed cue interrelationships). Fourth, the effect of the judgment response scale on the judgment policy has not been considered even though this may have consequences for the functional relationship between cues and criterion and therefore for the appropriateness of judgment policies. Fifth, prior research has been limited to compensatory judgment models while more complex models may be needed to model judgments in today's audit environment. Finally, changes in audit regulation and audit practice may have changed both the judgment task (i.e., the task characteristics and the accompanying appropriate judgment policies) and the behavior of the judge (e.g., increased use of appropriate judgment policies compared to findings in prior research). These issues will be discussed below.

3.5.1 A Closer Look at the Findings of Configularity in Internal Control Judgments

Brown and Solomon (1990) is the only study of how cue interrelationships affect the functional form of the judgment policy in auditors internal control judgments. The results were summarized as "40.5% of the auditor's judgment-models attributed a significant portion of judgment variance to one or both of the expected interactions". Results are referred to in a similar manner in Hooper and Trotman (1996), Trotman (1996, 105), Trotman (2005), and Bonner (2007, 154). It is therefore not unreasonable to believe that the general understanding of the results is that 40.5% of the auditors have either one or both of the predicted interactions in their judgment model, and in the correct form. This is, as will be discussed below, an inappropriate understanding of the results which may lead to an overly optimistic view of auditors' ability to process control cues configurally.

Brown and Solomon (1990) used a six-cue policy capturing experiment where cue interactions were manipulated so that cue D and E were expected to have a compensatory form interaction while cue D and F were expected to have an amplifying form interaction. An appropriate judgment policy would therefore include both of these interactions in their correct form. Results show that one or both of the interactions were statistically significant in 30 of the 74 auditor judgment models (i.e., 40.5%). A closer look at the results reveals the following:

- 19 auditors had only interaction DE (i.e., the compensating control)
- 8 auditors had only interaction DF (i.e., the amplifying control)
- 3 auditors had both interaction DE and DF

A further analysis of the auditors with configural models reveals the following:

- Of the 22 auditors (i.e., 19+3) having interaction DE, 13 had it in the appropriate compensating form while 9 had it in the inappropriate amplifying form.
- Of the 11 auditors (i.e., 8+3) having interaction DF, 9 had it in the appropriate amplifying form while 2 had it in the inappropriate compensating form.
- Reported results do not provide information on whether any of the three auditors that had both interactions actually had none, one, or both of them in the correct form. The interactions from these three auditor models are, however, accounted for above (i.e. there was a total of 33 interactions, including the interactions from the three auditors that had both interaction DE and DF).

In other words, focusing on the appropriateness of the form of interaction:

- Of the 15 compensating interactions, only 13 (i.e., from DE) were appropriate, while 2 (i.e., from DF) were inappropriate.
- Of the 18 amplifying interactions, only 9 (i.e., from DF) were appropriate, while 9 (i.e., from DE) were inappropriate.

The headline findings can therefore be summarized as follows:

- 13 out of 74 auditors (i.e., 17.6%) had the predicted compensating DE interaction in the correct form.
- 9 out of 74 auditors (i.e., 12.2%) had the predicted amplifying DF interaction in the correct form.
- The 3 out of 74 auditors (i.e., 4.1%) that had both interactions are accounted for in the two bullet points above. Based on reported results, it is, however, not possible to discern whether they had any of their interactions in a correct form or not

The results suggest that very few auditors processed the control cues in the appropriate configural manner (17.6% for compensating cues and 12.2% for amplifying cues, with a maximum of 4.1% getting both of them right). Prior evidence on auditor's ability to process control cues configurally in internal control judgments may therefore be weaker than previously believed (i.e., the possible current understanding of 40.5% of the auditor's using an appropriate configural judgment policy for at least one of the control interactions is possibly a misinterpretation of the results in Brown and Solomon 1990). An updated study of configularity in internal control judgments may therefore be more warranted than previously considered.

3.5.2 Control Interrelationships vs Cue Interrelationships

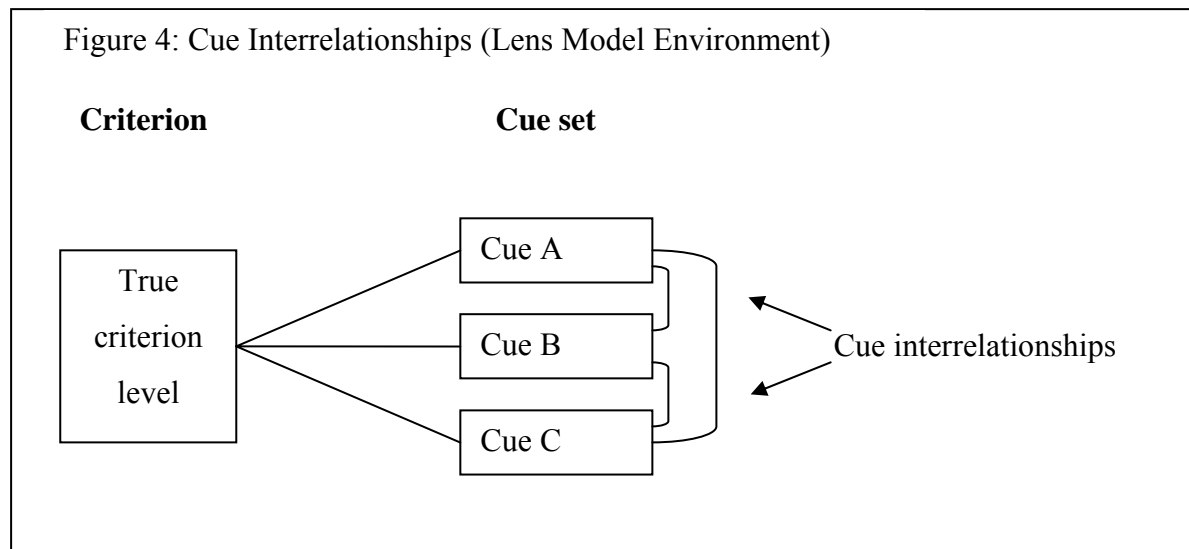
In internal control judgment tasks internal controls serve as cues and the criterion is audit specific (e.g., control risk). Internal controls may be interrelated in many ways (e.g., compensating and amplifying controls), and how they interrelate has been an important task characteristic in audit research (e.g., Brown and Solomon 1990).

At the generic level, cue interrelationships regard how cues interact in their relation to a criterion (Bonner 1994). It may therefore be useful to clarify whether control interrelationships and cue interrelationships are essentially the same thing, or whether, in fact, they are two different constructs. If they are the same thing, and control interrelationships are merely an operationalization of cue interrelationships, then the distinction is less important. If, however, they are two different constructs, then care must be taken when operationalizing cue interrelationships in a control judgment setting, and when generalizing findings from such a setting.

The existence of control interrelationships is clearly recognized in audit regulation. For example, ISA 315.44 and ISA 315.54 (IFAC 2008) require judging controls in combination with other controls when it is appropriate. Audit regulation furthermore recognizes various forms of control interrelationships. For example, PCAOB AS5.76 states that “*The auditor should evaluate the effect of compensating controls*”. PCAOB AS5.12 states that control objectives can be achieved by alternative (i.e., substitutable) internal controls “(...)
implement alternative controls to achieve its control objectives (...)”. Control interrelationships are therefore specific to an audit setting and tied to general concepts in auditing (Brown and Solomon 1990, 21). More generally, audit regulation gives support to the notion that control interrelationships can be defined through how controls interact in their relationship to risk: “*The auditor should focus on whether the selected controls, individually or in combination, sufficiently address the assessed risk of misstatement of a given relevant assertion*”, “(...)
evaluating controls individually or in consideration with other controls” (ISA 315.44 and ISA 315.54, IFAC 2008). In auditing, risk is conceptualized through the audit risk model. It is therefore reasonable to assume that control interrelationship concepts are based on some form of risk scale – typically a percentage point or categorical risk scale (e.g., high, medium, low).³²

Cue interrelationships, however, are more generic in that they can be defined by how cues interact in their relation to a criterion (i.e., the state of an external reality) (see figure 4 below) (Libby 1981, p8, item B3). While control interrelationships are tied to a risk scale, the general construct of cue interrelationships may be tied to any kind of scale.

³² Audit regulation acknowledges that the auditor may use various judgment response scales in control judgments: “*The risks of material misstatement may be expressed in quantitative terms, such as in percentages, or in non-quantitative terms*” (ISA 200.A17 ED, IFAC 2008).



It is therefore of value to know whether control interrelationships and cue interrelationships always represent the same construct. If not, then generalization of findings from internal control judgment studies may be questionable. Similarly, it would be problematic to assume that findings from generic judgment research would automatically be valid for an audit setting. The question is therefore posed: What is the relationship between control interrelationships and cue interrelationships?

Early control judgment research (i.e., prior to 1990) with findings relevant for cue processing did not discuss interrelationships on the environment side of the Lens Model (i.e., it was not an issue since cues were independent) (Brown and Solomon 1990). Brown and Solomon (1990) is therefore the only internal control judgment study that provides an idea about the relationship between control interrelationships and cue interrelationships. Although they did not explicitly discuss the general relationship between control- and cue interrelationships it is clear that they implied a one-to-one relationship; compensating (amplifying) controls is the same thing as compensating (amplifying) cues. In their study the functional form of the judgment policy is predicted based on audit concepts of amplifying and compensating controls (e.g., compensating controls result in compensatory form judgment policies). The functional form of the judgment policy is then observed through the forms of interactions in judgment models, and these judgment models are defined in terms of effects of cues: “*Ordinal relations require that higher (lower) values of an information cue imply higher (lower) judgment values, regardless of the values of other information cues*” (Brown and Solomon 1990, 22). It is therefore reasonable to state that prior research

assumed a one-to-one relationship between control interrelationships and cue interrelationships.

This raises two questions: First, do compensating (amplifying) controls always translate to compensating (amplifying) cues? (2) Is the relationship one-to-one for other potentially important control- and cue- interrelationships (e.g., substitutable controls/cues)?

This dissertation suggests that control interrelationships and cue interrelationships are two different constructs. A sufficient argument for this view is that a change in the judgment response scale may change cue interrelationships even if controls, and thus control interrelationships, remain the same. For example, if each of two controls is related to a different subset of material invoices, then they are independent. Both controls are, however, necessary for a positive binary judgment about overall sufficiency of controls. Although controls are independent in their effect on risk (i.e., invoices), the cues are completely-dependent in the binary judgment setting. Control interrelationships and cue interrelationships must therefore be two different constructs. Note that it is not suggested that controls and cues are two different constructs; controls are operationalizations of cues. However, the interrelationships are two different constructs, and it is the interrelationships that are the independent variables in this dissertation. The mechanism through which the judgment response scale affects cue interrelationships is further developed under the discussion of the impact of the judgment response scale below.

This dissertation therefore proposes that control interrelationships do not always translate directly to cue interrelationships and that, in fact, they are two different concepts. The relationships will be further developed below when the concepts of control interrelationships, cue interrelationships, the judgment response scale and the criterion scale are discussed.

3.5.3 Incomplete Range of Control- and Cue Interrelationships

Prior to 1990, audit research was descriptive and variation in cue interrelationships was not systematically built into experiments. Auditor's therefore appropriately used linear judgment models (Brown and Solomon 1990). Brown and Solomon (1990) introduced some variation in cue interrelationships through compensating and amplifying controls. However, all prior internal control studies have used judgment tasks with controls at the transaction level (e.g.,

payroll, accounts receivable, sales and purchasing) (Trotman and Wood 1991). Today, audit regulation and practice is putting increased emphasis on entity-level controls in order to increase audit efficiency and effectiveness (PCAOB AS5.16; ISA 315, IFAC 2008).³³ An analysis of control interrelationships in some important entity-level controls (e.g., risk management and control monitoring) shows that they can be characterized as multi-step controls where each control step is completely dependent on the previous step in such a manner that the multi-step control is deficient unless each step is effective (i.e., control steps are completely dependent on one another and all control steps are necessary if any risk reduction is to occur at all).³⁴ This represents a fundamentally different control, and cue, interrelationship than what is found in the transaction level controls employed in prior research.

Furthermore, audit research has not systematically studied the potentially important effect of the degree of compensation between controls.³⁵ It was, however, briefly recognized in Brown and Solomon (1990, 22); “(...) control risk would be judged to be as low as (...) assuming full compensation, or would be judged somewhere between (...) assuming only partial compensation”. In Brown and Solomon (1991, 104), the two interacting cues (in experiment one) were intended to be fully substitutable; “(...) either procedure alone can provide sufficient competent evidence (...)”. However, in the next sentence they describe the second cue as providing “little incremental benefit”; the term “little” is, however, more relevant for compensating controls, while the term “no incremental benefit” is appropriate

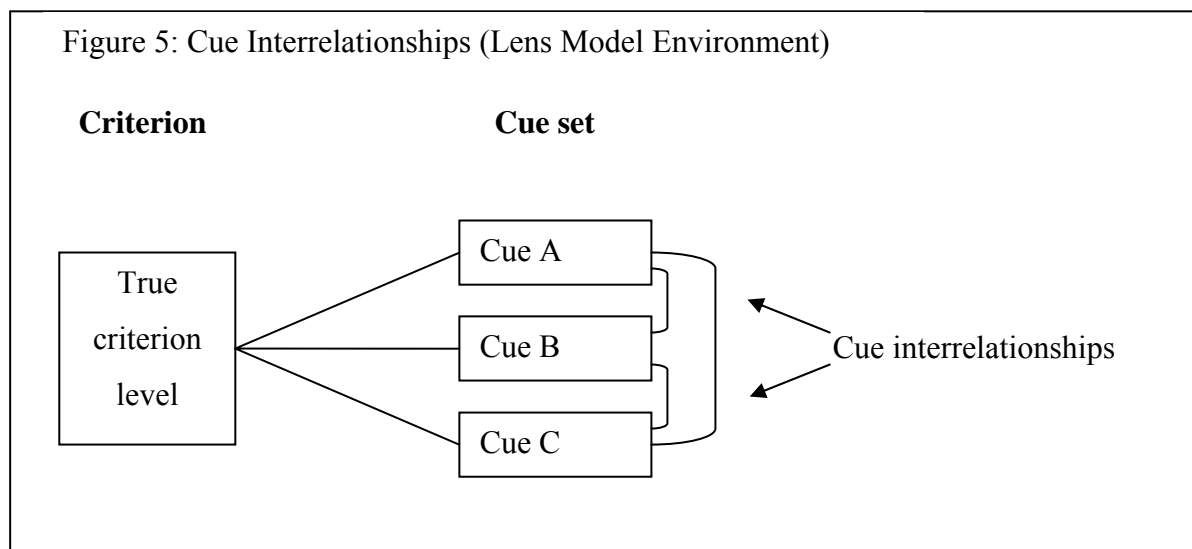
³³ “When using a top-down approach, the auditor identifies the controls to test by starting at the top—the financial statements and company-level controls (...) Following the top-down approach helps the auditor focus the testing on the right controls—those controls that are important to the auditor's conclusion (...)” (PCAOB 2006, 5).

³⁴ The clients’ risk assessment process is an internal control component (COSO 1992). Risk management is a multi step process with several sequential steps (see COSO 2004 p.3-4). PCAOB AS5.18 and ISA 315.43b refer to “the company’s risk assessment process”. AICPA audit guide para 4.45 (2006) and ISA 315.77 detail the steps in the risk assessment process “The auditor when evaluating the client’s risk assessment process should consider how client management: (1) Identifies business risks relevant to financial reporting, (2) Estimates the significance of the risks, (3) Assesses the likelihood of their occurrence, and (4) Decides upon actions to manage them.” ISA 315 appendix 2.5 further describes risk assessment as a process.

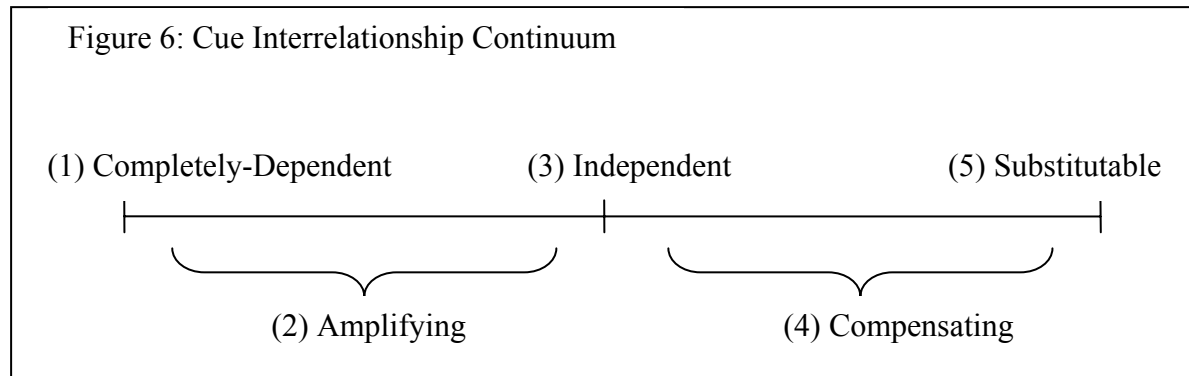
³⁵ Hooper and Trotman (1996, 134) analyzed subject’s reasons for judgments and found that the perceived degree of compensation between controls varied from “none” through “slight” and “some” up to “substitutable” (i.e., complete compensation). Furthermore the main difference between configural versus non-configural judges was that the former considered cues to be interrelated while the latter did not. Thus, degree of cue interrelationships (i.e., compensation) seems to be related to the presence and magnitude of configural judgment policies. However, no systematic contrast of substitutable and partly compensating controls has been done. The relevance of fully compensating/substitutable/alternative controls is also recognized by the SEC (2007, 25); “(...) when more than one control exists that individually addresses a particular risk (i.e., redundant controls) (...)”.

for fully substitutable controls. Furthermore, the hypothesized judgment models in Brown and Solomon (1990 and 1991) (i.e., compensatory form ordinal) are relevant for partly compensating controls, and not for fully substitutable controls which, as will be discussed below, require disjunctive judgment policies. The degree of compensation between controls may therefore be important for the functional form of the judgment policy.

Cue interrelationships are defined through how cues interact in their relation to the criterion (Libby 1981, p8, item B3). This can be illustrated by a figure of the environment side of the Lens Model, consisting of the underlying reality/criterion and the cue set (see figure 5 below). The solid lines illustrate the independent and interrelated relationships between the cues and the criterion.



Cues may interact in many forms. A framework is suggested where the range of cue interrelationships be described along a continuum (see figure 6 below). The framework assumes a continuous criterion scale of some kind (e.g., control risk on 100-point percentage scale) and binary cue levels (e.g., cues are absent/present). The continuum has five sections. Although other cue interrelationships may exist, these are not developed here.



The continuum is developed for binary cues since: (1) prior research on internal control judgments has used tasks with binary cue levels (e.g., Ashton 1974; Brown and Solomon 1990, 1991), and (2) binary cues are similar to the audit ecology where a given internal control is either present or not. In general, when cue levels are binary, they are typically either above or below a judgment relevant criterion (e.g., an individual control is either present/absent or deficient/effective). The figure can be read as follows; independent cues are at the neutral center (i.e., no interaction). To the left, cues interact positively and to the right cues interact negatively in their effect on the criterion. At the extremes of the line, interaction is complete. Between the center and the extremes, interactions are partial.

The cue interrelationship continuum can furthermore be illustrated mathematically by use of functions. Assume that “y” represents the criterion level and that “y” is a function of cues “c_i”:

Let $y = f(c_i)$, where there are two cues ($i=1,2$) and cues take on the values 0 or 1 ($c_i = 0,1$)

Assume

$$0 = 0/100 \leq f(c_1, c_2) \leq 100/100 = 1$$

and

$$f(0,0) = 0 \text{ for all functions}$$

This provides the following cue interrelationship functions (see table 1 below):

Table 1: Mathematical Representation of Cue Interrelationship Functions

Cue interrelationship	Mathematical Function	Description
1. Completely-dependent	$f(1,0) = f(0,1) = 0$, and $f(1,1) > 0$	Complete positive interrelationship
2. Amplifying	$0 < f(0,1) + f(1,0) < f(1,1)$ $0 < f(1,0), 0 < f(0,1)$	Partial positive interrelationship
3. Independent	$0 < f(0,1) + f(1,0) = f(1,1)$ $0 < f(1,0), 0 < f(0,1)$	No interrelationship
4. Compensating	$0 < f(1,1) < f(0,1) + f(1,0)$ $0 < f(1,0), 0 < f(0,1)$	Partial negative interrelationship
5. Substitutable	$0 < f(1,0) = f(0,1) = f(1,1)$	Complete negative interrelationship

An intuitive interpretation of the mathematical functions and the continuum is provided through an audit example (see below). In this example, the criterion is control risk reduction on a continuous scale and the cues represent controls. The relationship between control risk and controls is such that all controls may have an effect on control risk (i.e., no controls are trivial). This effect can be independent (i.e., for each control) or interrelated (i.e., controls have combined effects).

1. Completely-dependent control cues: Such individual control cues only have an effect on control risk if all other control cues are present: $f(1,1) > 0$, but $f(1,0) = f(0,1) = 0$. In this setting $f(1,1) > 0$ can be interpreted as control risk reduction being bigger than 0%: control risk is thus $< 100\%$. $f(1,0) = f(0,1) = 0$ can be interpreted as no control risk reduction occurring: control risk is thus $= 100\%$. Such control cues do not have main (i.e., individual) effects; they only have a combined effect (i.e., an interaction effect). This is a corner solution that represents a complete positive interrelationship.

A typical example of completely dependent control is a multi-step risk assessment process. Unless all steps are present (i.e., risk identification, assessment of likelihood and impact, and appropriate mitigating action), the process as a whole fails.

2. Amplifying control cues. Such control cues have both individual and interactive effects on control risk. Individual effects imply that $0 < f(1,0)$ and $0 < f(0,1)$. An interactive effect of amplifying form implies that the effect of two control cues combined is larger than the sum of the two individual cue effects: $0 < f(0,1) + f(1,0) < f(1,1)$. For example, control “c₁” working alone (i.e., without control “c₂”) may reduce control risk by 20%. Similarly control “c₂” working alone (i.e., without control “c₁”) may reduce control risk by 25%. However, if both controls are present, control risk may, for example, be reduced by 60%. The combined effect of the two controls is therefore larger than the sum of their individual effects; $60\% > 20\% + 25\% = 45\%$.
3. Independent control cues. Such control cues have positive individual effects, but no interactive effects on control risk. Individual effects imply that $0 < f(1,0)$ and $0 < f(0,1)$. The lack of interactive effects imply that the combined effect of the controls is equal to the sum of their individual effects: $0 < f(0,1) + f(1,0) = f(1,1)$. The effect of a control cue on control risk is therefore independent of the presence/absence of other control cues. For example, control “c₁” working alone (i.e., without control “c₂”) may reduce control risk by 20%. Similarly control “c₂” working alone (i.e., without control “c₁”) may reduce control risk by 25%. However, if both controls are present, control risk is reduced by 45%. The combined effect of the two controls is therefore equal to the sum of their individual effects: $45\% = 20\% + 25\% = 45\%$.
4. Compensating control cues. Such control cues have both individual and interactive effects on control risk. Individual effects imply that $0 < f(1,0)$ and $0 < f(0,1)$. An interactive effect of compensating form implies that the effect of two control cues combined is smaller than the sum of the two individual cue effects: $0 < f(1,1) < f(0,1) + f(1,0)$. For example, control “c₁” working alone (i.e., without control “c₂”) may reduce control risk by 20%. Similarly control “c₂” working alone (i.e., without control “c₁”) may reduce control risk by 25%. However, if both controls are present, control risk may, for example, be reduced by 30%. The combined effect of the two

controls is therefore smaller than the sum of their individual effects; $30\% < 20\% + 25\% = 45\%$. Note that adding a second control has a positive effect as control risk is lower with two controls combined (30% reduction) than with any control working alone (20% or 25% reduction). Although the interaction effect is negative, the total effect is positive when a second control is added. This is due to the positive individual effects of each control. Compensating controls have individual but partly overlapping contribution to control risk reduction

5. Substitutable control cues. Such control cues have both individual and interactive effects. Furthermore, the individual effects are identical: $0 < f(1,0) = f(0,1)$. The interactive effects is such that the effect of two control cues combined is the same as the individual contribution of one control cue (i.e., cues can substitute for each other, but they do not add incremental effect if another cue is already present): $0 < f(1,0) = f(0,1) = f(1,1)$. For example, control “c₁” working alone (i.e., without control “c₂”) may reduce control risk by 20%. Similarly control “c₂” working alone (i.e., without control “c₁”) may reduce control risk by 20%. However, if both controls are present, control risk is also reduced by 20%. The combined effect of the two controls is therefore equal to one of the individual effects (which are equal). Substitutable controls therefore have individual but completely overlapping contribution to control risk reduction.

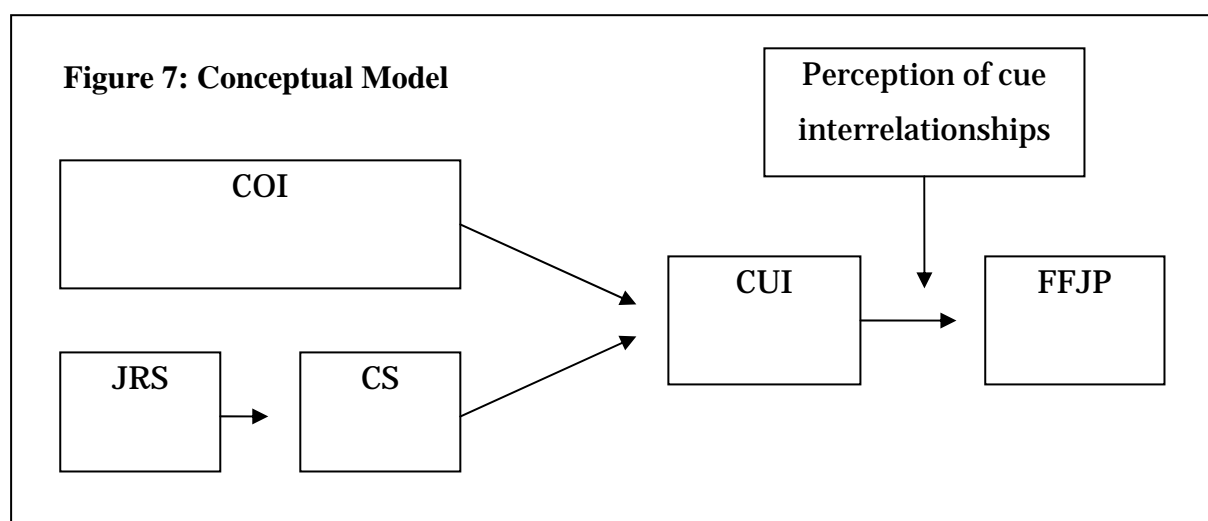
The relevance of such a control interrelationship continuum is supported in audit regulation: ISA 315.44 and ISA 315.54 require judging controls in combination with other controls when it is appropriate “*The auditor should focus on whether the selected controls, individually or in combination, sufficiently address the assessed risk of misstatement of a given relevant assertion*”, “*(...) evaluating controls individually or in consideration with other controls*”. PCAOB AS5.A9 states that “*Effective internal control over financial reporting often includes a combination of preventive and detective controls*”. PCAOB AS5.76 states that “*The auditor should evaluate the effect of compensating controls*”. PCAOB AS5.12 states that control objectives can be achieved by alternative internal controls “*(...) implement alternative controls to achieve its control objectives (...)*”. PCAOB AS5.47 states that control strength depends on “*the degree to which the control relies on the effectiveness of other controls (...)*”. Regulatory guidance therefore has a clear concept of

controls being individual or interacting with other controls, and that control interactions can be alternative (i.e., substitutable), compensating and amplifying, in addition to multi-step.

This dissertation proposes that the continuum is relevant for describing the range of both control interrelationships and cue interrelationships (i.e., they can be described along the same scale). Both controls and cues can interact positively or negatively in their contribution to control risk, and interactions can be partial or complete. Note, however, that when a control serves as a cue, control interrelationships and cue interrelationships may, or may not, be the same (i.e., they are assessed on the same scale/continuum, but may have different values on the scale/continuum).

3.5.4 Cue Interrelationships May Not Be Well Understood by Participants

Prior studies have not directly measured if participants perceive cues to be interrelated in the manner intended by the researcher. For example, Brown and Solomon (1990) manipulated two of the cues to be compensating and then observed the judgments made by participants. However, if the judgment model of an auditor did not include the relevant interaction it could not be known whether this was due to the auditor misunderstanding the cue interrelationships (e.g., understanding the cues as independent), or whether it was due to an inability to process cues configurally given a correct understanding of cue interrelationships. Figure 7 (below) incorporates the judge's perception of cue interrelationships into the conceptual model of this dissertation.



Hooper and Trotman (1996), however, measured this indirectly. They asked participants to give reasons for judgments, and post hoc analysis revealed that; (1) understanding of cue interrelationships varied; the perceived degree of compensation between controls ranged from “none” through “slight” and “some” up to “substitutable” (i.e., complete substitution), and (2), overall, subjects that understood cues as interrelated used configural judgment models, while subjects that understood cues as independent used linear judgment models.

The consequence of varying perceptions of cue interrelationships becomes clearer when one considers the structure of a policy capturing study: Hypotheses predict how actual (i.e., apriori designed or intended) cue interrelationships should affect judgment models. Hypotheses are then tested by comparing predictions for a given cue interrelationship treatment with regression model parameters from cases that correspond to that specific cue interrelationship treatment. The risk judgments for a given cue interrelationship treatment are, however, based on how participants perceive cue interrelationships. For example; hypotheses about the effect of compensating cues on the judgment model is tested against regression model parameters obtained from the cases that were designed to include compensating cues. If participants perceive these cues to be interrelated in another manner, for example amplifying, then the treatment of this condition fails. Participant’s lack of understanding of cue interrelationships may therefore be one potential explanation for the low levels of configularity in Brown and Solomon (1990).

When the purpose of the study is to examine the effect of cue interrelationships on the functional form of the judgment model, measures should therefore be taken to increase the likelihood that participants perceive cue interrelationships as intended. Furthermore, it may be reasonable to omit from further analysis participants reporting deviating cue interrelationship perceptions since it is more likely that they did not absorb the treatment. However, since self reporting may be noisy, and since no measurement instrument has been developed and tested in prior research, automatically omitting them can not be recommended. Omitting them should therefore be based on a thorough study of their judgment responses, looking for systematic response patterns indicating deviating cue interrelationship perceptions. The question of how well auditors perceive cue interrelationships is a separate research question that does not require a policy capturing approach. It is, however, clear that more research is needed on this issue, and that a

measurement instrument for cue interrelationship perception would be useful for future research on configularity.

A study where participant understanding of cue interrelationships is both facilitated and directly measured may therefore be warranted (e.g., as in the current study by having subjects explicitly classify cue interrelationships before making judgments).³⁶ This would provide evidence on whether judges apply appropriate judgment models conditional on understanding cue interrelationships (i.e., separate from the question of whether judges actually understand cue interrelationships). Such knowledge may be important for deciding where to direct judgment improvement initiatives such as feedback: If judges have problems understanding cue interrelationships, then decision aid in the form of task property feedback may be most appropriate, but if judges have problems using appropriate judgment policies (i.e., given that they understand cue interrelationships), then decision aid in the form of cognitive feedback may be most appropriate (see Leung and Trotman 2005). It may therefore be appropriate and useful to split the question of the impact of task characteristics on judgment policies into two components; understanding cue interrelationships vs. applying correct judgment policies.³⁷

3.5.5 The Importance of the Judgment Response Scale

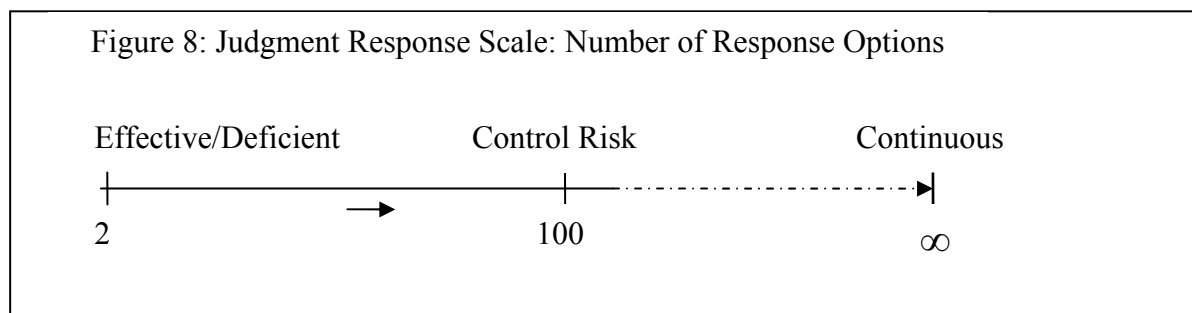
Audit regulation acknowledges that the auditor may use various judgment response scales in control judgments: “*The risks of material misstatement may be expressed in quantitative terms, such as in percentages, or in non-quantitative terms*” (ISA 200.A17 ED, IFAC 2008). This dissertation defines the judgment response scale as the nature of the auditors’ judgment response. Consider, for example, a percentage point risk judgment; the nature of the scale could, for example, be (1) the risk level or (2) the degree of risk reduction. This may seem like the same thing, but a closer look reveals that the scales have two different natures. A control system with low risk should get a low judgment if the nature of the scale is the “risk

³⁶ Direct measurement of participant understanding of cue interrelationships creates the need for a framework for classifying cue interrelationships. Such a framework is proposed in this study.

³⁷ This would be one potential answer to the call made by Leung and Trotman (1996) and Brown and Solomon (1990 and 1991) for research on (a) what factors increase the likelihood of configularity (answer: understanding cue interrelationships), and (b) why some auditors process cues configurally while others do not process cues configurally (answer: because they understand interrelationships or not).

level”, but a high score if the nature of the scale is “risk reduction”. Although the response to the question reveals the same information, the percentage scores are different, and the underlying cue combination and weighting is different. The nature of the scale could therefore have an impact on the judgment, and consequently on the judgment policy.

The range of the scale is the number of judgment response options the auditor has available for a given judgment task. In the simplest form, a judgment response scale is binary (e.g., yes/no, effective/deficient, acceptable/unacceptable). As more response options become available, the scale approaches continuity (see figure 8). In auditing, response scales typically approach continuity in the form of percentage point risk judgments (e.g., the audit risk model; inherent risk, control risk, detection risk).



Audit research on internal control judgments has not applied binary response scales in the past. This is likely due to the fact that the judgment task studied essentially involved continuous risk assessments for audit planning purposes. Today, however, internal control judgments are also important for reporting purposes. Reporting on internal control over financial reporting, for example, requires a binary judgment: “a control deficiency exists when the design or operation of a control does not allow management or employees, in the normal course of performing their assigned functions, to prevent or detect misstatements on a timely basis” (PCAOB AS5.A3). ISA 315.54 has a similar requirement: “Obtaining an understanding of internal control involves evaluating the design of a control and determining whether it has been implemented (i.e., or not).”³⁸ These regulatory requirements and definitions show that there is no degree regarding the existence of a deficiency (although

³⁸ ISA 330 has a similar requirement regarding judging operational effectiveness of internal controls over financial reporting.

there are degrees regarding their importance).³⁹ Judging if a control deficiency exists (i.e., in design or operation) is therefore a binary judgment; the control is effective or deficient.

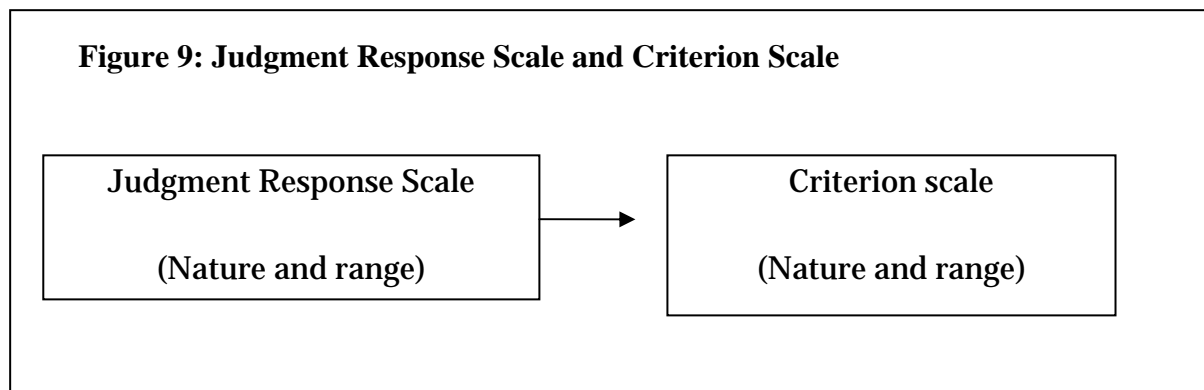
Audit research does not provide theory on cue combination in binary judgments even though such judgments are important for audit practice.⁴⁰ Psychology research has, however, recognized that the judgment response scale is important for theory because the appropriate form of the judgment policy may depend on it (Elrod et al. 2004). For example, Elrod et al. (2004, 5-6) developed different judgment models for varying forms of observed dependent variables (i.e., the dependent variable was binary or ordinal evaluation). Of interest for this dissertation are the two additive noncompensatory models developed for binary accept/reject judgments based on conjunctive or disjunctive judgment policies. Both of these binary models may be relevant for binary auditor judgments, and they are different from the additive compensatory models used in prior audit research to study judgments on continuous judgment response scales. How then does the judgment response scale generally affect the appropriateness of judgment policies in auditing?

First of all it is important to recognize that control interrelationships do not always translate directly into cue interrelationships. Control interrelationships are defined through how controls interact in their relation to risk, while cue interrelationships are defined through how cues interact in their relation to a criterion. This study assumes that when the judgment response scale changes, the range of the criterion changes (see figure 9 below). For example, when the judgment response scale is continuous (e.g., control risk), the range of the criterion is continuous. When the judgment response scale is binary (e.g., acceptable/unacceptable overall control), the range of the criterion is binary. In general, it is assumed that the scale for the judgment response determines the scale for the criterion in such a manner that the criterion scale mirrors the judgment response scale. One potential explanation for why this happens is that the judge reframes his perception of reality (i.e., the criterion) to fit to the judgment response scale. For example, when assessing control risk, the auditor frames the

³⁹ The judgment of whether a control deficiency exists is a separate judgment that is performed independently of, and prior to, the materiality/significance judgment. The materiality/significance judgment is an important audit judgment task, but it is not the focus of this study.

⁴⁰ To the best of my knowledge, no cue combination studies have applied binary internal control judgment tasks.

true risk level in terms of percentage points, but when performing an overall assessment of whether controls are acceptable, the reality is framed as acceptable or unacceptable.



Mathematically, the model above can be formulated as:

$$CS = h(JRS),$$

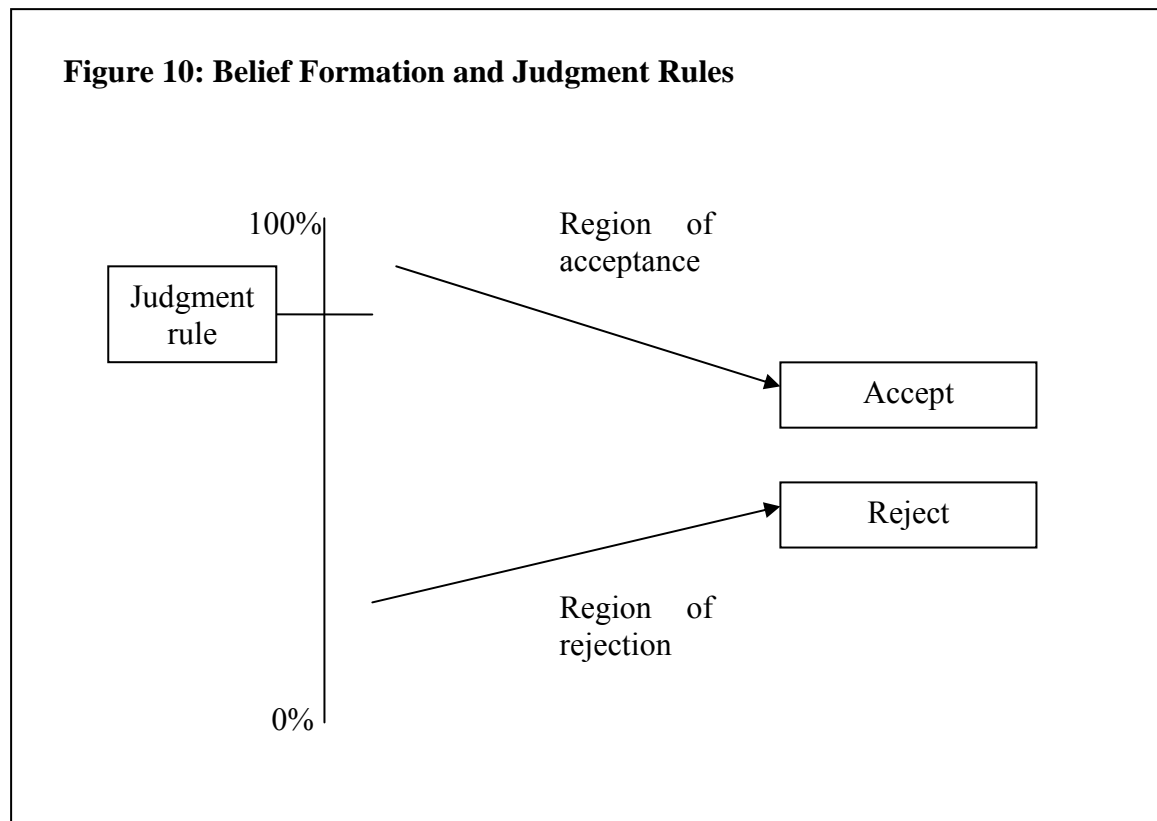
Where:

JRS = Judgment Response Scale

CS = Criterion Scale

It may be argued that a change from a continuous scale to a binary scale is simply the introduction of a cutoff point on continuous scale, and not a change in the scale. This argument may imply that the criterion scale does not change, but that the judgment rule changes: The judge forms his belief about the criterion on a continuous scale, and the belief is mapped onto a binary judgment rule (see figure 10 below).⁴¹

⁴¹ Figure 10 is based on a figure and comments provided by a discussant of the study (Robert W. Knechel) at the European Audit Research Network (EARNET) 2007 conference in Aarhus, Denmark.



The belief formulation process thus involves integrating cues/controls and therefore may exhibit the configural information processing strategies described in this dissertation, but the judgment depends on the relationship between the belief and the judgment rule. In this case, this would imply potentially complex cue processing in order to form a belief about a risk percentage, but a simple accept/reject judgment about that risk percentage.

There are reasons to believe that this might not be the case. Such a judgment process would be unnecessarily demanding: *“Most would agree that determining the proper trade-offs in a compensatory model is the most difficult activity in decision making”* (Libby, 1981, 44). Conjunctive and disjunctive judgment strategies are less demanding cognitively (Elrod et al. 2004). The general finding from the heuristics and biases literature also supports the notion that linear judgment is cognitively and emotionally demanding, and that judges in many circumstances have difficulty with the trade-off involved in compensatory cue weighting (Elrod et al. 2004; Hogarth and Karelaia 2007). It is therefore hard to believe that auditors generally perform cognitively demanding cue processing, by integrating all cues, when it is not necessary.

In binary judgments cues are either sufficient or necessary. A binary judgment can therefore be made by assessing the presence of one sufficient cue or the absence of one necessary cue (i.e., and not assessing all other cues). The judgment processes in binary judgments may therefore be simpler and less demanding for the judge if the belief is formed directly on the binary scale and not on a continuous scale. For example, in a control judgment case with five control cues, the auditor would need to integrate all five cues in order to make a judgment about a control risk percentage (i.e., a continuous scale). However, for an accept/reject judgment the auditor could reach a reject judgment by identifying the absence of one necessary cue, or an accept judgment by identifying the presence of one sufficient cue. Although it cannot be ruled out that the auditor first performs a more demanding cue integration process and then maps this to a cutoff rule, it is not unreasonable to assume that the auditor uses the least demanding judgment strategy (i.e., a judgment based on a binary criterion scale).⁴²

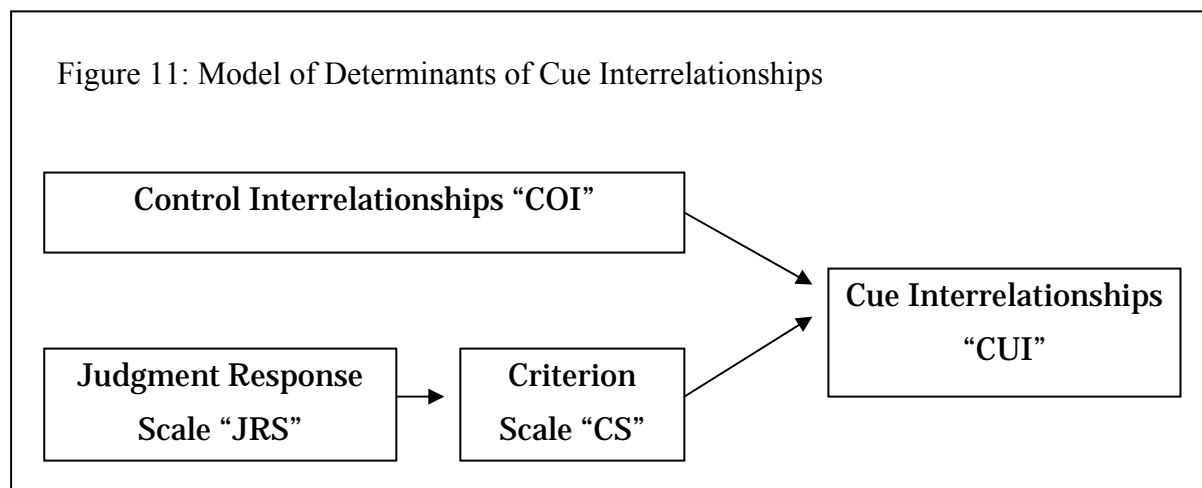
A middle road may also exist. Assume that the auditor, before assessing controls, has an explicit control risk cutoff; e.g., if control risk is above 10% controls are judged to be insufficient. A rational judgment process would then assess cue interrelationships and choose one of two strategies: (1) If control interrelationships indicate that individually sufficient controls for control risk being below 10% exist, find at least one of these being effective, or (2) if control interrelationships indicate that individually necessary controls for control risk being below 10% exist, find at least one of these being deficient. In essence, this judgment process applies a disjunctive judgment policy (i.e., in 1) by finding one individually sufficient control being effective, or a conjunctive policy (i.e., in 2) by finding at least one necessary control being deficient. Furthermore, the criterion scale is, in essence, binary since the criterion is either below or above the chosen cutoff. The change in the judgment response scale thus changes the criterion scale; even though the nature of the scale is still “control risk”, the range of the scale is now binary. Finally, such a judgment policy could be used without assessing all cues. The judge would not need to make a judgment

⁴² In this study’s post experimental questionnaire, participants are asked to explain their thought process for the binary judgment. Responses revealed that 67% of the participants made the binary judgment directly while 33% based the binary judgment on the percentage risk score. In this study the judgment response scale was manipulated within subjects. The percentage risk score was therefore available to participants when making the binary judgment. If this had not been available, it is not unreasonable to believe that more than 67% would have made the binary judgment directly since this would have been less demanding cognitively. However, this is an open research question.

about a specific control risk for the process. In general, it seems reasonable that the judge would not continue cognitively demanding cue processing, by processing additional cues, after it is clear what the judgment would be.

A change in the judgment response scale is therefore assumed to cause a change in the criterion scale and not just impose a cutoff judgment rule on an unchanged scale. Policy capturing does, unfortunately, not provide tools for resolving this issue, since, in essence, the methodology correlates input (i.e., cues) and output (i.e., judgments) without observing what goes on within the mind of the judge.

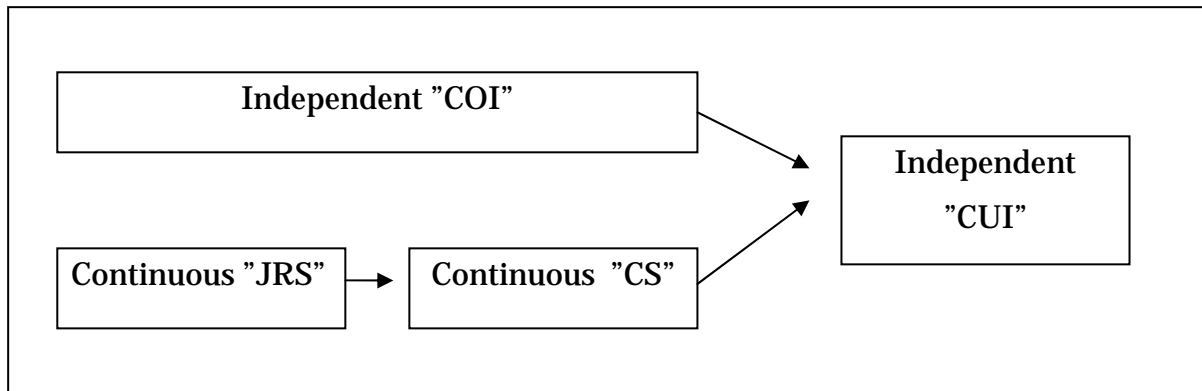
Assuming that a change in the judgment responses scale causes a change in the criterion scale (i.e., $CS = h(JRS)$), the functional form of the cues relationship to the criterion may also change. In other words, cue interrelationships are determined by control interrelationships and the judgment response scale (i.e., $CUI = g[COI, CS] = g[COI, h(JRS)]$) (see figure 11 below):



For example; when controls are independent and the judgment response scale is continuous the criterion is continuous and cues are independent (i.e., interrelated in a linear manner) (see figure 12 below):

$$\begin{aligned} \text{Independent CUI} &= g[\text{Independent COI, Continuous CS}] \\ &= g[\text{Independent COI, } h(\text{Continuous JRS})] \end{aligned}$$

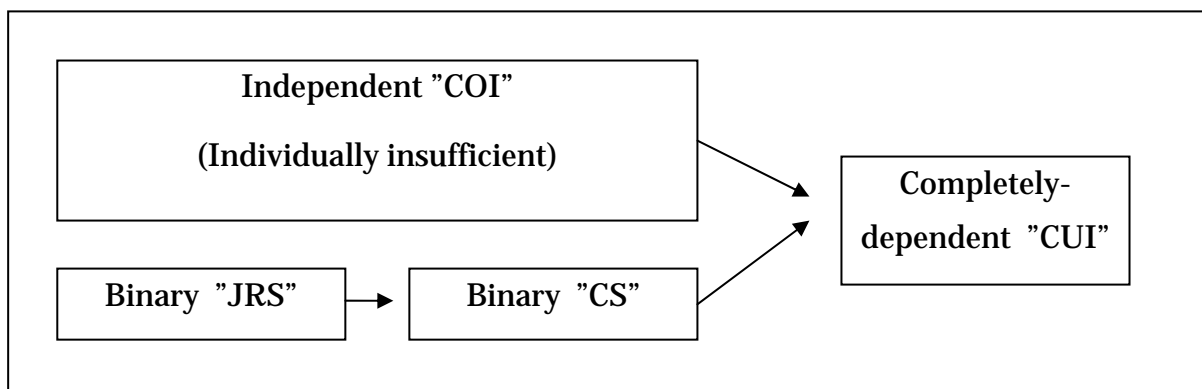
Figure 12: Independent cue interrelationships



If the judgment response scale (“JRS”) becomes binary (i.e., a change in task characteristics), the criterion (“CS”) also becomes binary. The relationship of the same cues to the criterion (“CUI”) however changes; if two independent control cues do not individually provide sufficient risk reduction, but together provide sufficient risk reduction, then both controls need to be effective if the binary judgment is to be positive (i.e., that controls are overall effective). The cue interrelationship in the binary scale judgment would then be that cues are completely dependent (note that the exact same cues were independent in a continuous scale judgment) (see figure 13 below).

$$\begin{aligned} \text{Completely-dependent CUI} &= g[\text{Independent COI, Binary CS}] \\ &= g[\text{Independent COI, } h(\text{Binary JRS})] \end{aligned}$$

Figure 13: Completely-Dependent Cue Interrelationships



In general, when the response scale is continuous, control interrelationships (“COI”) translate directly into cue interrelationships (“CUI”):

Multi-step controls and other kinds of completely-dependent controls translate to completely-dependent cues:

Completely-dependent CUI = g[Completely-dependent COI, Continuous CS]
 = g[Completely-dependent COI, h(Continuous JRS)]

Amplifying controls translate to amplifying cues:

Amplifying CUI = g[Amplifying COI, Continuous CS]
 = g[Amplifying COI, h(Continuous JRS)]

Independent controls translate to independent cues:

Independent CUI = g[Independent COI, Continuous CS]
 = g[Independent COI, h(Continuous JRS)]

Compensating controls translate to compensating cues:

Compensating CUI = g[Compensating COI, Continuous CS]
 = g[Compensating COI, h(Continuous JRS)]

Substitutable controls translate to substitutable cues:

Substitutable CUI = g[Substitutable COI, Continuous CS]
 = g[Substitutable COI, h(Continuous JRS)]

However, independent, amplifying and compensating cues do not, by definition, exist in binary judgment tasks; cues that are independent, compensating or amplifying in continuous judgment tasks are either (1) individually sufficient in binary judgment tasks (and thus translate to substitutable cues), or (2) not individually sufficient in binary judgment tasks (and thus translate to completely-dependent cues). This can also be understood as sufficient (i.e., substitutable) or necessary (i.e., completely dependent) conditions for binary judgment outcomes.

In general, when the response scale is binary, control interrelationships (“COI”) do not translate directly into cue interrelationships (“CUI”):

Multi-step controls and other kinds of completely-dependent controls translate to completely-dependent cues (i.e., not individually sufficient):

$$\begin{aligned} \text{Completely-dependent CUI} &= g[\text{Completely-dependent COI, Binary CS}] \\ &= g[\text{Completely-dependent COI, } h(\text{Binary JRS})] \end{aligned}$$

Amplifying “A”, independent “I” or compensating “C” controls translate to completely-dependent cues (i.e., not individually sufficient) or to substitutable cues (i.e., individually sufficient):

$$\begin{aligned} \text{Completely-dependent CUI} &= g[\text{I, A or C COI, Binary CS}] \\ &= g[\text{I, A or C COI, } h(\text{Binary JRS})] \end{aligned}$$

or

$$\begin{aligned} \text{Substitutable CUI} &= g[\text{I, A or C COI COI, Binary CS}] \\ &= g[\text{I, A or C COI COI, } h(\text{Binary JRS})] \end{aligned}$$

Substitutable controls translate to substitutable cues (i.e., individually sufficient):

$$\begin{aligned} \text{Substitutable CUI} &= g[\text{Substitutable COI, Binary CS}] \\ &= g[\text{Substitutable COI, } h(\text{Binary JRS})] \end{aligned}$$

3.5.6 Judgment Policies and Judgment Models

Several psychology researchers have argued for the importance of judgment task characteristics for the form of the judgment policy, and for the importance of developing mathematical representations (i.e., models) of judgment policies (e.g., Einhorn 1970 and 1971; Brehmer 1994; Stewart et al. 1997; Elrod et al. 2004). A judgment policy is defined as the way in which cues are combined by the judge when making a judgment. A judgment model is a mathematical representation of the judgment policy. Valid mathematical models

of judgment policies are argued to be important because they provide a precise specification of theory (Elrod et al. 2004). A wide range of models have thus been developed.

Audit research has been limited to studying judgment tasks with independent, compensating or amplifying controls and searching for compensatory, linear judgment models with amplifying or compensatory form ordinal interactions (e.g., Brown and Solomon 1990 and 1991). Non-compensatory judgment policies have not been studied in internal control research, although they are known to audit literature (e.g., Libby 1981, 46).

This study proposes that the following models may be relevant for audit research. Assume that “y” represents the judgment and that “y” is a function of cues “c_i”:

Let $y = f(c_1, c_2)$, where $c_i = 0, 1$ and $i=2$,

and

$$0 = 0/100 < f(c_1, c_2) < 100/100 = 1$$

This provides the following judgment models (see table 2 below):

Table 2: Mathematical Representation of Judgment Models

Model	Mathematical Form	Main effects and interactions
1. Conjunctive	$f(1,0) = f(0,1) = 0$, and $f(1,1) > 0$	No main effects Complete positive interaction
2. Amplifying	$0 < f(0,1) + f(1,0) < f(1,1)$	Main effects and amplifying form ordinal interaction
3. Linear	$0 < f(0,1) + f(1,0) = f(1,1)$	Only main effects No interactions
4. Compensatory	$0 < f(1,1) < f(0,1) + f(1,0)$	Main effects and compensatory form ordinal interaction
5. Disjunctive	$0 < f(1,0) = f(0,1) = f(1,1)$	Main effects and interactions that nullify all but one of the (equal) main effects

The mathematical models of judgment policies are similar to the mathematical functions of cue integration in judgment policies presented in section 3.5.3. The reason for this is that this dissertation proposes that the judge perceives the cue interrelationships in the environment and applies a form of cue integration in his judgment policy that mirrors these cue interrelationships in the environment. Although the math is parallel, it represents different constructs. In the environment, the model represents the relationship between cues and the criterion (e.g., the effect of controls on risk). In the judgment policy, the model represents the weight the judge places on the various cues and their interactions (i.e., the effects of the cues on the judgment). The models are discussed below:

Additive, compensatory, linear models (non configural)

The linear compensatory model is the simplest and most used model in judgment research. It has three main features: (1) it is additive in its attributes, which means that the judgment is obtained by simply summing the assessments of each cue considered individually, (2) it is compensatory, which implies that the judgment, based on any cue, may be offset by considering one or more of the other cues, and (3) it is linear, which means that all cues relate in a linear manner to the judgments (Elrod et al. 2004).

Additive, compensatory, nonlinear models (configural)

If the third feature is relaxed, the mathematical model can still remain additive and compensatory, but it is no longer linear. This is the case if one or more cues interact, or if the form of a cues' relationship to the judgment criterion is of a quadratic, cubic or higher order polynomial form. The model can be represented as additive by forming product terms of the interacting cues. Even though interactions are formed by a multiplicative rather than additive organizing principle, the overall policy model is still additive and can be analyzed by ordinary regression procedures.⁴³ The two compensatory, additive nonlinear models most relevant to audit research are compensatory form ordinal and amplifying form ordinal:

Compensatory form ordinal models represent cues that contribute independently and interact in such a manner that the combined effect of both cues is smaller than the sum of their individual effects. In such models the main effects will be significant and interactions for compensating cues will be negative (i.e., ordinal form compensatory).

Amplifying form ordinal models represent cues that contribute independently and interact in such a manner that the combined effect of both cues is larger than the sum of their individual effects. In such models the main effects will be significant and interactions for amplifying cues will be positive (i.e., ordinal form amplifying).

Nonadditive, noncompensatory, nonlinear models (configural)

If the second feature is also relaxed, the model becomes noncompensatory. In a noncompensatory judgment policy, the judgment may be determined by the level of only one

⁴³ The model remains additive by adding polynomial and interaction terms to the mathematical model.

of the cues, irrespective of the level of other cues. The two types of noncompensatory judgment policies that have received the greatest attention are conjunctive and disjunctive (Einhorn 1970, 1971; Libby 1981, 46).

In a conjunctive judgment policy it is necessary that each cue be above a criterion level (i.e., all cues are necessary), thus the judgment may be based solely on the lowest/worst cue (i.e., if it is below the criterion level (e.g., absent)).⁴⁴ It is a form of multiple cutoff judgment, where each cue is required to exceed a minimum level. An example of this would be a control process that is judged to have deficient controls unless each step in the process is above a certain cutoff criterion. This is equivalent to judging based on the worst cue: it does not help if all other cues are above the criterion as long as the worst cue is under. Conjunctive models seem intuitively appropriate for binary judgments where all cues are necessary criteria for a given judgment; if one of the cues fail, the judgment is negative. Furthermore, conjunctive judgments require configularity.

Conjunctive models that have binary cues (e.g., 0/1 or no/yes) can be represented mathematically by a model with no main effects and only the highest order interaction term. In such a model, the interaction term would be zero unless all cues are positive (i.e., 1 or yes), thus it would be judged on the worst cue.

In a disjunctive judgment policy it is sufficient if at least one cue is above a criterion level (i.e., each cue is sufficient), thus the judgment can be based solely on the highest/best cue (i.e., if it is above the criterion level).⁴⁵ An example of this would be a judgment of control risk being below a certain acceptability level and where several alternative controls are assessed. As long as one of the controls reduces control risk sufficiently (e.g., the best control), the other controls do not matter (i.e., they cannot change the judgment). Disjunctive models seem intuitively appropriate for binary judgments where all cues are individually

⁴⁴ An example of a high level conjunctive judgment task is the overall SOX 404 judgment (AS5.2 PCAOB 2007): “(...) because a company’s internal control cannot be considered effective if one or more material weaknesses exist (...)”. Thus, if one or more material weaknesses exist, the other effective controls contribute nothing to the overall judgment.

⁴⁵ ISA 330.70 (IFAC 2008) provides an example of a judgment criterion; “The auditor should conclude whether sufficient appropriate audit evidence has been obtained to reduce to an acceptably low level the risk of material misstatement in the financial statements”. The sufficient and appropriate evidence requirement is an example of a criterion where the audit evidence cues (e.g., control test results) may be assessed according to conjunctive or disjunctive judgment policies. See further discussion in the hypotheses development.

sufficient criteria for a given judgment; if one of the cues is acceptable, the judgment is positive. Disjunctive judgments require configural cue processing.

Disjunctive 2-cue models with binary cues (e.g, 0;1 or no;yes cues) can be represented by a model where an interaction term is formed in addition to the main effects. The purpose of the interaction term would be to remove one of the main effects if both cues are positive, so that only one main effect remains. The model would then represent a judgment where only one of the cues impacts the judgment, even though they are both positive. The cue weights would be expected to be equal in size, but positive for main effects and negative for the interaction. The model can be generalized to judgments with more than 2 cues by adding interaction effects that eliminate the effects of having more than one positive cue.

It can be noted that the use of a conjunctive versus a disjunctive model may depend on the framing of the judgment criterion: If the judgment criterion is framed as “is there a material weakness in internal controls”, then it is sufficient to find one material weakness and a conjunctive judgment policy would be appropriate. If the judgment criterion is framed as “are controls free of material weaknesses”, then it is necessary that no controls are materially deficient and a disjunctive judgment policy would be appropriate.

Range of relevant models

The five models discussed above may all be potentially relevant for internal control judgments, as shown in the given examples of audit judgment tasks above and in previous research (e.g., Brown and Solomon 1990 and 1991). I refer to Elrod et al. (2004) for further generalization of mathematical models of noncompensatory judgments. The range of relevant judgment policies in internal control judgment tasks may therefore include:

- Linear
- Compensatory
- Amplifying
- Conjunctive
- Disjunctive

3.5.7 Changes in Audit Practice and Regulation

As discussed, prior audit research findings may not be sufficient for describing internal control judgments in today's audit environment. Under SOX 404 reporting (PCAOB 2007) and ISA 315.115 (IFAC 2008), auditors must make accept/reject judgments on the effectiveness of specific types of internal control.⁴⁶ The binary response scale used in this judgment (i.e., accept/reject) may require a different form of cue combination than what is used for control risk judgments made as part of audit planning (i.e., a continuous scale judgment). Furthermore, prior research did not examine the full range of cue interrelationships that result from the increased importance of controls such as the entity's risk assessment process and other entity-level controls. Finally, regulatory changes and changes in audit practice may have caused an increase in the proportion of auditors that combine cues in a configural manner.

Changes in audit regulation and audit practice may therefore have changed both the judgment task (i.e., the task characteristics and the accompanying appropriate judgment policies) and the behavior of the judge (i.e., increased use of appropriate judgment policies compared to findings in Brown and Solomon (1990)).

Current theory does therefore not provide normative evidence and/or descriptive findings on judgment policies as a function of the judgment response scale and/or the full range of cue interrelationships. Such a theory is needed to answer questions like how auditors combine cues when judging a client's multi-step risk assessment process, as required by AS5 (PCAOB 2007) and ISA 315 (IFAC 2008) for audit planning (i.e., continuous control risk judgment) and reporting purposes (i.e., binary effective/deficient control judgment).

⁴⁶ ISA 315.115 discusses judgments where substantive procedures are insufficient so that the auditor must accept/reject that internal controls over financial reporting reduce risk of material misstatement to an acceptable level. ISA 315.120 also requires reporting to management or those charged with governance on material weaknesses in the design or implementation of internal controls which have come to the auditor's attention. Similar requirements exist in ISA 330 for judging operational effectiveness of controls. While the required scope of internal controls judged under ISA 315 and ISA 330 is less extensive than that under AS5 (PCAOB 2007), there is no reason to expect differences in the form of cue integration for those controls that end up being judged (e.g., compare guidance on judging the entity's risk assessment process in ISA 315 versus AS5 (PCAOB 2007)). Similar arguments can be made for US audit of non public companies under AU 319 and AU 325. The task characteristics of binary judgment response scales and cue interrelationships should therefore also be relevant for judgments made under other regulatory regimes than the SOX 404.

3.6 Summary: Contribution from Literature Review

Prior findings on configurality in auditors internal control judgments might not be as strong and relevant as previously believed:

- The current understanding of the findings of configurality in prior research may be based on a misinterpretation of the reported results; auditors were not as able to apply configural judgment policies in control judgments as currently believed.
- Prior research is based on experimental tasks that may not represent the full range of relevant control- and cue interrelationships
- Prior research is based on experiments where it is not clear whether participants actually understand cue interrelationships before they make judgments (i.e., if they absorb the treatment of manipulated cue interrelationships)
- Prior research is based on experimental tasks limited to continuous judgment response scales (as opposed to binary judgment response required in current practice)
- Prior research is based on analysis using compensatory judgment models, while noncompensatory models may be relevant
- Prior research is based on outdated audit regulation and audit practice; both the judgment task (i.e., the task characteristics and the accompanying appropriate judgment policies) and the behavior of the judge may have changed (i.e., increased use of appropriate judgment policies compared to findings in Brown and Solomon (1990)).

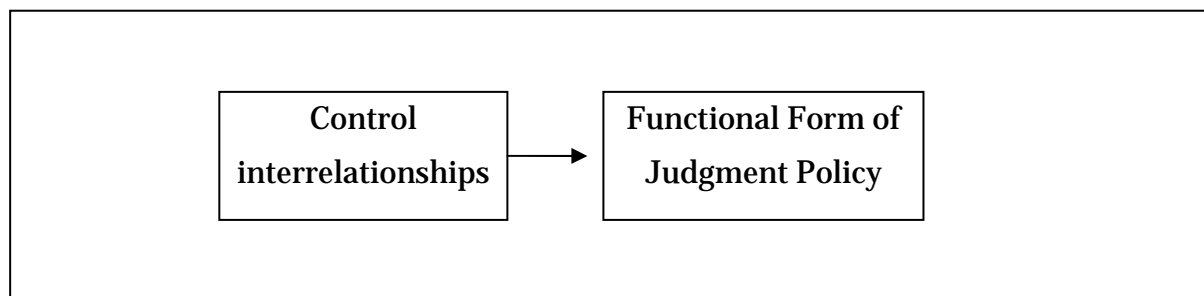
The relevance of prior research findings for describing and/or improving auditor judgment behavior in today's audit environment may therefore be questionable.

4. Conceptual Model and Research Hypotheses

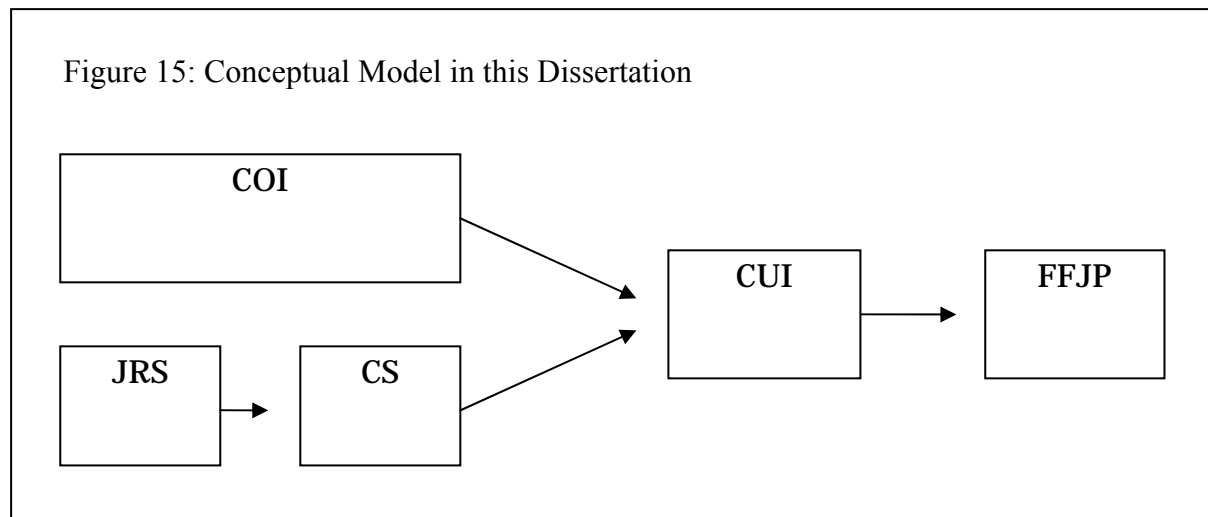
4.1 Conceptual Model and Theory

Prior research has shown that judgment policies with an appropriate functional form may be identified through analysis of control interrelationships (i.e., which were assumed equivalent to cue interrelationships since the judgment response scale was always continuous) (see figure 14 below). Furthermore, some auditors use judgment policies with an appropriate form (e.g., Brown and Solomon 1990, 1991; Hooper and Trotman 1996; Leung and Trotman 2005; Bonner 2007, 154).

Figure 14: Conceptual Model in Prior Studies



This dissertation develops theory that, in essence, states that auditors should combine cues in a similar manner as to how cues relate to the criterion on the environment side of the Lens Model. In addition, theory is developed for how audit task characteristics determine cue interrelationships. The precise model is pictured below (see figure 15 below).



The proposed conceptual model is an illustration of the following theoretical propositions:

- P1: A normatively appropriate functional form of the judgment policy (FFJP) can be derived from studying cue interrelationships (CUI).
- P2: Cue interrelationships (CUI) can be derived from studying control interrelationships (COI) and the judgment response scale (JRS).
- P3: The judgment response scale (JRS) determines the criterion scale (CS).

P1, P2 and P3 have been discussed in the preceding section of the dissertation. In addition, the following proposition is made about audit practice:

- P4: Auditors make judgments by using judgment policies that have normatively appropriate forms of cue integration given the cue interrelationships that result from control interrelationships and the judgment response scale.

P4 can be empirically tested through development of specific hypotheses and a policy capturing approach.

In policy capturing research, ANOVA or regression models are constructed of participant's judgments, and judgment variance attributable to interaction terms is used to examine the extent and form of configularity. Relationships between cue interrelationships and judgment policies are thus consistent with specific cue weight patterns in the mathematical judgment models (i.e., nature, sign and magnitude of cue weights in judgment models). Below,

specific hypotheses are developed in order to test the relationship between the varying forms of control interrelationships, judgment response scales and judgment policies.

4.2 Hypotheses

In order to test the effect of the full range of control interrelationships and relevant judgment response scales on judgment policies, ten hypotheses are formulated. Each hypothesis is accompanied by specific judgment model predictions and a description of the underlying judgment logic.⁴⁷

4.2.1 H1: Multi-Step (i.e., Completely-Dependent) Controls

Multi-step controls result in completely-dependent cues regardless of the judgment response scale. A conjunctive judgment policy is then appropriate and expected to be observed.

Judgment logic:

A multi step control is deficient by definition unless each step is effective; controls are therefore completely-dependent since an individual control step does not reduce risk unless all other control steps are effective. This holds regardless of whether the criterion is binary (i.e., effective/deficient) or continuous (i.e., control risk %-score). Cues are therefore completely dependent, and it is appropriate to use a conjunctive judgment policy (i.e., if one control step is below criterion, the multi-step control as a whole does not function). This result should hold for both continuous and binary judgment response scales.

H1a:

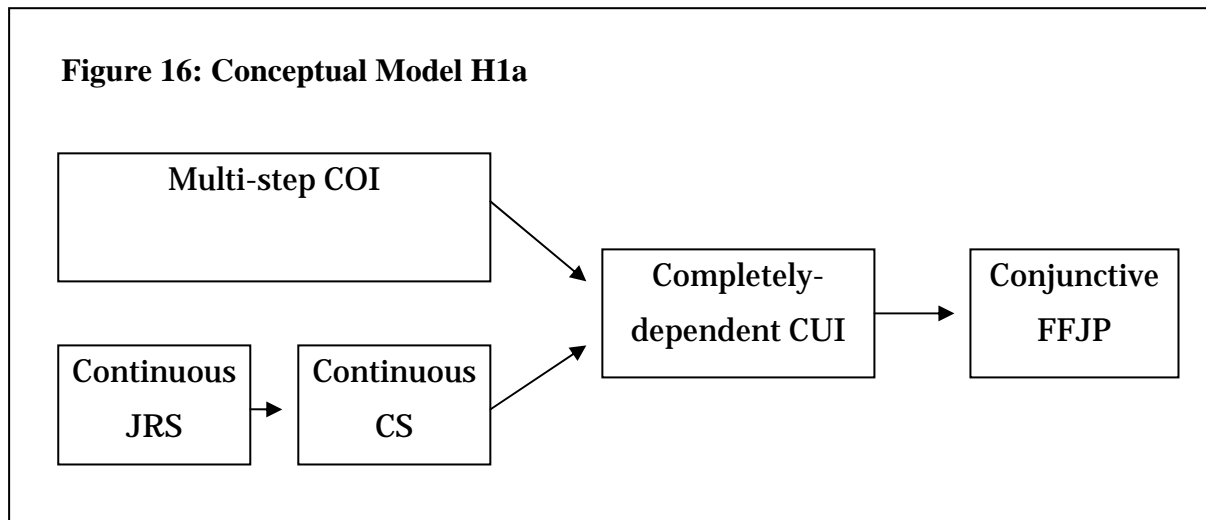
If the judgment response scale is continuous and if controls are multi-step, then the judgment policy will be conjunctive:

Conjunctive FFJP = f{Completely-dependent CUI}

⁴⁷ It is assumed that cues are binary and that there are no trivial cues (i.e., that all cues are relevant in the sense that they could change the judgment independently or depending on the level of other cues).

$$= f\{g[\text{Multi-step COI}, \text{continuous CS}]\}$$

$$= f\{g[\text{Multi-step COI}, h(\text{continuous JRS})]\}$$



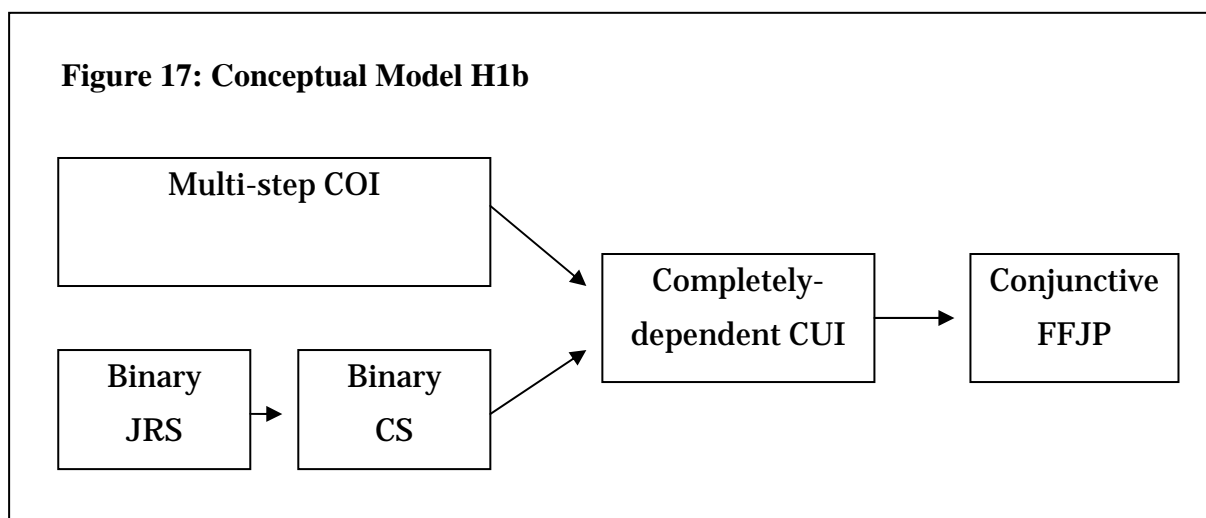
H1b:

If the judgment response scale is binary and if controls are multi-step, then the judgment policy will be conjunctive:

$$\text{Conjunctive FFJP} = f\{\text{Completely-dependent CUI}\}$$

$$= f\{g[\text{Multi-step COI}, \text{Binary CS}]\}$$

$$= f\{g[\text{Multi-step COI}, h(\text{Binary JRS})]\}$$



Judgment model prediction:

Verbally, the model can be explained as follows: In a conjunctive model only the highest order interaction effect will be significant, no main effects or lower order interactions will be significant. The mathematical properties of such a judgment model can be expressed as below:

Assume a two-cue judgment task where “y” represents the judgment and where “y” is a function of controls “c_i”:

Let $y = f(c_i)$, where there are two cues ($i=1,2$) and cues take on the values 0 or 1 ($c_i = 0,1$)

In a judgment task with a continuous judgment response scale, the model should have the following form:

$$0 = 0/100 \leq f(c_1, c_2) \leq 100/100 = 1$$

$$f(0,0) = f(1,0) = f(0,1) = 0, \text{ and } f(1,1) > 0$$

In a judgment task with a binary judgment response scale (i.e., 0/1), the model should have the following form:

$$f(c_1, c_2) = 0, 1$$

$$f(0,0) = f(1,0) = f(0,1) = 0, \text{ and } f(1,1) = 1$$

In a judgment task with a binary response scale, the function thus takes on the value 0 or 1.

4.2.2 H2: Substitutable Controls

Substitutable controls result in substitutable cues regardless of the judgment response scale. A disjunctive judgment policy is then appropriate and expected to be observed.

Judgment logic:

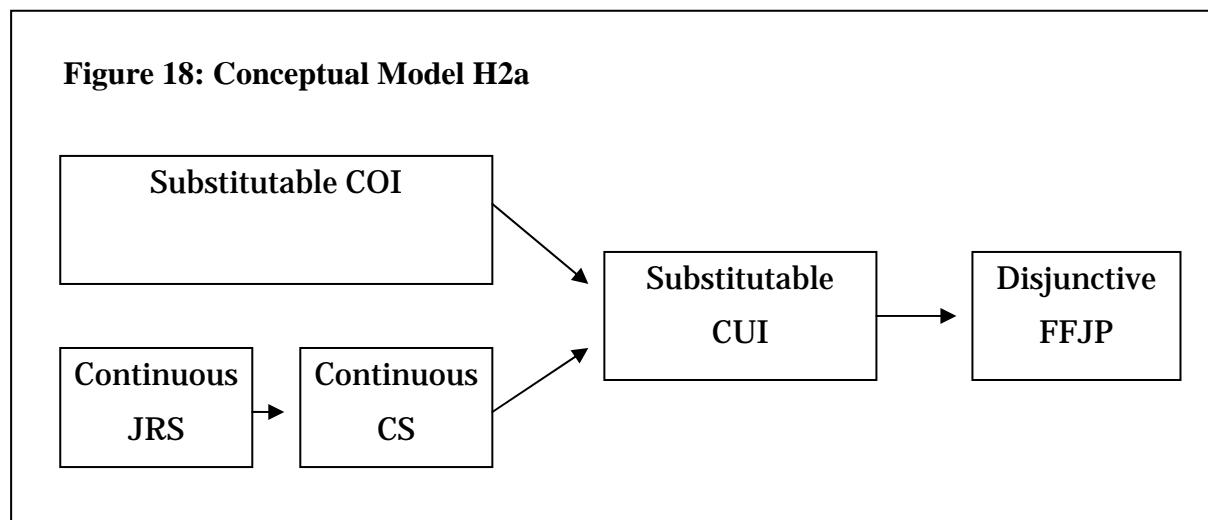
If controls are substitutable, then it is a necessary and sufficient criterion for controls to be judged as effective if at least one of the substitutable controls are above criterion (i.e., effective). If more than one substitutable control is above criterion, it will not contribute to control risk reduction since it does not add to what the first control is already doing (i.e., it

will not change the judgment). It therefore seems appropriate to use a disjunctive judgment policy. This result should hold for both continuous and binary judgment response scales.

H2a:

If the judgment response scale is continuous and if controls are substitutable, then the judgment policy will be disjunctive:

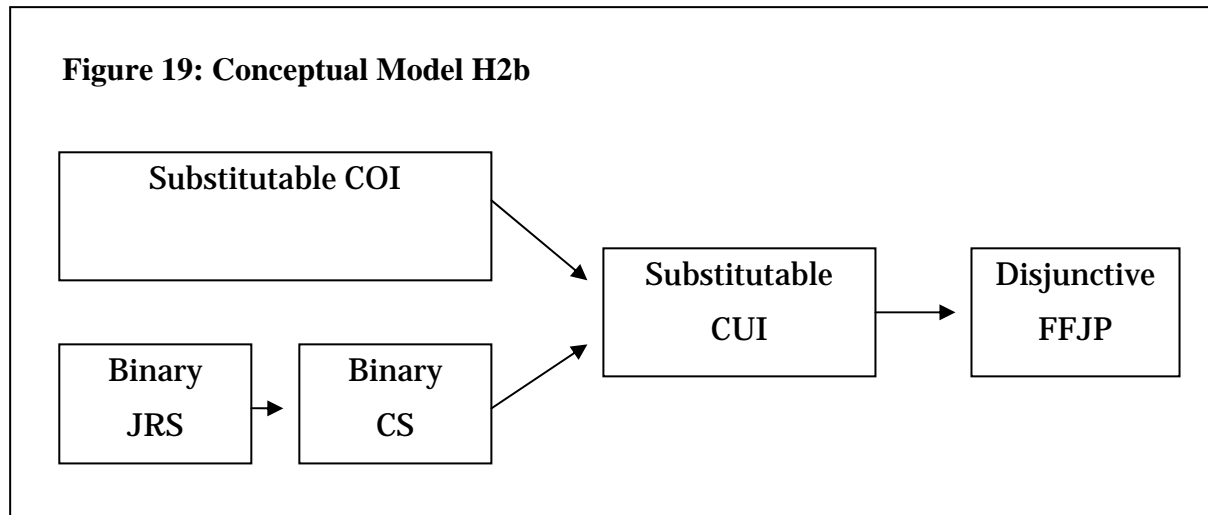
$$\begin{aligned} \text{Disjunctive FFJP} &= f\{\text{Substitutable CUI}\} \\ &= f\{g[\text{Substitutable COI, continuous CS}]\} \\ &= f\{g[\text{Substitutable COI, } h(\text{continuous JRS})]\} \end{aligned}$$



H2b:

If the judgment response scale is binary and if controls are substitutable, then the judgment policy will be disjunctive:

$$\begin{aligned} \text{Disjunctive FFJP} &= f\{\text{Substitutable CUI}\} \\ &= f\{g[\text{Substitutable COI, Binary CS}]\} \\ &= f\{g[\text{Substitutable COI, } h(\text{Binary JRS})]\} \end{aligned}$$



Judgment model predictions:

Predictions for a disjunctive model differ depending on the judgment response scale. The model properties can be expressed as follows:

Assume a two-cue judgment task where “y” represents the judgment and where “y” is a function of controls “c_i”:

Let $y = f(c_i)$, where there are two cues ($i=1,2$) and cues take on the values 0 or 1 ($c_i = 0,1$)

In a judgment task with a continuous judgment response scale, the model should have the following form:

$$0 = 0/100 \leq f(c_1, c_2) \leq 100/100 = 1$$

$$f(0,0) = 0, 0 < f(1,0) = f(0,1) = f(1,1)$$

Verbally, the model can be explained as follows: If the response scale is continuous, main effects and interaction(s) among substitutable control cues will be significant. The main effects of each of the substitutable control cues will be equal since by definition each contributes equally to control risk reduction. The interaction term(s) will serve to remove all but one of the main effects from the substitutable control cues so that the net effect will equal one main effect (since by definition, adding additional substitutable control cues does not contribute to control risk reduction). The interaction term will thus have the opposite sign of the main effects from substitutable cues.

In a judgment task with a binary judgment response scale (i.e., 0/1), the model should have the following form:

$$f(c_1, c_2) = 0, 1$$

$$f(0,0) = 0, f(1,0) = f(0,1) = f(1,1) = 1$$

In a judgment task with a binary response scale, the function thus takes on the value 0 or 1. Verbally, the model can be explained as follows: If the response scale is binary, it is sufficient if at least one of the substitutable controls is above criterion. Adding a second control above criterion will have no effect on the judgment.

If other cues than substitutable cues are present, it is necessary to know the interrelationships of all cues before making specific predictions.⁴⁸

4.2.3 H3: Independent Controls

Continuous judgment response scale

Independent controls result in independent cues when the judgment response scale is continuous, thus resulting in a linear judgment policy being appropriate and expected to be observed.

Judgment logic:

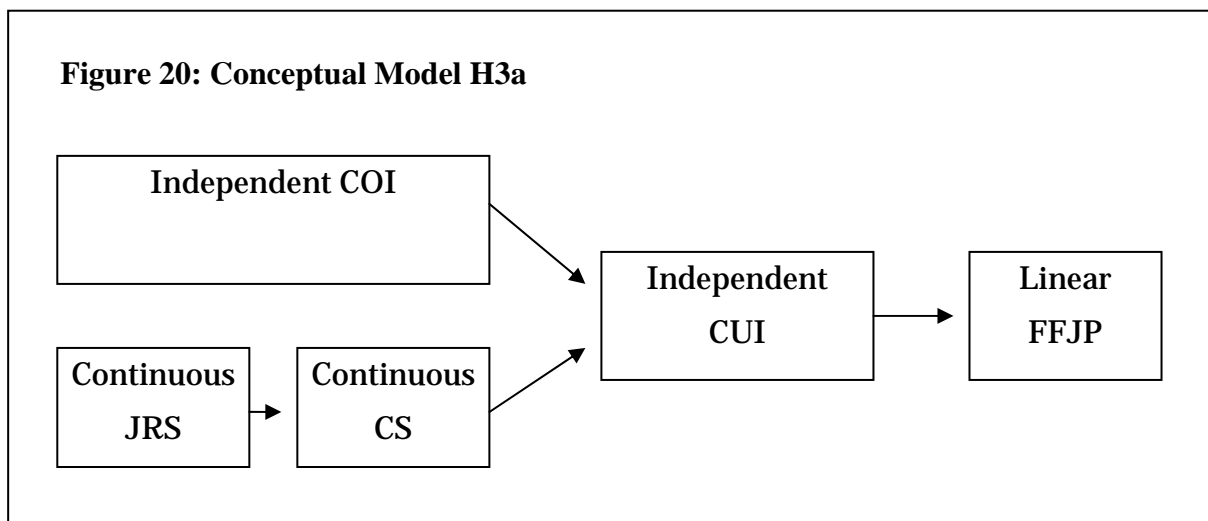
Independent controls do, by definition, only have a linear effect on control risk. They do not interact. In a continuous scale judgment, cue interrelationships are therefore independent and a basic compensatory, additive, linear judgment policy is appropriate (this is referred to as a linear policy in the following).

⁴⁸ A simple example could be the case of two substitutable cues and one independent cue. It would then be necessary and sufficient for effective controls that the independent cue and one of the substitutable cues be effective. Predicted interactions would then be all possible combinations of this; two-way interactions involving the independent cue and one substitutable cue would be significant. Other effects would not be significant (i.e., the three main effects and the two-way interaction among the substitutable cues are not sufficient, and the three-way interaction is not necessary). Predictions for more complex examples can be deducted from the hypotheses; it is therefore not done here. Please note that such a three-cue task also tests for the prediction that only main effects would be significant if all cues are substitutable; it is predicted that interactions involving more than one substitutable cue are not significant, while all interactions involving one substitutable cue are significant.

H3a:

If the judgment response scale is continuous, and if controls are independent, then the judgment policy will be linear:

$$\begin{aligned} \text{Linear FFJP} &= f\{\text{Independent CUI}\} \\ &= f\{g[\text{Independent COI, continuous CS}]\} \\ &= f\{g[\text{Independent COI, } h(\text{continuous JRS})]\} \end{aligned}$$

**Judgment model prediction:**

In compensatory, additive, linear judgment models, the main effects will be significant. No interactions will occur. The mathematical properties of such a judgment model can be expressed as below:

Assume a two-cue judgment task where “y” represents the judgment and where “y” is a function of controls “c_i”:

Let $y = f(c_i)$, where there are two cues ($i=1,2$) and cues take on the values 0 or 1 ($c_i = 0,1$)

$$0 = 0/100 \leq f(c_1, c_2) \leq 100/100 = 1$$

$$f(0,0) = 0, 0 < f(1,0), 0 < f(0,1), f(1,0) + f(0,1) = f(1,1)$$

Binary judgment response scale

Independent controls result in completely dependent cues when the judgment response scale is binary, thus resulting in a conjunctive judgment policy being appropriate and expected to be observed.

Judgment logic:

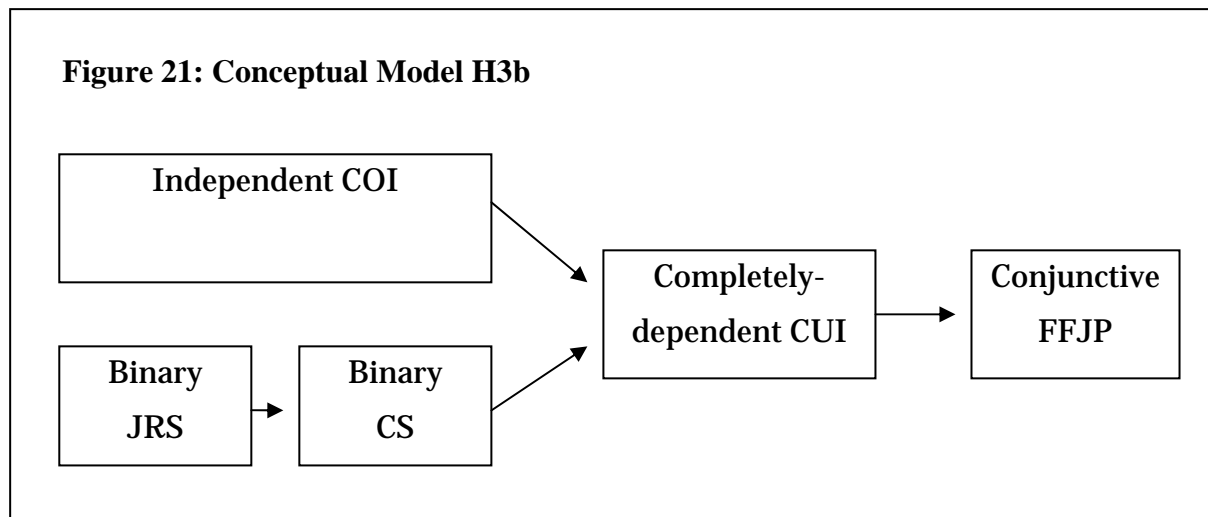
When controls are independent and the judgment response scale changes to binary (i.e., from continuous), the functional form of the relationship of the cues to the criterion changes. In general, when the response scale is continuous, control interrelationships translate directly into cue interrelationships. However, independent cues do not, by definition, exist in binary judgment tasks; cues that are independent in continuous judgment tasks are either (1) individually sufficient in binary judgment tasks (and thus become substitutable), or (2) not individually sufficient in binary judgment tasks (and thus become completely-dependent). This can also be understood as individually sufficient (i.e., substitutable) or necessary (i.e., completely-dependent) conditions for binary judgment outcomes.

For example: if two independent controls do not individually provide sufficient risk reduction, then both controls need to be effective if the binary judgment is to be positive (i.e., that controls are overall effective). The cue interrelationship in the binary judgment would then be that cues are completely-dependent (note that the exact same cues were independent in a continuous scale judgment).

H3b:

If judgment response scales are binary, and if controls are independent, then the judgment policy will be conjunctive:

$$\begin{aligned}
 \text{Conjunctive FFJP} &= f\{\text{Completely-dependent CUI}\} \\
 &= f\{g[\text{Independent COI}, \text{Binary CS}]\} \\
 &= f\{g[\text{Independent COI}, h(\text{Binary JRS})]\}
 \end{aligned}$$



Judgment model prediction:

In a conjunctive model only the highest order interaction effect will be significant, no main effects or lower order interactions will be significant for such controls. The mathematical properties of such a judgment model can be expressed as below:

Assume a two-cue judgment task where “y” represents the judgment and where “y” is a function of controls “c_i”:

Let $y = f(c_i)$, where there are two cues ($i=1,2$) and cues take on the values 0 or 1 ($c_i = 0,1$)

$f(c_1, c_2) = 0,1$

$f(0,0) = f(1,0) = f(0,1) = 0$, and $f(1,1) = 1$

In a judgment task with a binary response scale, the function thus takes on the value 0 or 1.

4.2.4 H4: Compensating Controls

Continuous judgment response scale

Compensating controls result in compensating cues when the judgment response scale is continuous. A compensatory, additive and nonlinear judgment policy with a compensatory form ordinal interaction is then appropriate and expected to be observed.

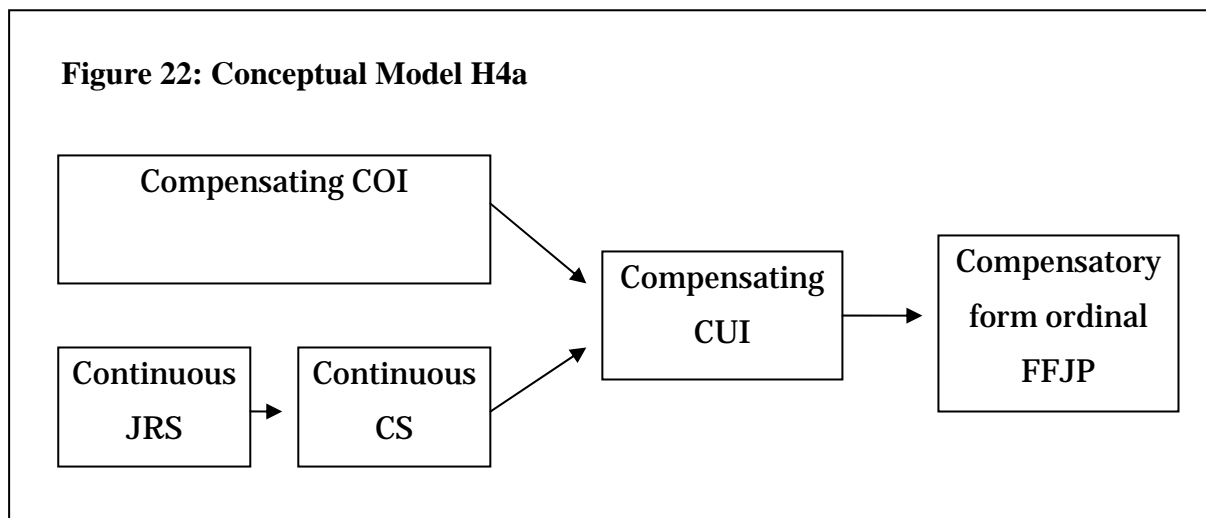
Judgment logic:

If judgments are made on a continuous response scale (i.e., regarding control risk), compensating controls by definition have an individual contribution to control risk reduction and in addition interact in such a manner that the combined effect of both controls is smaller than the sum of their individual effects.

H4a:

If judgment response scales are continuous and if controls are compensating (i.e., compensating cues), then the judgment policy will be compensatory, additive and nonlinear with a compensatory form ordinal interaction:

$$\begin{aligned} \text{Compensatory form ordinal FFJP} &= f\{\text{Compensating CUI}\} \\ &= f\{g[\text{Compensating COI, continuous CS}]\} \\ &= f\{g[\text{Compensating COI, } h(\text{continuous JRS})]\} \end{aligned}$$

**Judgment model prediction:**

A compensatory, additive and nonlinear judgment model with a compensatory form ordinal interaction will have significant main effects and an interaction between the compensating control cues. The interaction will have the opposite sign of the main effects. The mathematical properties of such a judgment model can be expressed as below:

Assume a two-cue judgment task where “y” represents the judgment and where “y” is a function of controls “c_i”:

Let $y = f(c_i)$, where there are two cues ($i=1,2$) and cues take on the values 0 or 1 ($c_i = 0,1$)

$$0 = 0/100 \leq f(c_1, c_2) \leq 100/100 = 1$$

$$f(0,0) = 0, \quad 0 < f(1,0), \quad 0 < f(0,1), \quad 0 < f(1,1) < f(1,0) + f(0,1)$$

Binary judgment response scale

Compensating controls result in completely dependent cues when the judgment response scale is binary. A conjunctive judgment policy is then appropriate and expected to be observed.

Judgment logic:

When controls are compensating and the judgment response scale changes to binary (i.e., from continuous), the functional form of the relationship of the cues to the criterion changes. In general, when the response scale is continuous, control interrelationships translate directly into cue interrelationships. However, compensating cues do not, by definition, exist in binary judgment tasks; cues that are compensating in continuous judgment tasks are either (1) individually sufficient in binary judgment tasks (and thus become substitutable), or (2) not individually sufficient in binary judgment tasks (and thus become completely-dependent). This can also be understood as individually sufficient (i.e., substitutable) or necessary (i.e., completely-dependent) conditions for binary judgment outcomes.

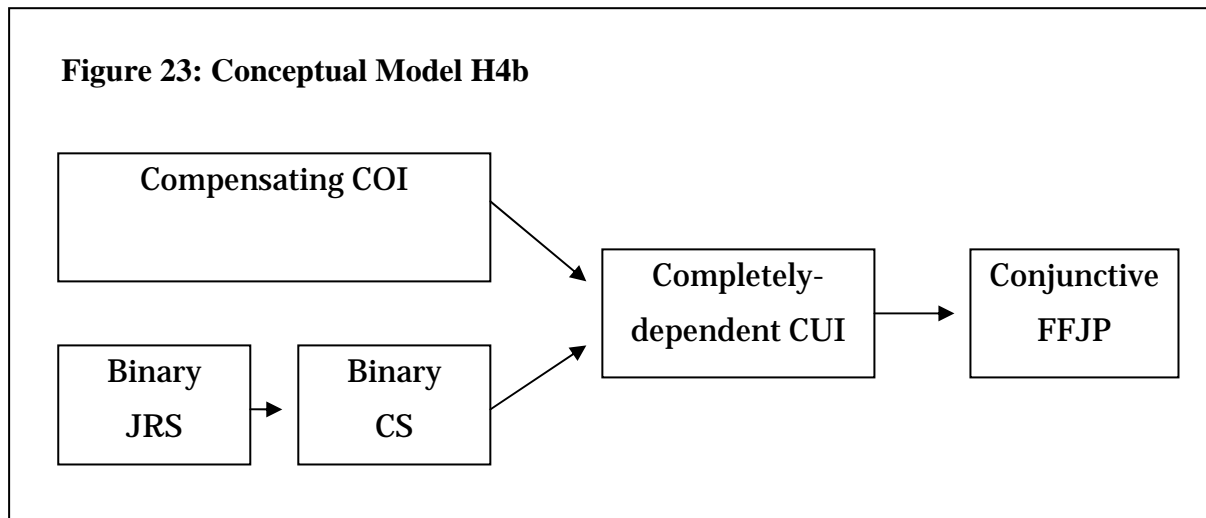
For example: if two compensating controls do not individually provide sufficient risk reduction, then both controls need to be effective if the binary judgment is to be positive (i.e., that controls are overall effective). The cue interrelationship in the binary judgment would then be that cues are completely-dependent (note that the exact same cues were independent in a continuous scale judgment).

H4b:

If judgment response scales are binary, and if controls are compensating, then the judgment policy will be conjunctive:

$$\begin{aligned} \text{Conjunctive FFJP} &= f\{\text{Completely-dependent CUI}\} \\ &= f\{g[\text{Compensating COI, Binary CS}]\} \end{aligned}$$

$$= f\{g[\text{Compensating COI}, h(\text{Binary JRS})]\}$$



Judgment model prediction:

In a conjunctive model only the highest order interaction effect will be significant, no main effects or lower order interactions will be significant for such controls. The mathematical properties of such a judgment model can be expressed as below:

Assume a two-cue judgment task where “y” represents the judgment and where “y” is a function of controls “c_i”:

Let $y = f(c_i)$, where there are two cues ($i=1,2$) and cues take on the values 0 or 1 ($c_i = 0,1$)

$$f(c_1, c_2) = 0, 1$$

$$f(0,0) = f(1,0) = f(0,1) = 0, \text{ , and } f(1,1) = 1$$

In a judgment task with a binary response scale, the function thus takes on the value 0 or 1.

4.2.5 H5: Amplifying Controls

Continuous judgment response scale

Amplifying controls result in amplifying cues when the judgment response scale is continuous. A compensatory, additive and nonlinear judgment policy with an amplifying form ordinal interaction is then appropriate and expected to be observed.

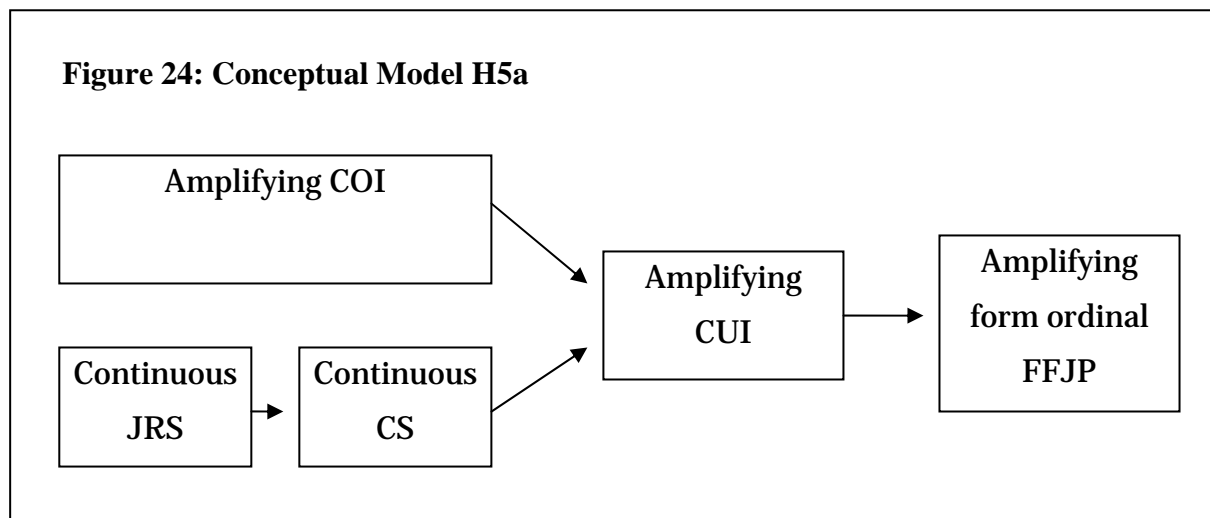
Judgment logic:

If judgments are made on a continuous response scale (i.e., regarding control risk), amplifying controls by definition have an individual contribution to control risk reduction and in addition interact in such a manner that the combined effect of both controls is larger than the sum of their individual effects.

H5a:

If judgment response scale is continuous and if controls are amplifying, then the judgment model will be compensatory, additive and nonlinear with an amplifying form ordinal interaction:

$$\begin{aligned} \text{Amplifying form ordinal FFJP} &= f\{\text{Amplifying CUI}\} \\ &= f\{g[\text{Amplifying COI, continuous CS}]\} \\ &= f\{g[\text{Amplifying COI, } h(\text{continuous JRS})]\} \end{aligned}$$

**Judgment model prediction:**

A compensatory, additive and nonlinear judgment model with an amplifying form ordinal interaction will have significant main effects and an interaction between the amplifying control cues. The interaction will have the same sign as the main effects. The mathematical properties of such a judgment model can be expressed as below:

Assume a two-cue judgment task where “y” represents the judgment and where “y” is a function of controls “c_i”:

Let $y = f(c_i)$, where there are two cues ($i=1,2$) and cues take on the values 0 or 1 ($c_i = 0,1$)

$$0 = 0/100 \leq f(c_1, c_2) \leq 100/100 = 1$$

$$f(0,0) = 0, \quad 0 < f(1,0), \quad 0 < f(0,1), \quad \text{and} \quad 0 < f(1,0) + f(0,1) < f(1,1)$$

Binary judgment response scale

Amplifying controls result in completely dependent cues when the judgment response scale is binary. A conjunctive judgment policy is then appropriate and expected to be observed.

Judgment logic:

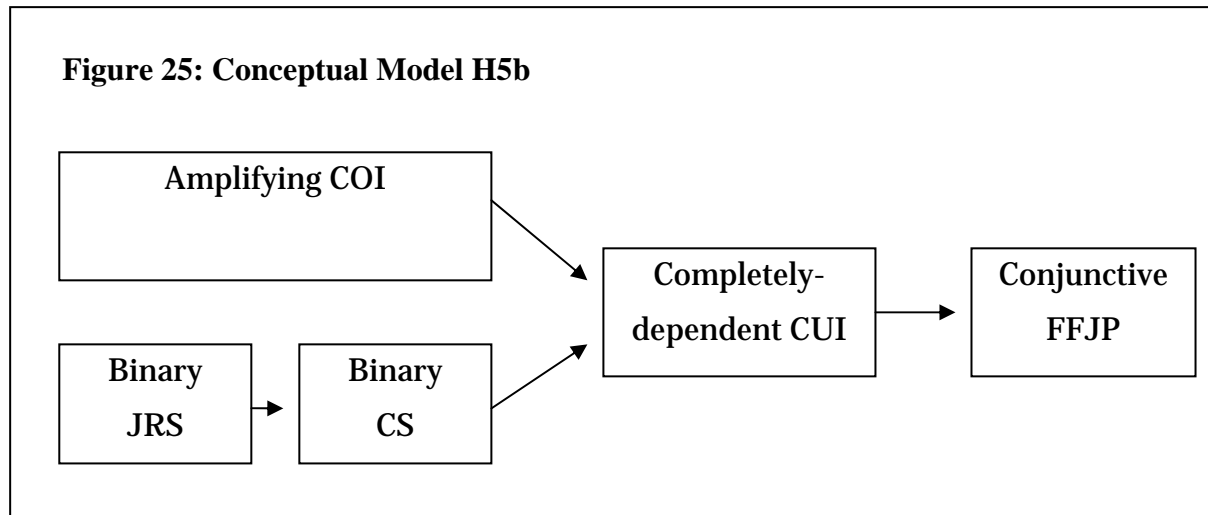
When controls are amplifying and the judgment response scale changes to binary (i.e., from continuous), the functional form of the relationship of the cues to the criterion changes. In general, when the response scale is continuous, control interrelationships translate directly into cue interrelationships. However, amplifying cues do not, by definition, exist in binary judgment tasks; cues that are amplifying in continuous judgment tasks are either (1) individually sufficient in binary judgment tasks (and thus become substitutable), or (2) not individually sufficient in binary judgment tasks (and thus become completely-dependent). This can also be understood as individually sufficient (i.e., substitutable) or necessary (i.e., completely-dependent) conditions for binary judgment outcomes.

For example: if two amplifying controls do not individually provide sufficient risk reduction, then both controls need to be effective if the binary judgment is to be positive (i.e., that controls are overall effective). The cue interrelationship in the binary judgment would then be that cues are completely-dependent (note that the exact same cues were amplifying in a continuous scale judgment).

H5b:

If judgment response scales are binary, and if controls are amplifying, then the judgment policy will be conjunctive:

$$\begin{aligned} \text{Conjunctive FFJP} &= f\{\text{Completely-dependent CUI}\} \\ &= f\{g[\text{Amplifying COI, Binary CS}]\} \\ &= f\{g[\text{Amplifying COI, } h(\text{Binary JRS})]\} \end{aligned}$$



Judgment model prediction:

In a conjunctive model only the highest order interaction effect will be significant, no main effects or lower order interactions will be significant for such controls. The mathematical properties of such a judgment model can be expressed as below:

Assume a two-cue judgment task where “y” represents the judgment and where “y” is a function of controls “ c_i ”:

Let $y = f(c_i)$, where there are two cues ($i=1,2$) and cues take on the values 0 or 1 ($c_i = 0,1$)

$$f(c_1, c_2) = 0, 1$$

$$f(0,0) = f(1,0) = f(0,1) = 0, \text{ and } f(1,1) = 1$$

In a judgment task with a binary response scale, the function thus takes on the value 0 or 1.

4.2.6 Comparison of Hypotheses with Prior Research

Hypotheses in prior research (i.e., Brown and Solomon 1990 and 1991) did not make explicit assumptions about the judgment response scale and the degree of compensation/amplification between controls/cues. In this study such assumptions are made. First, the “a”-hypotheses are limited to continuous judgment response scales and the “b”-hypotheses are limited to binary judgment response scales. Second, the degree of compensation or amplification is taken into account. In other words, while Brown and Solomon (1991) had one hypothesis (i.e., H2) for judgments about cues designed to interact

positively, this study has four hypotheses for the same judgment setting: H1a, H1b, H4a, and H4b. This has been achieved by refining the hypotheses regarding the judgment response scale and the degree of compensation.

4.2.7 Predictions for Three-Cue Judgment Tasks

General three-cue model

Specific predictions are made for judgment tasks with three cues (e.g., cue A, cue B, cue C) since it is the minimum number of cues that permit predictions of specific cue interactions; two cues are needed to show a specific interaction and a third cue is needed to show that it does not interact with the two cues that interact. As such, the third cue is a counterfactual in that since it is not interrelated (i.e., it is independent) it does not interact.

Predictions for more than three cues can be developed using the hypotheses and cue interaction definitions. The three-cue predictions can thus be viewed as construct-level generic predictions of how cue interrelationships affect the functional form of the judgment policy. In section five of this dissertation the generic three-cue predictions are operationalized using specific internal controls.

In three-cue judgment tasks a mathematical model of the judgment policy can potentially have three main effects (A, B, C), three two-way interactions (AB, AC, BC), and one three-way interaction (ABC). The generic judgment model will then have the following form:

$$\text{Judgment} = \alpha_1A + \alpha_2B + \alpha_3C + \alpha_4AB + \alpha_5AC + \alpha_6BC + \alpha_7ABC$$

where,

Cue	Cue weight	Interaction
A, B, C	$\alpha_1, \alpha_2, \alpha_3$	Main effect
AB, AC, BC	$\alpha_4, \alpha_5, \alpha_6$	Two-way interaction effect
ABC	α_7	Three-way interaction effect

Development of predictions for three-cue judgment tasks

In order to develop specific predictions for three-cue judgment tasks, five separate judgment tasks (i.e., series) are created (see table 3 below), where controls are completely dependent, amplifying, independent, compensating or substitutable (i.e., variation in the task characteristics of control interrelationships). Recall that control interrelationships are defined for continuous judgment response scales.

Table 3: Variation in Control Interrelationships in Three-Cue Judgment Tasks

	Series 1	Series 2	Series 3	Series 4	Series 5
Cue A	Multi-step	Independent	Independent	Compensating	Substitutable
Cue B	Multi-step	Amplifying	Independent	Independent	Substitutable
Cue C	Multi-step	Amplifying	Independent	Compensating	Independent

When the judgment tasks (i.e., each series) include both binary and continuous judgments (i.e., variation in the task characteristic of judgment response scale), the ten hypotheses result in the following specific cue weight predictions (see table 4 below). The listed cues are predicted to be significant, and the unlisted cues are predicted to be insignificant. Predictions for magnitude of cue weights are given where these are part of the hypothesized model predictions.

Table 4: Overview of Hypotheses and Predicted Coefficients in Three-Cue Judgment Tasks

Control Interrelationships	Multi-Step	Amplifying	Independent	Compensating	Substitutable
Predictions Continuous Response Scale	H1a $\alpha_7 < 0$	H5a $\alpha_1, \alpha_2, \alpha_3 < 0$ $\alpha_6 < 0$	H3a $\alpha_1, \alpha_2, \alpha_3 < 0$	H4a $\alpha_1, \alpha_2, \alpha_3 < 0$ $\alpha_5 > 0$	H2a $\alpha_1, \alpha_2, \alpha_3 < 0$ $\alpha_1 = \alpha_2 = -\alpha_4$
Predictions Binary Response Scale	H1b α_7	H5b α_7	H3b α_7	H4b α_7	H2b α_5, α_6

Note:

Each cell in the table regards a specific control interrelationship treatment (see column headings) and contains the applicable hypothesis and judgment model predictions for that treatment. Listed alphas are predicted to be significant when regressing judgments on cues in the given treatment. Unlisted alphas are predicted to be non-significant. For the continuous response scale, negative alphas imply that control risk is reduced. For the binary response scale, a listed coefficient implies that it is predicted to be a sufficient criterion for a “yes” judgment. For definitions of control interrelationships see section 3.5.3 and for response scales see section 3.5.5.

$\alpha_1, \alpha_2, \alpha_3$ = main effects of cues

$\alpha_4, \alpha_5, \alpha_6$ = two-way interactions between cues

α_7 = three-way interaction between cues

For example: in table 1, series 4, control B is independent and control A and control C are compensating. Furthermore the judgment response scale is continuous. H4a then predicts that $\alpha_1, \alpha_2, \alpha_3$, and α_5 will be significant and that $\alpha_5 < 0$ (i.e., a compensatory form ordinal interaction). The judgment model is therefore consistent with a judgment policy that is compensatory, additive and nonlinear in a compensatory ordinal form.

Another example is series 5 in table 1; control A and B are substitutable and control C is independent. In the case of a binary response scale, it is hypothesized that it is necessary and sufficient that control C and at least one of the substitutable controls are effective (i.e., main effects are not sufficient and the three-way interaction is not necessary). The predicted coefficients (i.e., α_5 and α_6) are all possible combinations of this (i.e., A and C, or B and C).

5. Methodology and Research Design

5.1 Participants

Twenty-eight audit managers from one Big 4 firm in Norway served as participants in this study. Audit managers are expected to be familiar with overall internal control judgments since they have responsibility for review of audit files and they often perform internal control judgment tasks.⁴⁹ Mean audit experience of the participants was 7.3 years, with minimum of 2.5 years and maximum 12 years. The managers were recruited for participation through an email from the national leader of the firm.⁵⁰

The Norwegian firm uses U.S. developed audit manuals and audit guidance that is common to the firm worldwide. These are in English and based on international audit standards (i.e., ISA 315 and 330 for internal controls in a financial statement audit). Norwegian audit standards are translations of ISA's and thus include a requirement to judge design and implementation of internal control over financial reporting (i.e., as in ISA 315) and to test operational effectiveness of controls that are relied upon to reduce control risk (i.e., as in ISA 330). Material internal control weaknesses are required to be reported to those charged with governance (i.e., as in ISA 315), and may also, under certain conditions, alter and/or be reported in the audit opinion (see discussion in section 2.3 above). Any differences between the audit practices of the Norwegian Big-4 firm and their U.S. counterparts regarding internal control judgments is thus expected to be negligible even though some differences in the scope and purpose of the internal control audit exists.⁵¹ Any bias is expected to work

⁴⁹ Audit manager familiarity with the task was confirmed by the lead technical partner of the participating firm and the firm assurance services leader. Furthermore, Brazel et al (2004) state that audit managers use up to 50% of their available time performing reviews. The reason for not using seniors was that familiarity with entity-level controls was assumed higher for managers.

⁵⁰ Fifty-nine emails were sent out by using a database with the name of all audit managers at the office where the study was conducted. Twenty-three of the managers were not physically working at the office at the time of the study due to maternity leave, overseas work, etc. Of the remaining 36 potential participants, 28 participated, thus resulting in a response rate of 78%.

⁵¹ See Trotman (1996, 86) for a further discussion of why population validity is less of a concern in audit research; potential biases in this study include using auditors from one office of one firm in one country, and using only audit managers.

against the hypotheses since U.S. auditors on average are expected to be more experienced in internal control judgments due to SOX 404.⁵²

5.2 Experimental Design and Procedures

5.2.1 Research Design

The research design is a within participants $5 \times 2 \times 2^3$ design yielding 40 cases with two responses for each case: 5 control interrelationships (i.e., multi-step, amplifying, independent, compensating and substitutable) \times 2 judgment response scales (i.e., binary and continuous) \times 2^3 (full factorial manipulation of three binary cues). For each of the five control interrelationship manipulations (i.e., series), the participants thus provided two responses to each of the eight cases, yielding eighty observations per participant.

5.2.2 Independent and Dependent Variables

Dependent Variable: judgment policy

The dependent variable constructs (i.e., the five judgment policies: (1) conjunctive, (2) disjunctive, (3) linear, (4) nonlinear compensatory form ordinal, and (5) nonlinear amplifying form ordinal) are operationalized by the judgment models (i.e., the nature, sign and magnitude of cue weights of main effects and interactions in the mathematical models of the captured judgment policies – see table 4 and hypotheses with accompanying model predictions).

Independent Variable: judgment response scale

The independent variable construct of judgment response scale was operationalized by (1) a control related risk judgment on a 0-100 point percentage scale (i.e., a continuous judgment response scale), and (2) a “yes”/“no” (i.e., binary) answer to the question “Does the client

⁵² Although there are similarities to the SOX 404 regime regarding judgments about internal control over financial reporting, there is no separate requirement to test operational effectiveness of internal controls over financial reporting such as in SOX 404.

(i.e., all material locations) have sufficient operationally effective controls for the given audit area?

In the continuous judgment response scale the participants were asked to judge “control risk”. “Control risk” was loosely defined as the risk of error after the company has performed controls. This is similar to the judgment response scale applied in Brown and Solomon (1990), which is the only prior study of internal control judgments finding evidence of configularity. Their control risk judgment scale was defined as “Given the controls as represented above, assess the risk that cash disbursements could be materially misstated as a result of checks being written and/or distributed for improper (unauthorized and/or invalid) purposes”.

Participants were informed that “inherent risk had been fixed at 100% and should therefore not impact the control risk judgment. It was furthermore referred to “CR” in the Audit Risk Model: $AR=IR*CR*DR$, where AR =Audit Risk, IR=Inherent Risk, CR=Control Risk, DR=Detection Risk (see ISA 200.29 (IFAC 2008) for a description of the model).

Independent Variable: control interrelationships

Interrelationships of controls are manipulated using three-cue judgment tasks as described in table 3. The five treatment levels of interrelationships of controls are (see series in table one): (1) one series with three multi-step control cues that are completely dependent (series one), (2) one series containing three independent controls (series three), and (3) three series containing one independent control each and two controls that were either amplifying (series two), compensating (series four) or substitutable (series five). An independent control was included in series two, four and five so that specific predictions could be made about which control cues interact and which do not. The order of the independent control was varied between series two, four and five (see table 3); in series two, Cue A was the independent control. In series four, Cue B was the independent control. In series five, cue C was the independent control. Since each cue is binary, a full factorial manipulation of the three cues within a series yields eight cases (i.e., $2^3=8$).

Through the various combinations of control interrelationships and judgment response scales, the five types of cue interrelationships described in the cue interrelationship

framework are achieved. The cue interrelationships are in other words operationalized through the operationalization of control interrelationships and judgment response scales.

Setting: control interrelationships

Controls over accuracy of booked invoices in the purchasing cycle were selected as the primary setting for operationalizing control cues.⁵³ Limiting the control setting to one cycle and one control objective facilitates participant understanding of control interrelationships and reduces potential noise. A hypothetical audit client with three similar locations and three accounting controllers was developed. Cue levels were manipulated through varying audit test results from testing of each controller (i.e., each controller represents one of three cues), where each controller may perform one or two controls (see example in Table 5 below):

- In the base case, each controller performs accuracy controls over invoices from one location (i.e., independent controls).
- Substitutable controls are achieved by having a second controller double check all of the first controller's invoices (i.e., complete overlap in work performed).
- Compensating controls are achieved by having a second controller double check half of the first controller's invoices (i.e., partial overlap in work performed).
- Multi-step controls are operationalized by having each controller perform one control step in a basic three step risk management process (i.e., a fully joint effort).⁵⁴
- Amplifying controls are operationalized by having a controller perform a compound control where part of the control is independent and part of the control is a joint effort. Amplification is achieved since the risk effect (i.e., reduction in control risk) of the first controller's effort increases if the second controller performs his part of the joint effort.

⁵³ This is assumed to be a basic business cycle well known to audit managers. Overview tables of relevant controls from an audit textbook (Eilifsen, Messier, Glover and Prawitt 2006, p388), along with firm audit manual guidance, served as the main source for generating potential control cues.

⁵⁴ The three risk steps are based on the generic risk process described in AICPA audit guide para 4.45 (2006) and ISA 315.77. Participants are furthermore assumed familiar with this process due to the existence of COSO Enterprise Risk Management – Integrated Framework (COSO 2002).

Table 5: Example of Compensating Control Case

Audit Area: accuracy of booked incoming invoices			
Control Design:	Control Objective	Operationally Effective? (Audit test result)	
1. Accounting <u>controller "A"</u> reviews <u>all</u> (i.e., 100) booked incoming invoices from <u>location "A"</u> to check that they are booked in the accounting system with the correct amount.	Accuracy	YES	
2. Accounting <u>controller "B"</u> reviews <u>all</u> (i.e., 100) booked incoming invoices from <u>location "B"</u> to check that they are booked in the accounting system with the correct amount. In addition, accounting <u>controller "B"</u> double checks <u>half of all</u> (i.e., 50) booked incoming invoices from <u>location "C"</u> to check that they are booked in the accounting system with the correct amount.	Accuracy	YES	
3. Accounting <u>controller "C"</u> reviews <u>all</u> (i.e., 100) booked incoming invoices from <u>location "C"</u> to check that they are booked in the accounting system with the correct amount. In addition, accounting <u>controller "C"</u> double checks <u>half of all</u> (i.e., 50) booked incoming invoices from <u>location "B"</u> to check that they are booked in the accounting system with the correct amount.	Accuracy		NO

Letting controllers perform one or two controls and focusing audit test results on the controller (instead of on each individual control) allows flexibility in constructing control interrelationships while keeping the setting constant (i.e., controllers, locations and invoices) and using only three cues (i.e., controllers). If focus had been on individual controls (assuming that a constant setting was desired), five controls would have to be manipulated in the compensating, amplifying and substitutable series, thus increasing the number of cases in each of these series from 8 to 32. This choice is a trade off; keeping the setting constant and using only three cues makes the interpretation of cue levels more demanding, since each cue level may relate to two controls. However, participants still have to process information regarding control interrelationships in order to make judgments (consistent with the purpose of the study). If five controls were used, boredom/fatigue would become a serious issue since 112 cases would be needed ($3 \times 32 + 2 \times 8$). If the setting was allowed to vary, three cues could be used, but this would create a need for much more extensive background material, thus increasing the risk of noise and/or alternative explanations for judgments. The chosen

method of operationalization thus seems most efficient in facilitating understanding of cue interrelationships and reducing potential noise while obtaining the purpose of the study.

The hypothetical audit client is constructed based on reasonable real life traits: A setting with several similar locations and dedicated controllers could exist in retail businesses like fast food franchises. Double checks could be used for sample based follow up of new controllers, internal audit, control monitoring or for control of transactions/areas where double controls are desired (e.g., by dual signature). Cooperation (i.e., amplification) could be required if each controller possesses a necessary but not sufficient competence (e.g., language, GAAP, IT, business knowledge). Focusing on the percentage of invoices (i.e., transactions) controlled for accuracy per location is also considered realistic and consistent with AS5 (PCAOB 2007) which focuses on transactions, classes of transactions, assertions and locations (see AS5.A5, AS5.29, AS5.35, and AS5.B10) (e.g., the auditor ultimately makes a judgment about the accuracy of an account that is populated of individual invoices with individual values coming from different locations).

5.2.3 Materials and Procedures

Setting

The experiment was administered by the author in one sitting at the audit firm's office (i.e., a controlled setting). Participants were told that a survey was being conducted in order to understand auditors' internal control judgments. In order to increase accountability, participants were asked to write their name on the folder including the materials.

Materials

Each participant received a folder with an introduction and six numbered envelopes. Participants were instructed to read the introduction first. The introduction contained instructions regarding the task to be performed and brief background information about the hypothetical company being audited. Participants were told that they were audit managers and that they were to review the documented results from testing of operational effectiveness of internal controls in order to make two judgments; (1) control risk in the area being audited, and (2) does the client (i.e., all material locations) have sufficient operationally effective controls for the given audit area? The instructions also included a framework for

classifying cue interrelationships (i.e., in the form of five definitions of controller cooperation). For complete instructions and background material, see appendix 1.

After reading the instructions, participants opened envelope one and documented their responses to the questions being asked. After responding, participants sealed their responses in envelope one and repeated the same procedure for envelope two to five (i.e., five envelopes requiring responses to cases).

Envelopes

Each envelope contained nine pages. The first page contained a description of three control cues (i.e., the control work performed by the three controllers) with accompanying control objectives.

AS5.A2 (PCAOB 2007) defines a control objective as: “A **control objective** provides a specific target against which to evaluate the effectiveness of controls. A control objective for internal control over financial reporting generally relates to a relevant assertion and states a criterion for evaluating whether the company's control procedures in a specific area provide reasonable assurance that a misstatement or omission in that relevant assertion is prevented or detected by controls on a timely basis”

Control objectives were indicated in order to make control interrelationships more salient (i.e., if controls relate to different control objectives they are independent, but if they relate to the same control objective they could be interrelated). Inclusion of control objectives is consistent with how the participating audit firm documents assessments of internal controls, and thus makes the task more realistic (i.e., ecologically valid). Furthermore it is consistent with the emphasis Brown and Solomon (1991, 103) place on control objectives when discussing control interrelationships. Making control interrelationships more salient is not thought to create a demand effect per se (Schepanski et al. 1992).

Participants were asked to classify each cue as belonging to one of the five definitions in the controller cooperation framework (e.g., compensating, independent). The primary purpose of the classification exercise was to stimulate participants to engage in deeper processing of the interrelationships between given control cues prior to completing the cases, but without creating a demand effect.

The purpose of the classification exercise can be explained by a metaphor; if studying how fast subjects run, subjects must first be stimulated to run instead of walking, but without the stimulation affecting how fast they run. Hooper and Trotman (1996) found that one of the main differences between configural and non-configural processors was that the former believed cues were interrelated while the latter considered cues to be independent. Auditor perception of interrelated cues may therefore increase the general likelihood of configural cue processing occurring. The classification exercise in this study thus makes the treatment more salient, and this may increase the likelihood of configural cue processing. Since cue interrelationships in general are made more salient, there is no reason to expect a bias as to how different cue interrelationships impact the form of the judgment policy (i.e., the likelihood of presence of configurality in general is increased (given interrelated cues), but the kind of configural cue processing should not be differentially affected between treatments). This is consistent with Schepanski et al. (1992) who noted that, “after all, cue salience is not viewed as a causal variable per se”, and that “although salience may alter effect sizes of a variable that is causally relevant, it would not do so for a variable that is causally irrelevant”. Any potential effect size impact of cue interrelationship salience is irrelevant to this study since separate analysis is conducted for each cue interrelationship treatment.

The classification exercise also served a secondary purpose; as a direct measurement of participant understanding of control interrelationships (i.e., as a manipulation check of the control interrelationship treatment). However, due to potential difficulties in understanding the framework the classification responses may be a noisy measure of the understanding of control interrelationships.

Pages

Each of the following eight pages contained the manipulation of the cue levels within the series (i.e., a “YES” or “NO” response to whether each cue was operationally effective), a control objective for each cue (i.e., accuracy), a requirement to document a control risk judgment as a percentage score, and a requirement to document an overall control sufficiency judgment by circling either “YES” or “NO” (i.e., effective or deficient). Each envelope thus contained a full factorial manipulation of three binary cues representing one of the five control interrelationship treatments (i.e., each page was a case and each envelope was a series of cases) (an example of a case from each series is listed in appendix two).

The response scale treatment was thus manipulated within subjects within the same case (i.e., page). This allowed keeping everything constant from case to case apart from control cue levels (within an envelope) and control interrelationships (between envelopes). The alternatives of (1) manipulating response scales between subjects or (2) repeating cases twice within subjects but with varying response scales is thought to be less efficient by using more subjects or more subject time, and furthermore risk creating more noise or memory carry-over effects. Again, making variation in the response scale variable more salient is not, per se, thought to create a demand effect (Schepanski et al. 1992). Furthermore, since auditors make both judgments in practice, external validity is not impaired by measuring both responses within the same case. Finally, it can be noted that the responses to each response scale are expected to be naturally correlated (e.g., overall deficient controls should lead to higher control risk than overall effective controls).

Within one envelope, participants were allowed to respond to the cases in the order they wished, and to change their previous answers if they wished (pages were stapled together to preserve the order, but participants could flip back and forth as they wished). However, when one envelope was completed, the answers were sealed in the supplied envelope and could not be changed.

Allowing subjects to change answers within an envelope (i.e., for one control interrelationship condition) is thought to reduce noise since subjects are allowed to process variation in control cue levels taking into account the entire variation in the cue set. Furthermore, this should not create a bias since cue level variation is not part of the hypotheses (i.e., the research question is about how different cue interrelationships affect judgments, not about how cue levels affect judgments). Subjects could therefore not alter judgments in a completed and sealed envelope (i.e., for different cue interrelationship conditions).

Prior research has shown that participants are relatively successful in applying judgment policies appropriate to independent and compensating cues (e.g., Brown and Solomon 1990). In order to allow participants to practice on the assumed easiest cases, all participants therefore started with the envelope containing three independent control cues and then continued with the envelope containing one independent and two compensating control cues. The order of the three remaining envelopes was varied in the six possible orders (see table 6

below). Six repetitions of each envelope order were created, thus creating thirty-six folders in total (e.g., six folders with envelope order A).

Table 6: Order of Envelopes in Experiment

Order A	Order B	Order C	Order D	Order E	Order F
Independent	Independent	Independent	Independent	Independent	Independent
Compensating	Compensating	Compensating	Compensating	Compensating	Compensating
Substitutable	Substitutable	Multi-Step	Amplifying	Multi-Step	Amplifying
Multi-Step	Amplifying	Substitutable	Substitutable	Amplifying	Multi-Step
Amplifying	Multi-Step	Amplifying	Multi-Step	Substitutable	Substitutable

The envelopes were then filled with cases. The order of the three cues within a case was fixed, but the order of cases within an envelope was systematically varied: First, a basic order of cases was developed and each case was numbered (see table 7 below):

Table 7: Order of Cases within an Envelope

Case nr	Cue A	Cue B	Cue C
1	Y	Y	Y
2	N	Y	Y
3	Y	N	Y
4	Y	Y	N
5	N	Y	N
6	Y	N	N
7	N	N	Y
8	N	N	N

Second, four orders of cases were then developed by using a random number generator (see table 8 below; case nr. from the table above is indicated in each cell):

Table 8: Order of Cases

order 1	order 2	order 3	order 4
2	3	5	6
8	5	8	2
1	2	4	3
6	8	6	8
4	6	3	7
7	4	1	1
3	1	7	4
5	7	2	5

Recall that each envelope order was repeated six times. When filling the envelopes, the envelope order was crossed with the case order in order to first produce 24 basic orders (i.e., 6 envelope orders multiplied by 4 case orders) (see table 9 below). The first six folders (i.e., one from each envelope order) were filled in the following manner (see table below): “i1” indicates independent cues in order one, “m3” indicates multi-step cues in order three, “s2” indicates substitutable cues in order 2, etc:

Table 9: Crossing of Envelope Order with Case Order

Order A	Order B	Order C	Order D	Order E	Order F
i1	i2	i3	i4	i1	i2
c2	c3	c4	c1	c2	c3
s3	s4	m1	a2	m3	a4
m4	a1	s2	s3	a4	m1
a1	m2	a3	m4	s1	s2

For the next six folders, the number in each cell in the table above was increased with one (e.g., order A now started with i2 and continued with c3). This was done two more times so that the 24 basic orders were repeated. The remaining two folders in each order were filled in the following manner: For order A, the order of the first and second round of filling was repeated. For order B, the order of the second and third round of filling was repeated, etc.

This resulted in 24 basic orders, of which half were repeated once – thus resulting in 36 folders. 28 participants took part in the experiment. The 24 basic orders were first handed out. For the remaining four participants, one folder from order A to D was handed out. The 24 basic orders were thus used, and four of these were repeated once. Order effects are therefore unlikely to affect results since they must be due to the four cases that were repeated

once. No testing of order effects was done since the number of observations for each order is too low (i.e., 16 observations for a repeated order would need to be compared to the observations for each of the 23 other possible orders).

Participants were allowed to access the introductory materials during the whole experiment (i.e., the task description, background information and definitions of control interrelationships).

Envelope six, which was opened after envelope five was sealed, contained a post experimental survey and manipulation checks (see appendix three). After all envelopes were completed, they were put in the folder and returned to the author.

5.2.4 Pilot Testing

The introductory materials, a selection of 20 cases and the post experimental survey were pre-pilot tested on 12 consultants from the participating firm (all of which had relevant internal control experience). A final, and full, pilot test was performed with 11 audit seniors from the participating firm. The final cues are listed in appendix two.

6. Hypotheses Testing

6.1 Introductory Discussion

6.1.1 Statistical Analysis

Level of analysis

In an analysis of individual judgment policies (i.e., ideographic analysis) the number of cases determines the number of observations for each analysis (i.e., analysis is done for each individual by regressing the individuals judgments on the cues).

In individual level analyses using three cues without replication of cases, each cell has only one observation; eight cases are used to estimate an intercept term and seven effect terms. The models will then be fully determined and have no error estimate. Formal significance testing can therefore not be performed. However, prior research has used an arbitrary 4% criterion for significance of interaction terms, and shown the relevance of this criterion for judgment quality (Brown and Solomon 1990 and 1991; Hooper and Trotman 1996). Such an approach would give comparability with prior research regarding extent of configularity (i.e., number of judges with configural judgment policies), but statistical testing of the form of configularity would be based on an arbitrary criterion. Furthermore aggregation of results regarding form of configularity would be difficult. Since the focus of this study is on the form of configularity, such an analysis is not performed for hypotheses testing.⁵⁵

With regression techniques aiming for stable regression weights, the preferred ratio of observations to the number of factors is 10:1, with the minimum being 5:1 (Cooksey 1996, 123; Karren and Barringer 2002; Aiman-Smith et al. 2002). Since this study applies three-cue judgment tasks with seven potential factors for each judgment, individual-level statistical analysis is not feasible: The minimum 5:1 ratio would require $8 \times 5 = 40$ cases for each of the five series, thus requiring subjects to judge 200 cases each. Furthermore,

⁵⁵ In this study, the cue effect sizes would be very sensitive to the percentage of invoices each controller checked or double checked. The 4% arbitrary significance criterion from prior research would therefore not be appropriate.

achieving 40 cases with three cues would require repeating cases on average five times within a series, since three cues only yield $2^3=8$ cases. This would create an unacceptable risk of memory carry-over effects and boredom/fatigue (Cooksey 1996, 123). Finally, when cues are binary, adding a cue doubles the amount of possible cases, but the number of interactions increases faster; each new cue introduces a main effect and potential first order interactions with the other cues. This doubles the number of cases needed and the net ratio gain is zero. If higher order interactions are relevant for the study, the number of cases needed grows faster than the doubling of cases provided by the additional cue. When cues are binary and higher order interactions are relevant, increasing the number of cases by increasing the number of cues is therefore not a viable strategy. Statistical analysis can therefore only be performed for all participants aggregated (i.e., nomothetic analysis).

Method of analysis

The choice of variables restricts the choice of statistical method of analysis.⁵⁶ When judgment responses are elicited on a continuous scale, ANOVA or OLS (Ordinary Least Squares Regression) may be used. However, when a binary response scale is used (i.e., controls being deficient/efficient), two key OLS assumptions are violated; (1) the model error distribution is discrete, therefore, non-normal, and (2) error or residual variance is not homogeneous. Binary logistic regression is then appropriate (Cooksey 1996, 273-280).

This study uses OLS to estimate judgment models for the continuous response scale. The main argument for doing so is that this is more efficient since interaction coefficients can be directly tested. In ANOVA, contrast coding would have to be performed to test for lower order interaction effects when the three-way interaction turns out significant. The binary judgment models are estimated using binary logistic regression. This results in a total of ten models being estimated in order to test the hypotheses; one for each hypothesized combination of response scale and cue interrelationship.

⁵⁶ If cues are continuous, OLS is generally used. If cues have few levels, ANOVA is generally used (Cooksey 1996, 273-280). If cues and responses are categorical, logistic regression or the nonmetric Lens Model may be used (Cooksey 1996,280). All previous audit internal control studies conducted using the policy capturing paradigm have, to the best of the authors knowledge, used binary cues, continuous dependent variables, and ANOVA.

Potential issue with binary logistic regression

One potential issue with binary logistic regression arises if one of the independent variables perfectly, or near perfectly, predicts the dependent variable. The estimation of coefficients in binary logistic regression is performed in order to find the slope of the factor that best fits the data. When a factor is able to (near) perfectly predict the data, this implies that, in principle, a (near) infinite number of slopes fit the model for this factor. If the statistical software program (in this case SPSS) is forced to estimate the slope, the estimate will be meaningless (i.e., have no practical interpretation).

In this study, the three-way interaction is hypothesized to perfectly predict the judgment for H1b, H3b, H4b and H5b. For H2b, similar issues arise with two-way interactions. The experimental results show that practically all participants judged controls to be sufficient when all cues were effective (i.e., a three-way interaction).⁵⁷ The three-way interaction thus perfectly, or near perfectly, predicts effective controls. SPSS (v.16) reacted by estimating p-values between 0.998 and 1.00 for all coefficients.⁵⁸ The binary logistic regression did therefore not provide usable results.

Instead of conducting a statistical analysis at the aggregated level, a detailed analysis of individual judgment policies can be performed. Since regression is not feasible since each participant has only eight cases per series, another approach must be used to test the hypotheses about binary judgments.

In this study, a conjunctive judgment policy is hypothesized for four of the five binary hypotheses (i.e., for H1b, H3b, H4b and H5b which provide predictions for binary judgments with multi-step, independent, compensating, and amplifying controls respectively). Such a judgment policy implies that all controls must be effective for a positive judgment to occur. If one or more controls are deficient, then a negative judgment should occur. With the three-cue judgment tasks used in this study, such a judgment policy can easily be observed from the case responses; only one case should be answered positively (i.e., the one with all cues being effective). The remaining seven cases should be answered negatively. The number of participants having such a case response pattern was therefore

⁵⁷ The two only exceptions being one participant in the independent series and one participant in the compensating series

⁵⁸ Similar results were achieved when performing analysis in STATA.

counted. No statistics is needed at the individual judge level to observe whether this one case is judged positively and whether the seven other cases are judged negatively.

Similarly, for H2b, the hypothesized judgment policy implied that three cases should be answered positively; the one with three effective controls, and the two where the independent control and one of the substitutable controls were effective. The remaining five cases should be answered negatively. The number of participants having such a case response pattern was therefore counted. No statistics is needed to observe at the individual level whether these three cases are judged positively and whether the five other cases are judged negatively.

It was thereafter tested whether the number of participants showing the predicted case response pattern (i.e., judgment policy) was significantly different from random by using the test approach in Brown and Solomon (1990, 30). The procedure is a binomial test and was applied as follows: First it was assumed that if participants used a random policy, the probability of having the hypothesized policy versus any other policy would be 50%. A binomial distribution could then be applied to calculate whether the number of auditors using the appropriate policy was statistically greater than 50% and thus not due to randomness. For example, the probability of observing 18 of the hypothesized judgment policy and three other judgment policies, if auditors choose whether to use the hypothesized policy or any other policy at random, is <0.001 . It is therefore very unlikely that the use of the hypothesized judgment policy is due to randomness. The test of hypotheses section below presents the results, including observed exceptions from predictions.

6.1.2 Control Interrelationship Treatment

Prior research has shown that auditor's perception of cue interrelationships may vary and that this may be important for the extent and form of configurality in judgment models (Hooper and Trotman 1996). Differences in cue interrelationship perceptions also indicate that the researcher's manipulation may not be having the desired effect. This may again pose a threat to construct validity and internal validity of the research.

In this study, hypotheses predict how actual (i.e., apriori designed or intended) control (and cue) interrelationships should affect judgment models. Hypotheses are tested by comparing predictions for a given control (and cue) interrelationship treatment with regression model

parameters from cases that correspond to that specific control (and cue) interrelationship treatment. The simultaneous testing of control and cue treatments is due to the fact that a manipulation of control interrelationships and judgment response scales imply a specific cue interrelationship. The judgments for a given cue interrelationship treatment are, however, based on how participants perceive cue interrelationships. For example; the hypotheses about the effect of compensating cues (i.e., achieved by compensating controls and a continuous judgment response scale) on the judgment model is tested against regression model parameters obtained from the cases that were designed to include compensating cues (i.e., the compensating control series). If some participants perceive these cues to be interrelated in another manner, for example amplifying, then the compensating cue treatment is not absorbed, and the observations from these participants should be deleted from further analysis. In general, unless cue interrelationships are perceived as intended, the cue interrelationship treatment can not be said to be absorbed. Participants with cue interrelationship perceptions that deviate from the intended should therefore, in general, be omitted from further analysis.

One way of checking the validity of the case design is to observe whether only interactions representing the intentionally designed cue interrelationships are statistically significant (Brown and Solomon 1990, 32). In this study, the regression models on the full sample of participants (i.e., all 28) included statistically significant, non-intended interactions. This was an indication of potential issues with the case design (i.e., participant understanding of cue interrelationships). An analysis of individual judgments was therefore conducted.

Individual responses to cases revealed that seven out of 28 participants had responded to the control risk judgment by using mostly 0% or 100% responses. Such a response pattern indicates a systematically different perception of cue interrelationships than what was intended. These seven participants were therefore contacted and asked open questions about their understanding of the background material in the experiment and their judgment logic. Responses revealed that they had misinterpreted the background information on inherent risk; the inherent risk level was interpreted as all locations having at least one material error or as all transactions being erroneous prior to company control. The intention was for participants to perceive inherent risk as there being one material error somewhere in the

company (i.e., one of the locations had one material error). This misinterpretation of inherent risk results in cue interrelationships being different from intended.⁵⁹ These seven subjects were therefore deleted from further analysis. It is, however, commented below whether inclusion of the seven participants would have changed the results. Any other issues with perception of individual cues are discussed in the results section.

The regression models on the remaining sample of participants (i.e., n=21) did not include non-predicted interactions. It is therefore assumed that no systematic misinterpretation of the case design occurred for these participants. Any deviating results are discussed below.

6.2 Post Experimental Questions

Table 10: Summary of Post Experimental Question Responses

#	Question content	Scale / Anchor	Average	Std.dev.	Max.	Min.
1	Years of experience as an auditor	Years	7.3	2.5	12	2.5
2	Time to complete the study	Minutes	50.6	8.1	65	35
3	Case materials were easy to understand	1=Never, 7=Always	5.1	1.2	7	3
4	Case materials were realistic	1=Never, 7=Always	4.7	1.4	6	2
5	Difficulties due to materials being in English	1=None, 7=Very Difficult	1.8	0.9	5	1
6	Understanding of framework	1=Poor, 7=Excellent	5.3	0.8	7	4
7	Effort needed to understand and complete materials	1=Easy, 7=Difficult	4.1	1.0	6	2

For question #3 to #7 a seven point scale anchored in the ends was used. Responses are from the 21 participants included in the judgment analysis.

Twenty-one participants were included in the judgment analysis. All of the participants were audit managers, and average work experience as an auditor was 7.3 years (#1 in table 10). An average of 51 minutes was spent on the entire experiment (#2 in table 10).

Participants on average reported the case materials to be easy to understand (5.1 on a 7 point scale; #3 in table 10) and realistic (4.7 on a 7 point scale: #4 in table 10). Providing the

⁵⁹ For example; when participants interpreted inherent risk as all transactions being erroneous prior to control, an appropriate judgment policy would be to require all locations to have sufficient control if control risk is to be below 100%. Such a judgment policy implies that all cues in the cases that were intended to be independent, amplifying, compensating and multi-step were perceived as completely-dependent (i.e., since all control cues are required to be effective for control to be sufficient).

instructions in English did not cause difficulties (1.8 on a 7 point scale: #5 in table 10). Participants generally assessed their understanding of the controller interrelationship framework as good (5.3 on a 7 point scale: #6 in table 10). When asked to circle any framework elements they found difficult to understand or apply, 11 participants reported difficulties with amplifying controls; 6 participants had difficulties with both understanding and applying the definition, 3 participants had difficulties understanding the definition (i.e., but did not report difficulties applying the definition), and 2 participants had difficulties applying the definition (i.e., but did not report difficulties understanding the definition). Furthermore, two participants reported difficulties applying the definition of multi-step controls, while one participant reported difficulties applying the definition of substitutable controls. Participants reported that, overall, understanding and completion of the case materials required some effort (4.1 on a 7 point scale: #7 in table 10).

Overall this indicates that the experimental task was not perceived as trivial. Furthermore, responses indicate that participants generally understood the materials and that the overall effort was not too difficult. It can, however, be noted that issues with understanding and applying the amplifying cue definition were reported by some participants.⁶⁰

Judgment logic in binary judgments

Since participants also made control risk judgments for each case, it is possible that the binary judgment was made based on the control risk percentage judgment, and not by reassessing cues. This would imply a different judgment policy (i.e., a simple accept/reject judgment about a percentage score) than the hypothesized conjunctive or disjunctive policy applied directly to the cues. The construct validity of the dependent variable operationalization (i.e., FFJP) would thus be threatened (i.e., the regression coefficients would not represent the underlying judgment policy).⁶¹

⁶⁰ No covariate analysis was performed for the following reason: The dependent variable in this study is the functional form of the judgment policy. This is only observed at the aggregate level for all participants. It is therefore not feasible to analyze the effect of individual level covariates on the aggregate policy.

⁶¹ As noted in the review of psychology literature this is an inherent weakness with policy capturing since one does not observe what goes on within the mind of the judge. The models from the regression of cue levels on judgments are only paramorphic (i.e., surface) representations of judgment policies (Hoffman, 1960; Ashton, 1982; Trotman, 1996).

In the post experimental questions, participants were therefore asked the following question:

It is of special interest to this study to understand your thought process when you were judging the “YES/NO” question of whether sufficient controls were in place. Please read both options below and think carefully about the way you responded to that question in the survey. Circle the option (below) that best describes your thought process. Both options lead to appropriate judgments, so please think back to how you actually made the judgments in your mind:

- (1) I simply looked at the control test results and considered whether there were sufficient controls in each location. That is, in your mind you were not just comparing the overall control risk score of the client to a threshold.*
- (2) I simply looked at the percentage score in the client’s overall control risk judgment and considered whether this was above or below my threshold. That is, in your mind you were not thinking in terms of sufficiency of controls in locations, but rather in terms of percentage scores. Please indicate the control risk threshold you applied: _____%*

From a normative perspective, both strategies should result in identical judgments. It is, however, more efficient to use a conjunctive (disjunctive) judgment strategy since one only needs to find enough deficient (efficient) controls to judge controls insufficient (sufficient) (e.g., if all cues are perceived as necessary (individually sufficient), then it is sufficient to find one deficient (effective) control in order to make a negative (positive) judgment).

Responses revealed that 14 auditors (i.e., 67%) made their judgments based on the cues (option 1 above), while 7 auditors (i.e., 33%) made their judgments based on the percentage score (option 2 above). If it is assumed that the auditor chooses one of these judgment processes by random (i.e., 50% chance for each), then the likelihood of 14 out of 21 auditors basing their judgments on the cues directly being due to randomness is <0.039 . Statistically, it is therefore supported at the 5% level that auditors, on average, use option 1 deliberately (i.e. they judge cues directly).

The six auditors basing their judgments on percentage scores had thresholds between 0% and 15%, while one auditor reported a threshold of 25%. For the majority of the auditors (67%), the description of their judgment policy supports the validity of the dependent variable construct operationalization.

6.3 Test of Hypotheses and Discussion

The hypotheses for the continuous judgment response scale (i.e., the “a” hypotheses) are tested by comparing the predicted coefficients with the results from the regression of cue values on participant judgments. Since the binary logistic regression does not provide usable results, the hypotheses for the binary judgment response scale (i.e., the “b” hypotheses) are not tested at the aggregate level. Instead, judgment policy details are reported at an individual level, and these provide a clear picture of the applied judgment policies. Table 11 below repeats the hypotheses and predictions:

Table 11: Overview of Hypotheses and Predicted Coefficients in Three-Cue Judgment Tasks:

Control Interrelationships	Multi-Step	Amplifying	Independent	Compensating	Substitutable
Predictions Continuous Response Scale	H1a $\alpha_7 < 0$	H5a $\alpha_1, \alpha_2, \alpha_3 < 0$ $\alpha_6 < 0$	H3a $\alpha_1, \alpha_2, \alpha_3 < 0$	H4a $\alpha_1, \alpha_2, \alpha_3 < 0$ $\alpha_5 > 0$	H2a $\alpha_1, \alpha_2, \alpha_3 < 0$ $\alpha_1 = \alpha_2 = -\alpha_4$
Predictions Binary Response Scale	H1b α_7	H5b α_7	H3b α_7	H4b α_7	H2b α_5, α_6

Note:

Each cell in the table regards a specific control interrelationship treatment (see column headings) and contains the applicable hypothesis and judgment model predictions for that treatment. Listed alphas are predicted to be significant when regressing judgments on cues in the given treatment. Unlisted alphas are predicted to be non-significant. For the continuous response scale, negative alphas imply that control risk is reduced. For the binary response scale, a listed coefficient implies that it is predicted to be a sufficient criterion for a “yes” judgment. For definitions of control interrelationships see section 3.5.3 and for response scales see section 3.5.5.

$\alpha_1, \alpha_2, \alpha_3$ = main effects of cues

$\alpha_4, \alpha_5, \alpha_6$ = two-way interactions between cues

α_7 = three-way interaction between cues

6.3.1 H1: Multi-Step (i.e., Completely Dependent) Controls

Continuous judgment response scale

H1a predicts that if judgments are about controls that are multi-step (i.e., cues that are completely dependent on one another), then the judgment policy will be conjunctive. Such a judgment policy is consistent with a judgment model where only the highest order interaction effect (i.e., α_7 : the ABC-interaction) is significant. An identical judgment policy is predicted for the binary response scale.

For the continuous response scale, the results show that the three-way interaction (i.e., α_7) is the only significant coefficient ($p < 0.01$) (see table 12 below). Findings for the continuous response scale are therefore consistent with the predictions in H1a.

Table 12: Results from Linear Regression H1a

H1a: Multi-step (completely-dependent) controls
Dependent variable: Control risk percentage judgment (n=21)

Cue	Coefficient	Prediction	Coefficients	Std. Error	t	Sig.
	Intercept		0.99	0.03	34.77	0.00 ***
A	α_1	0	-0.06	0.04	-1.56	0.12
B	α_2	0	-0.02	0.04	-0.53	0.60
C	α_3	0	-0.04	0.04	-0.95	0.35
AxB	α_4	0	-0.05	0.06	-0.85	0.40
AxC	α_5	0	0.01	0.06	0.24	0.81
BxC	α_6	0	0.04	0.06	0.71	0.48
AxBxC	α_7	(-)	-0.85	0.08	-10.55	0.00 ***

Model Summary:

R-sq=0.852; Adjusted R-sq=0.845; F(7, 167)=131.2, $p < 0.01$

*** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$

Note:

Cue A, Cue B and Cue C are completely-dependent

$\alpha_1, \alpha_2, \alpha_3 =$ main effects of cues

$\alpha_4, \alpha_5, \alpha_6 =$ two-way interactions between cues

$\alpha_7 =$ three-way interaction between cues

The regression coefficients can be interpreted as how much risk reduction they contribute to over and above other coefficients. $\alpha_7 = -0.85$ can be interpreted as risk is reduced by 85% if the three way interaction (i.e., AxBxC) is significant. In other words, risk is reduced by 85%

if all cues are effective. If less than all cues are effective, no risk reduction occurs. This is reasonable since no risk reduction should occur unless all cues are effective.

If the seven deleted participants are included in the analysis, the three-way interaction remains significant with an unchanged coefficient and p-value ($\alpha_7=-0.85$ and $p<0.01$). However the main effect of risk identification (α_1) becomes marginally significant even though the coefficient is unchanged ($\alpha_1=-0.06$, $p=0.09$). A closer examination of the individual judgment policies reveals that all seven deleted participants responded as hypothesized. The reason for α_1 becoming significant is that the standard error is reduced when seven participants without noise in responses are included in the analyses (reduced from 0.040 to 0.035). Further analyses of the 21 participants originally included in the analysis show that four of them reduced risk to some degree when cue A was effective, even though one or two of the other cues were deficient. The reason for the change in results is therefore not the seven deleted participants having different policies, but rather the effect on the standard error. Overall 24 (17) of the 28 (21) participants responded as hypothesized.

Binary Judgment Response Scale

H1b predicts that if judgments are about controls that are multi-step (i.e., cues that are completely dependent on one another), then the judgment policy will be conjunctive for both response scales. A conjunctive judgment policy is consistent with a judgment model where only the highest order interaction effect will be significant (i.e., α_7).

Since statistical software packages are unable to provide meaningful output from binary logistic regression on the data, the judgment policies will be described at an individual judge level through the procedure described in section 6.1.1. In essence, the procedure implies counting the number of participants having a positive response for the case with three effective controls, and negative responses in the seven other cases. Such a case response pattern is consistent with the hypothesized conjunctive judgment policy.

The detailed analysis of individual judgments showed that:

- 20 auditors (95%) judged control to be sufficient only if all cues were effective. This is consistent with hypothesis H1b.

- 1 auditor (5%) judged controls to be sufficient in the two cases where cue A and B were effective (i.e., regardless of the effectiveness of cue C).

Overall, 95% (20) of the auditors applied the appropriate configural judgment policy. The remaining 5% (1) of the auditors (i.e., only one auditor) applied an inappropriate judgment policy. The likelihood of observing 20 out of 21 participants using the hypothesized judgment policy versus any other judgment policy being due to randomness is <0.001 (i.e., the number of auditors using the appropriate judgment policy is significantly greater than 50% ($p < 0.001$) and thus not due to randomness). Results are therefore consistent with H1a: if judgments are about controls that are multi-step (i.e., cues that are completely dependent on one another), then the judgment policy will be conjunctive for both response scales.

The seven deleted participants responded as hypothesized, without exceptions. Their explanation of their judgment logic revealed that they had interpreted cues as intended (i.e., completely dependent). Their case response pattern was therefore as expected.

6.3.2 H2: Substitutable Controls

Continuous judgment response scale:

H2a predicts that if judgments are about controls that are substitutable, then the judgment policy will be disjunctive. An identical judgment policy is predicted for the binary response scale.

A disjunctive judgment policy in a task with a continuous response scale is consistent with a judgment model with significant main effects (i.e., α_1 , α_2 , α_3) and a significant interaction between substitutable cues (i.e., $A \times B$ or α_4) with the opposite sign of the main effect from the substitutable cues. Furthermore, the main effects of each of the substitutable cues (i.e., α_1 and α_2) are predicted to be equal since they by definition contribute equally to control risk reduction. The interaction term (i.e., α_4) serves to remove all but one of the main effects from the substitutable cues so that the net effect will equal one main effect (i.e., $\alpha_1 = \alpha_2 = -\alpha_4$), since by definition, adding additional substitutable cues does not contribute to control risk reduction.

For the continuous response scale, the results show that all three main effects (i.e., α_1 , α_2 , α_3) and the interaction between the substitutable cues (i.e., α_4) are significant at the $p < 0.01$ level

(see table 12 below). Furthermore, when 95% confidence intervals are created for cue coefficients $\alpha_1 = \alpha_2 = -\alpha_4$ ($\alpha_1 = -0.58$, $\alpha_2 = -0.60$ and $\alpha_4 = 0.52$, with confidence intervals of +/- 0.006, 0.006 and 0.008 respectively). For substitutable cues, findings for the continuous response scale are therefore consistent with the predictions in H2a.

Table 13: Results from Linear Regression H2a

H2a: Substitutable controls
Dependent variable: Control risk percentage judgment
(n=21)

Cue	Coefficient	Predictio	Coefficients	Std. Error	t	Sig.
	Intercept		0.99	0.02	46.40	0.00 ***
A	α_1	(-)	-0.58	0.03	-19.06	0.00 ***
B	α_2	(-)	-0.60	0.03	-20.02	0.00 ***
C	α_3	(-)	-0.31	0.03	-10.33	0.00 ***
AxB	α_4	(+)	0.52	0.04	12.27	0.00 ***
AxC	α_5	0	-0.04	0.04	-0.94	0.35
BxC	α_6	0	-0.02	0.04	-0.37	0.71
AxBxC	α_7	0	0.05	0.06	0.78	0.44

Model Summary

R-sq=0.917; Adjusted R-sq=0.913; F(7, 167)=251.3, p<0.01

***p<0.01, **p<0.05, *p<0.10

Note:

Cue A and cue B are substitutable, cue C is independent.

$\alpha_1, \alpha_2, \alpha_3 =$ main effects of cues

$\alpha_4, \alpha_5, \alpha_6 =$ two-way interactions between cues

$\alpha_7 =$ three-way interaction between cues

The regression coefficients can be interpreted as how much risk reduction they contribute to over and above other coefficients. For example, $\alpha_1 = -0.58$, $\alpha_2 = -0.60$ and $\alpha_4 = 0.52$. This can be interpreted as follows: If cue A is effective risk reduction is 58% ($\alpha_1 = -0.58$). This is the main effect of cue A. If cue B is effective, risk reduction is 60% ($\alpha_2 = -0.60$). This is the main effect of cue B. If both cue A and cue B are significant, both main effects occur, but the combined effect (i.e., interaction effect AxB) reduces the risk reduction by 52% ($\alpha_4 = 0.52$) so that overall risk is reduced by 58%+60%-52% = 66% (i.e., $\alpha_1 + \alpha_2 + \alpha_4 = 0.66$). Given that cue A and cue B each represent completely overlapping control over

approximately 66% of the hypothetical audit client's invoices, the risk reduction percentages seem reasonable.

If the seven deleted participants are included in the analysis it is to be expected that the two remaining two-way interactions (AxC-interaction and BxC-interaction) become significant. The reason for this is that these participants explained that they interpreted inherent risk to be that all locations had material errors prior to controls. An appropriate judgment policy would, given this interpretation of inherent risk, be to judge control risk as either 0% (when all locations had 100% of invoices controlled) or 100% (when any location had less than 100% of invoices controlled). 100% invoice control occurs when the independent cue (cue C) and at least one of the substitutable cues (cue A and/or cue B) are effective. When these participants are included in the analysis, results show that the hypothesized effects remain significant at the $p < 0.001$ level. In addition, the AC-interaction and BC-interaction become significant ($p = 0.01$). This is consistent with the way the seven excluded participants explained that they had interpreted cue interrelationships. In fact, it is consistent with the binary judgment logic in H2b.

Binary judgment response scale

H2b predicts that if the judgment response scale is binary and if controls are substitutable, then the judgment policy will be disjunctive. In a disjunctive judgment policy it is necessary and sufficient if at least one of the substitutable controls is effective. In the three-cue judgment task in this study, it is also necessary that the independent cue is operationally effective. A disjunctive judgment policy will therefore be consistent with a judgment model with significant two-way interactions between the independent cue and each of the substitutable cues (i.e., α_5 and α_6).

Since statistical software packages are unable to provide meaningful output from binary logistic regression on the data, the judgment policies will be described at an individual judge level through the procedure described in section 6.1.1. In essence, the procedure implies counting the number of participants having a positive response for the three cases where the independent control and one or both of the substitutable controls are effective and negative responses in the five other cases. Such a case response pattern is consistent with the hypothesized disjunctive judgment policy.

The detailed analysis of individual judgments showed that:

- 18 auditors (85%) judged controls to be sufficient if the independent cue and at least one of the substitutable cues were effective. This is consistent with H2b.
- 2 auditors (10%) judged controls to be sufficient only if all cues were effective.
- 1 auditor (5%) judged controls to be sufficient as long as any two cues are effective.

Overall, 85% (18) of the auditors applied the appropriate configural judgment policy. The remaining 15% (3) of the auditors applied inappropriate judgment policies. The likelihood of observing 18 out of 21 participants using the hypothesized judgment policy versus any other judgment policy being due to randomness is <0.001 (i.e., the number of auditors using the appropriate judgment policy is significantly greater than 50% ($p < 0.001$) and thus not due to randomness). Results are therefore consistent with H2b: if the judgment response scale is binary and if controls are substitutable, then the judgment policy will be disjunctive.

The seven deleted participants that were deleted from analysis had the following judgment policies:

- 5 auditors judged controls to be sufficient if the independent cue and at least one of the substitutable cues were effective. This is consistent with H2
- 2 auditors judged controls to be sufficient only if all cues were effective.

If the deleted participants had been included in the analysis results would not change: the likelihood of 23 out of 28 policies being as predicted due to randomness has a probability of <0.001 .

6.3.3 H3: Independent Controls

Continuous judgment response scale

H3a predicts that if judgment response scales are continuous (i.e., regarding control risk), and if controls are independent, then the judgment policy will be linear. A linear judgment policy is consistent with a judgment model with significant main effects (i.e., α_1 , α_2 , α_3) and no interactions.

The results show that all three main effects (i.e., α_1 , α_2 , α_3) are significant at the $p < 0.01$ level. No interactions are found to be significant (see table 14 below). Results are therefore consistent with H3a: if judgment response scales are continuous (i.e., regarding control risk), and if controls are independent, then the judgment policy will be linear.

Table 14: Results from Linear Regression H3a

H3a: Independent controls
Dependent variable: Control risk percentage judgment
(n=21)

Cue	Coefficient	Prediction	Coefficients	Std. Error	t	Sig.
	Intercept		0.99	0.02	45.56	0.00 ***
A	α_1	(-)	-0.33	0.03	-10.75	0.00 ***
B	α_2	(-)	-0.35	0.03	-11.39	0.00 ***
C	α_3	(-)	-0.33	0.03	-10.77	0.00 ***
AxB	α_4	0	0.04	0.04	0.86	0.39
AxC	α_5	0	0.04	0.04	0.86	0.39
BxC	α_6	0	0.05	0.04	1.09	0.28
AxBxC	α_7	0	-0.08	0.06	-1.26	0.21

Model Summary

R-sq=0.888; Adjusted R-sq=0.883; F(7, 167)=181.5, $p < 0.01$

*** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$

Note:

Cue A, cue B and cue C are independent

$\alpha_1, \alpha_2, \alpha_3 =$ main effects of cues

$\alpha_4, \alpha_5, \alpha_6 =$ two-way interactions between cues

$\alpha_7 =$ three-way interaction between cues

The regression coefficients can be interpreted as how much risk reduction they contribute to over and above other coefficients. For example, $\alpha_1 = -0.33$, $\alpha_2 = -0.35$ and $\alpha_3 = -0.33$. This can be interpreted as follows: If cue A is effective risk reduction is 33% ($\alpha_1 = -0.33$). This is the main effect of cue A. If cue B is effective, risk reduction is 35% ($\alpha_2 = -0.35$). This is the main effect of cue B. If cue C is effective risk reduction is 33% ($\alpha_3 = -0.33$). This is the main effect of cue C. Since no interaction terms are significant, this implies that when more than one cue is effective, their main effects can be added together to find the overall risk reduction. Given that each cue represents control over approximately 33% of the hypothetical audit client's invoices, the risk reduction percentages seem reasonable.

If the seven deleted participants are included in the analysis it is to be expected that the three-way interaction (ABC) becomes significant. The reason for this is that these participants explained that they interpreted inherent risk to be that all locations had material errors prior to controls. An appropriate judgment policy would, given this interpretation of inherent risk, be to judge control risk as either 0% (when all locations had 100% of invoices controlled) or 100% (when any location had less than 100% of invoices controlled). 100% invoice control occurs only if all three cues are effective. When these participants are included in the analysis, results show that the hypothesized effects remain significant at the $p < 0.001$ level. In addition, the ABC-interaction becomes significant ($p = 0.01$). This is consistent with the way the seven excluded participants explained that they had interpreted cue interrelationships. In fact, it is consistent with the binary judgment logic in H3b.

Binary judgment response scale

H3b predicts that if judgments are about controls that are independent and if judgment response scales are binary, then the judgment policy will be conjunctive. A conjunctive judgment policy is consistent with a judgment model where only the highest order interaction effect will be significant (i.e., α_7).

Since statistical software packages are unable to provide meaningful output from binary logistic regression on the data, the judgment policies will be described at an individual judge level through the procedure described in section 6.1.1. In essence, the procedure implies counting the number of participants having a positive response for the case with three effective controls, and negative responses in the seven other cases. Such a case response pattern is consistent with the hypothesized conjunctive judgment policy.

The detailed analysis of individual judgments showed that:

- 18 auditors (85%) judged controls to be sufficient only if all cues were effective. This is a conjunctive judgment policy, consistent with H3b.
- 2 auditors (10%) judged controls to be sufficient if any two cues were effective.
- 1 auditor (5%) judged controls to be insufficient in all cases.

Overall, 85% (18) of the auditors applied the appropriate configural judgment policy, while the remaining 15% (3) of the auditors applied inappropriate judgment policies. The

likelihood of observing 18 out of 21 participants using the hypothesized judgment policy versus any other judgment policy being due to randomness is <0.001 (i.e., the number of auditors using the appropriate judgment policy is significantly greater than 50% ($p < 0.001$) and thus not due to randomness). The findings are therefore consistent with H3b: If judgment response scales are binary, and if controls are independent, then the judgment policy will be conjunctive.

The seven deleted participants responded as hypothesized, without exceptions. Their explanation of their judgment logic revealed that they had interpreted cues as intended (i.e., completely dependent). Their case response pattern was therefore as expected.

6.3.4 H4: Compensating Controls

Continuous judgment response scale

H4a predicts that if judgment response scales are continuous (i.e., regarding control risk), and if controls are compensating, then the judgment policy will be compensatory, additive and nonlinear with a compensatory form ordinal interaction. Such a judgment policy is consistent with a judgment model with significant main effects (i.e., α_1 , α_2 , and α_3) and a significant, two-way interaction between compensating cues (i.e., α_5) of the opposite sign from main effects.

Results show that all three main effects (i.e., α_1 , α_2 , α_3) and the predicted interaction (i.e., α_5) are significant at the $p < 0.01$ level. Furthermore, as predicted, the interaction is of the opposite sign from main effects (i.e., the interaction is compensatory form ordinal). Findings are therefore consistent with the predictions in H4a (see table 15 below).

Table 15: Results from Linear Regression H4a

H4a: Compensating controls**Dependent variable: Control risk percentage judgment (n=21)**

Cue	Coefficient	Prediction	Coefficients	Std. Error	t	Sig.
	Intercept		0.98	0.02	40.28	0.00 ***
A	α_1	(-)	-0.45	0.03	-13.04	0.00 ***
B	α_2	(-)	-0.31	0.03	-8.96	0.00 ***
C	α_3	(-)	-0.46	0.03	-13.32	0.00 ***
AxB	α_4	0	-0.01	0.05	-0.15	0.88
AxC	α_5	(+)	0.26	0.05	5.43	0.00 ***
BxC	α_6	0	0.04	0.05	0.87	0.39
AxBxC	α_7	0	-0.05	0.07	-0.68	0.50

Model Summary

R-sq=0.870; Adjusted R-sq=0.864; F(7, 167)=152.9, p<0.01

***p<0.01, **p<0.05, *p<0.10

*Note:**Cue A and cue C are compensating, cue B is independent* $\alpha_1, \alpha_2, \alpha_3$ = main effects of cues $\alpha_4, \alpha_5, \alpha_6$ = two-way interactions between cues α_7 = three-way interaction between cues

The regression coefficients can be interpreted as how much risk reduction they contribute to over and above other coefficients. For example, $\alpha_1 = -0.45$, $\alpha_3 = -0.46$ and $\alpha_5 = 0.26$. This can be interpreted as follows: If cue A is effective risk reduction is 45% ($\alpha_1 = -0.45$). This is the main effect of cue A. If cue C is effective, risk reduction is 46% ($\alpha_3 = -0.46$). This is the main effect of cue C. If both cue A and cue C are effective, both main effects occur, but the combined effect (i.e., interaction effect AxC) reduces the risk reduction by 26% ($\alpha_5 = 0.26$) so that overall risk is reduced by 45%+46%-26% = 65% (i.e., $\alpha_1 + \alpha_3 + \alpha_5 = 0.65$). Given that cue A and cue C each represent control over half of the hypothetical audit client's invoices, the risk reduction percentages for main effects seem reasonable. The interaction effect is furthermore reasonable since, if both cue A and cue C are effective, 66% of invoices are controlled.

If the seven participants that are excluded from analysis are included, the findings do not change. Only hypothesized effects remain significant and signs do not change.

Binary judgment response scale

H4b predicts that if judgments are about controls that are compensating and if judgment response scales are binary, then the judgment policy will be conjunctive. A conjunctive judgment policy is consistent with a judgment model where only the highest order interaction effect will be significant (i.e., α_7).

Since statistical software packages are unable to provide meaningful output from binary logistic regression on the data, the judgment policies will be described at an individual judge level through the procedure described in section 6.1.1. In essence, the procedure implies counting the number of participants having a positive response for the case with three effective controls, and negative responses in the seven other cases. Such a case response pattern is consistent with the hypothesized conjunctive judgment policy.

The detailed analysis of individual judgments showed that:

- 13 auditors (62%) judged controls to be sufficient only if all cues were effective. This is consistent with H4b.
- 7 auditors (33%) judged controls to be sufficient in the three cases where cue B (i.e., the independent control) and at least one of either cue A or C (i.e., compensating controls) was effective. This judgment policy is appropriate if cue A and C are considered substitutable. See further discussion below.
- 1 auditor (5%) judged controls to be sufficient if any two cues were effective. This implies judging four out of the eight cases positively.

Overall, 62% of the auditors applied the appropriate configural judgment policy, while the remaining 38% of the auditors applied inappropriate judgment policies. The likelihood of observing 13 out of 21 participants using the hypothesized judgment policy versus any other judgment policy being due to randomness is <0.094 . The results for binary judgments about compensating controls are therefore marginally supportive of H4b.

The seven deleted participants responded as hypothesized, without exceptions. Their explanation of their judgment logic revealed that they had interpreted cues as intended (i.e., completely dependent). Their case response pattern was therefore as expected. If they had been included in the analysis, the likelihood of observing 20 out of 28 participants using the

hypothesized judgment policy versus any other judgment policy being due to randomness would have been <0.006 , and H4b would have been supported.

Further discussion: Responses from the 21 participants included in the main analysis indicate that seven auditors (33%), for the binary judgments, responded to cases in the manner that was expected when controls were manipulated to be individually sufficient (i.e., as in the substitutable series); controls were judged to be sufficient as long as the independent control (i.e., cue B) and one of the compensating controls (i.e., cue A or cue C) were effective. This raises the question of whether cue A and cue C were actually perceived as individually sufficient (i.e., substitutable) in the binary judgment, when they were manipulated to be perceived as individually insufficient.

One potential explanation for compensating controls being perceived as substitutable in a binary judgment may be found in audit regulation/practice: Internal controls only provide reasonable assurance, not absolute assurance (AS5.A7 PCAOB 2007). Controls that provide individually sufficient control risk reduction, given the reasonable assurance criterion, may therefore be considered as substitutable cues in a binary judgment task, even though they do not provide equal control risk reduction (i.e., they are not fully substitutable controls, but only sufficiently compensating controls). An example could be controls with a sufficiently high degree of compensation for both to be individually above the reasonable assurance criterion. In such a situation compensating controls would result in substitutable cues.

There are three ways to provide insight into the judgments of the seven participants under discussion: First, did they misclassify cues and/or report difficulties with understanding or applying the interrelationship framework? Second, how did they judge control risk? Were control risk judgments reasonable? Third, if they in the post experimental questions reported that the binary judgment was based on the percentage score in the control risk judgment, what cut-off did they report applying?

In the cue classification exercise all of these seven participants, and 20 out of the 21 participants, classified the cues correctly as compensating. In the post experimental questions, no participants reported issues with understanding or application of the compensating cues definition.

An inspection of the seven participant's individual control risk judgments shows that they judged control risk in the following manner:

- One participant judged control risk to be 0% in the three cases where cue B and cue A and/or cue C was effective. Although classifying cues as compensating, this person clearly made judgments as if cues were perceived to be substitutable.
- One participant judged control risk to be 10% in the two cases where cue B and cue A or cue C was effective.
- Five participants judged control risk to be between 15% and 17% in the two cases where cue B and cue A or cue C was effective.

This indicates that one of the seven participants perceived cue A and C to be substitutable, while the remaining six participants made reasonable control risk judgments in a manner consistent with perceiving cues as compensating.⁶² It may therefore be argued that the participant with the 0% judgments be deleted from analysis due to failing to absorb the cue interrelationship manipulation. If deleted, the likelihood of observing 13 out of 20 participants using the hypothesized judgment policy versus any other judgment policy being due to randomness would be <0.058 , and the conclusion of H3b being marginally supported would not be changed.

Of the six remaining participants, three reported that the binary judgment was based on the percentage score in the control risk judgment. Two of these participants reported using a 15% cut-off, while one reported using a 25% cutoff. In the two cases where cue B and either cue A or cue C was effective, the two participants reporting a 15% cutoff judged control risk to be 16% and 17%, while the participant reporting a 25% cutoff judged control risk to be 15%. Only the participant reporting applying a 25% cutoff was thus consistent between his reported cutoff, the control risk judgment and the binary judgment. His judgment policy implied that cue A and cue C were substitutable for the purpose of reducing control risk below his cutoff, and he did thus not perceive cues as intended in the binary judgment. The two others judged controls to be sufficient even though control risk was judged to be 16%

⁶² With the independent control effective, 33% of the invoices would be controlled. With one compensating control effective, an additional 50% of invoices would be controlled. In the cases with the independent control and one compensating control effective, it is therefore reasonable to judge control risk to be in approximately 17%.

and 17%, thus over their reported cutoff of 15%. It is therefore reasonable to assume that they had noise in their judgment policy due reasons like e.g., fatigue.

The remaining three participants reported basing the binary judgment on the cues directly, and did thus not report a cutoff. It is reasonable to assume that these participants either had a higher tolerance for risk, and thus perceived cue A and cue B as substitutable in the binary judgment, or that they just had noise in their judgments due to reasons like e.g., fatigue. However, it can not be ruled out that participants perceived cue interrelationships correctly in the binary judgments (i.e., that all cues were intended to be necessary conditions for a positive judgment), but just applied an inappropriate judgment policy.

Overall, the binary judgment responses from the seven participants indicate a potential systematic problem with their cue interrelationship perceptions. If they had been deleted from analysis, the probability of observing 13 out of 14 judgment policies being appropriate, due to randomness, would have been <0.001 , and H4b would have been supported.

6.3.5 H5: Amplifying Controls

Continuous judgment response scale:

H5a predicts that if judgment response scales are continuous (i.e., regarding control risk), and if controls are amplifying, then the judgment model will be compensatory, additive and nonlinear with an amplifying form ordinal interaction. Such a judgment policy is consistent with a judgment model with significant main effects (i.e., α_1 , α_2 , α_3) and a significant two-way interaction between amplifying cues (i.e., α_6) of the same sign as the main effects.

Results show that all three main effects (i.e., α_1 , α_2 , α_3) are significant at the $p < 0.01$ level.⁶³ The main effect coefficients can be interpreted as how much risk reduction each cue contributes with over and above other cues (e.g., the effect of cue A being efficient is to reduce risk with 33% ($\alpha_1=33\%$)). However, the predicted interaction is not significant ($p=0.45$) (see table 16 below). No interaction effects therefore occur. Findings are therefore

⁶³ One of the participants received an empty envelope for the amplifying series. The analysis of amplifying cues is therefore based on 20 participants. The empty envelope is due to manual error in sorting and filling 216 envelopes.

not consistent with the predictions in H5a. Instead, participants use a linear (i.e., nonconfigural) judgment model for this series of cases (i.e., with only main effects).

Table 16: Results from Linear Regression H5a

H5a: Amplifying controls

Dependent variable: Control risk percentage judgment (n=20)

Cue	Coefficient	Prediction	Coefficients	Std. Error	t	Sig.
	Intercept		0.98	0.03	34.89	0.00 ***
A	α_1	(-)	-0.33	0.04	-8.40	0.00 ***
B	α_2	(-)	-0.24	0.04	-5.94	0.00 ***
C	α_3	(-)	-0.35	0.04	-8.77	0.00 ***
AxB	α_4	0	0.06	0.06	1.05	0.30
AxC	α_5	0	0.02	0.06	0.28	0.78
BxC	α_6	(-)	-0.04	0.06	-0.76	0.45
AxBxC	α_7	0	-0.08	0.08	-0.96	0.34

Model Summary
R-sq=0.837; Adjusted R-sq=0.830; F(7, 167)=111,3,
p<0.01

***p<0.01, **p<0.05, *p<0.10

Note:

Cue B and Cue C are amplifying, Cue A is independent.

$\alpha_1, \alpha_2, \alpha_3$ = main effects of cues

$\alpha_4, \alpha_5, \alpha_6$ = two-way interactions between cues

α_7 = three-way interaction between cues

Two potential explanations arise: (1) cues were not perceived as amplifying, or (2) participants perceived cues as intended, but did not combine cues in the predicted manner.

The post experimental questions revealed that when asked to circle any framework elements found difficult to understand or apply, 11 (i.e., out of 20) participants reported difficulties with amplifying controls; 6 participants had difficulties with both understanding and applying the definition, 3 participants had difficulties understanding the definition (i.e., but did not report difficulties applying the definition), and 2 participants had difficulties applying the definition (i.e., but did not report difficulties understanding the definition).

Furthermore, the classification of cues in the amplifying control series reveals that:

- All 20 participants classified the independent cue correctly (i.e., cue A)

-
- 14 participants classified the first amplifying cue correctly (i.e., cue B)
 - 12 participants classified the second amplifying cue correctly (i.e., cue C)

The cue classification also revealed that:

- 10 participants (i.e., 50%) classified all cues correctly
- 4 participants (i.e., 20%) classified both amplifying cues wrong (i.e., cue B and C)
- 2 participants (i.e., 10%) classified amplifying cue B wrong, but cue C correctly
- 4 participants (i.e., 20%) classified amplifying cue C wrong, but cue B correctly

Furthermore, it can be noted that no systematic classification error occurred:

- Of the 6 participants that classified cue B as other than amplifying, 3 classified it as compensating, 1 as independent, and 2 as multi-step
- Of the 8 participants that classified cue C as other than amplifying, 2 classified it as compensating, 5 as independent, and 1 as multi-step

The participant's judgments, the cue classification exercise and the post experimental questions therefore indicate that a potential problem may exist with participant perception of amplifying cue interrelationships. This is a potential reason for the observation of a linear judgment policy even though cues were intended to be amplifying.

However, since 11 participants (53%) indicated difficulties with understanding and/or applying the amplifying cue definition, the cue classification exercise cannot be considered a precise measure (i.e., participants may have perceived the cue interrelationships correctly, but due to difficulties with the definition of amplifying cues, classification according to the definitions was noisy). It can therefore not be ruled out that participants perceived cues correctly even though 16 participants (i.e., 80%) reported either difficulty with the amplifying cue definition and/or classified at least one amplifying cue incorrectly. If participants perceived cues correctly, then the reason for the observed judgment policy (i.e., linear cue combination) is that the participating auditor's were not capable of the predicted form of configural cue processing.

If the seven deleted participants are included in the analysis it is to be expected that the three-way interaction (ABC) becomes significant. The reason for this is that these participants explained that they interpreted inherent risk to be that all locations had material errors prior to controls. An appropriate judgment policy would, given this interpretation of inherent risk, be to judge control risk as either 0% (when all locations had 100% of invoices controlled) or 100% (when any location had less than 100% of invoices controlled). 100% invoice control occurs only if all three cues are effective. When these participants are included in the analysis, results show that the ABC-interaction becomes significant ($p=0.02$). In addition, the three hypothesized main effects remain significant at the $p<0.001$ level, but the hypothesized two-way BC-interaction remains insignificant with $p=0.7$. This is also to be expected since the cue interrelationship perception of the seven additional participants did not make it appropriate to weight the BC-interaction in their judgment policy.

Binary judgment response scale

H5b predicts that if judgments are about controls that are amplifying and if judgment response scales are binary, then the judgment policy will be conjunctive. A conjunctive judgment policy is consistent with a judgment model where only the highest order interaction effect will be significant (i.e., α_7).

Since statistical software packages are unable to provide meaningful output from binary logistic regression on the data, the judgment policies will be described at an individual judge level through the procedure described in section 6.1.1. In essence, the procedure implies counting the number of participants having a positive response for the case with three effective controls, and negative responses in the seven other cases. Such a case response pattern is consistent with the hypothesized conjunctive judgment policy.

The detailed analysis of individual judgments showed that:

- 15 auditors (75%) judged controls to be sufficient only if all cues were effective. This is consistent with H5b.
- 3 auditors (15%) judged controls to be sufficient as long as cue A and C were effective.
- 1 auditor (5%) judged controls to be sufficient as long as cue B and C were effective.

- 1 auditor (5%) judged controls to be insufficient in all cases. This is an illogical policy.

Overall, 75% (15) of the auditors applied the appropriate configural judgment policy, while the remaining 25% (5) of the auditors applied inappropriate judgment policies. The likelihood of observing 15 out of 20 participants using the hypothesized judgment policy versus any other judgment policy being due to randomness is <0.006 (i.e., the number of auditors using the appropriate judgment policy is significantly greater than 50% ($p<0.006$) and thus not due to randomness). The findings are therefore generally consistent with H5b: If judgment response scales are binary, and if controls are independent, then the judgment policy will be conjunctive.

The seven deleted participants responded as hypothesized, without exceptions. Their explanation of their judgment logic revealed that they had interpreted cues as intended (i.e., completely dependent). Their case response pattern was therefore as expected.

6.4 Summary of Findings

Results of hypotheses tests are tabled below (see table 16). Apart from H5a, amplifying controls and continuous judgment response scale, which was clearly not supported, and H4b, compensating controls and binary judgment response scale, which was marginally supported, the remaining eight hypotheses were supported at the 5% significance level. Findings are further summarized in section 7.1 under RQ7.

Table 16: Summary of Results from Tests of Hypotheses:

Task Characteristics	Multi-Step	Amplifying	Independent	Compensating	Substitutable
Continuous Response Scale	H1a Supported	H5a Not supported	H3a Supported	H4a Supported	H2a Supported
Binary Response Scale	H1b Supported	H5b Supported	H3b Supported	H4b Marginally Supported	H2b Supported

7. Discussion and Implications

7.1 Answers to Research Questions

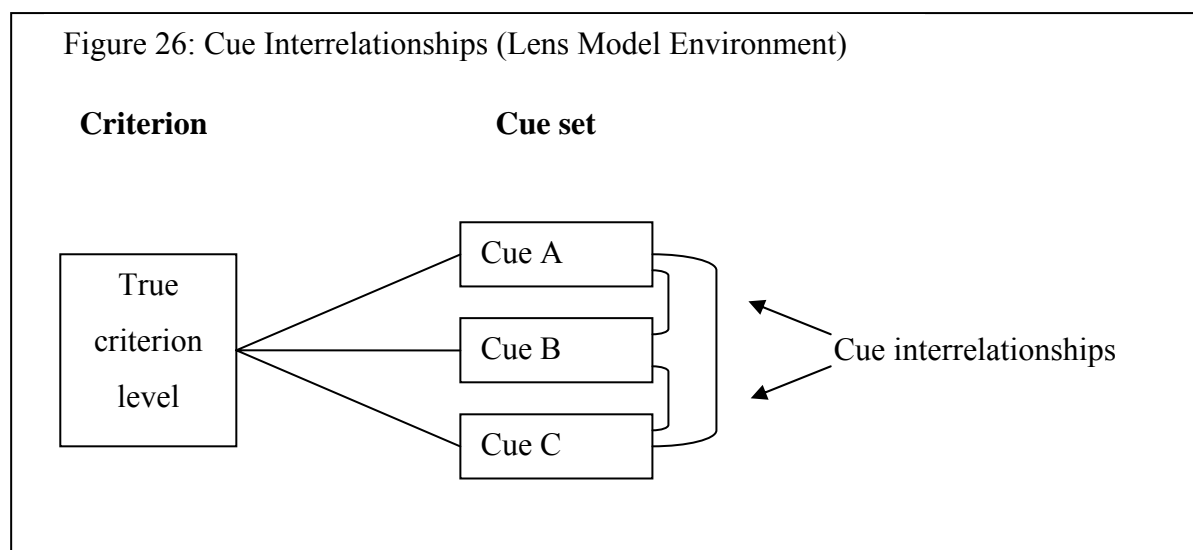
The following discussion summarizes the proposed answers to the research questions posed in the introductory section of this dissertation.

RQ1: What is the difference between control interrelationships and cue interrelationships?

This dissertation proposes that control interrelationships do not always translate directly to cue interrelationships and that, in fact, they are two different concepts.

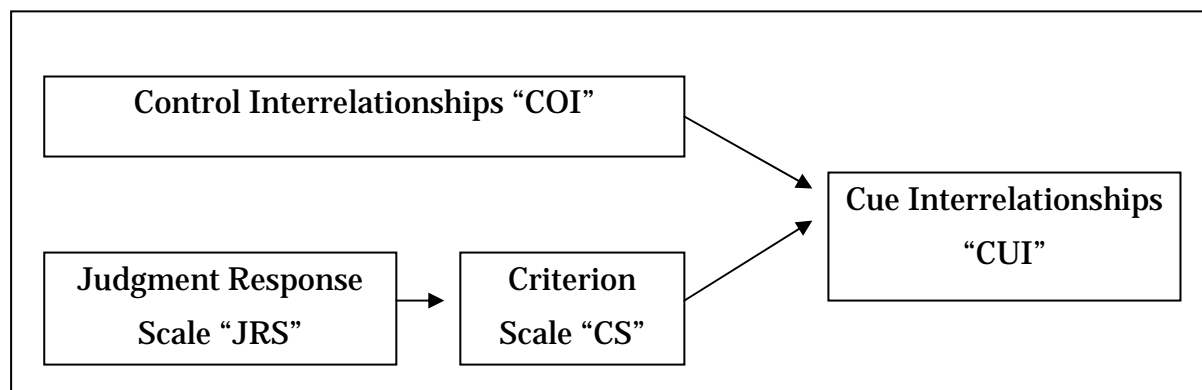
Control interrelationships are described in audit regulation and can be defined by how controls interact in their relation to risk (see discussion under section 3.5.2). Since, in auditing, risk is conceptualized through the audit risk model, it is furthermore reasonable to state that control interrelationship concepts are based on some form of continuous risk scale – typically a percentage point risk scale.

Cue interrelationships, however, are more generic in that they can be defined by how cues interact in their relation to a criterion (i.e., the state of an external reality) (see figure 26 below) (Libby 1981, p8, item B3). While control interrelationships are tied to a risk scale, the general construct of cue interrelationships may be tied to any kind of scale.



A sufficient argument for control interrelationships and cue interrelationships being two different concepts is that a change in the judgment response scale may change cue interrelationships even if controls, and thus control interrelationships, remain the same (see discussion under section 3.5.2). For example, if controls are independent and the judgment response scale is continuous, then cues are independent. If controls remain unchanged, but the judgment response scale changes to binary, then the cues will be completely-dependent, since all cues are necessary for a positive judgment (see discussion under section 3.5.2 and discussion of RQ2 and RQ3 below). In general, when controls serve as cues, a change in the judgment response scale may cause cue interrelationships to change (i.e., even though the cues (controls) remain constant). Control interrelationships are therefore one of the determinants of cue interrelationships, and they are therefore two different concepts (see figure 27 below and further discussion under the response to RQ6 below):

Figure 27: Model of Determinants of Cue Interrelationships

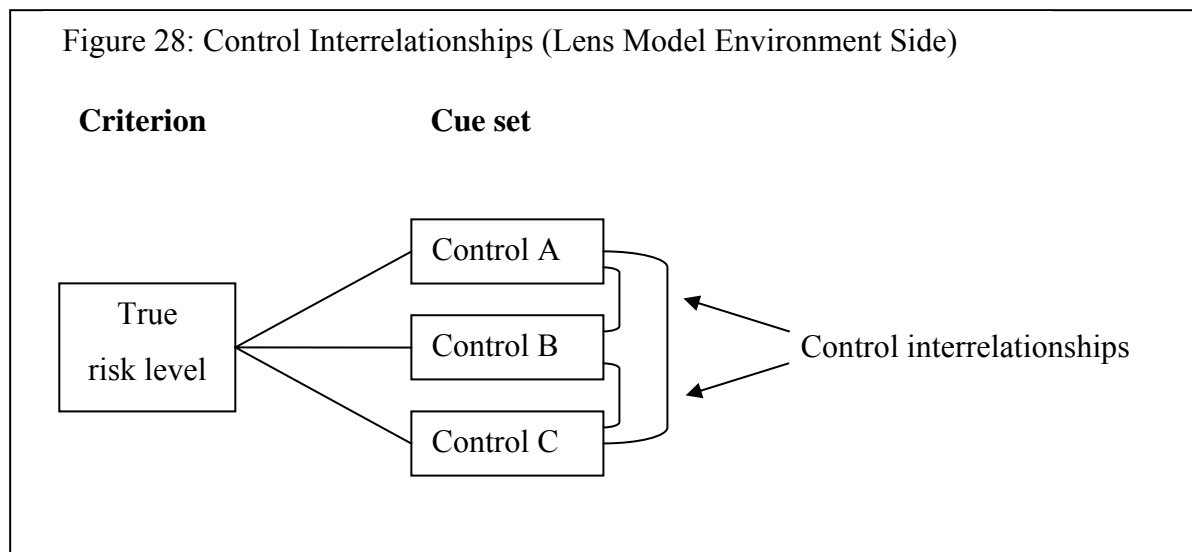


Note that it is not suggested that controls and cues are two different constructs; in audit judgment research, controls may serve as operationalizations of cues. However, the interrelationships are two different constructs, and it is the interrelationships that are the independent variables in this dissertation.

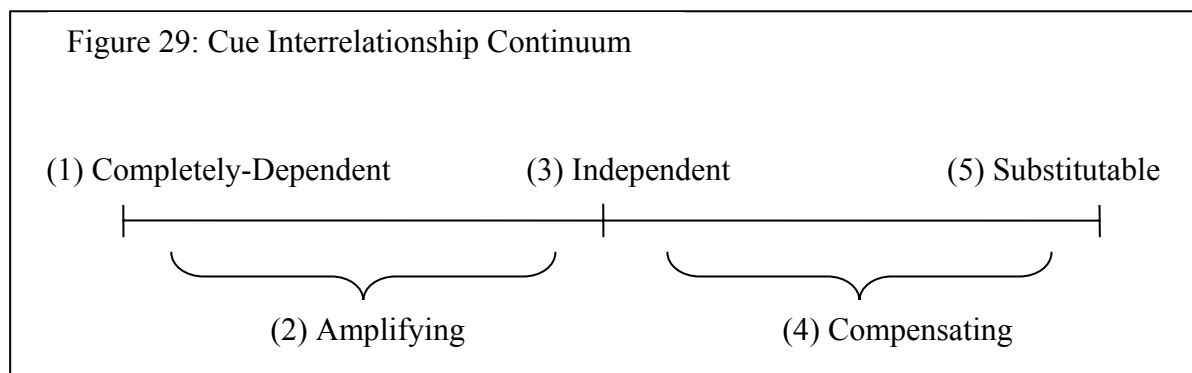
RQ2: What is the nature and range of variation in the control interrelationship variable?

In auditing, risk is conceptualized through the audit risk model. It is therefore reasonable to state that control interrelationship concepts are based on some form of continuous scale – typically a percentage point risk scale. This dissertation therefore proposes a framework

where control interrelationships are a function of how controls interact in their relationship to risk (see figure 28 below).



The range of variation in control interrelationships spans from completely-dependent to substitutable (see figure 29 below).



The range of variation in control interrelationships has five sections:

1. Completely dependent controls, where a control only has an effect if all other controls are present (i.e., all controls must be present if any effect is to occur; all controls are necessary)
2. Amplifying controls, where the effect of two controls combined is larger than the sum of the two individual control effects
3. Independent controls, where each control's effect is independent of the level (i.e., presence/absence) of other controls

4. Compensating controls, where the effect of two controls combined is smaller than the sum of the two individual control effects
5. Substitutable controls, where controls have the same individual effect and where the effect of two controls combined is the same as the individual contribution of one control (i.e., controls can substitute for each other, but they do not add incremental effect if another control is already present)

The control interrelationship continuum can furthermore be described mathematically by use of functions. Assume that “y” represents the risk criterion level and that “y” is a function of control cues “c_i”:

Let $y = f(c_i)$, where there are two cues ($i=1,2$) and cues take on the values 0 or 1 ($c_i = 0,1$)

and

$$0 = 0/100 \leq f(c_1, c_2) \leq 100/100 = 1$$

$f(0,0) = 0$ for all functions

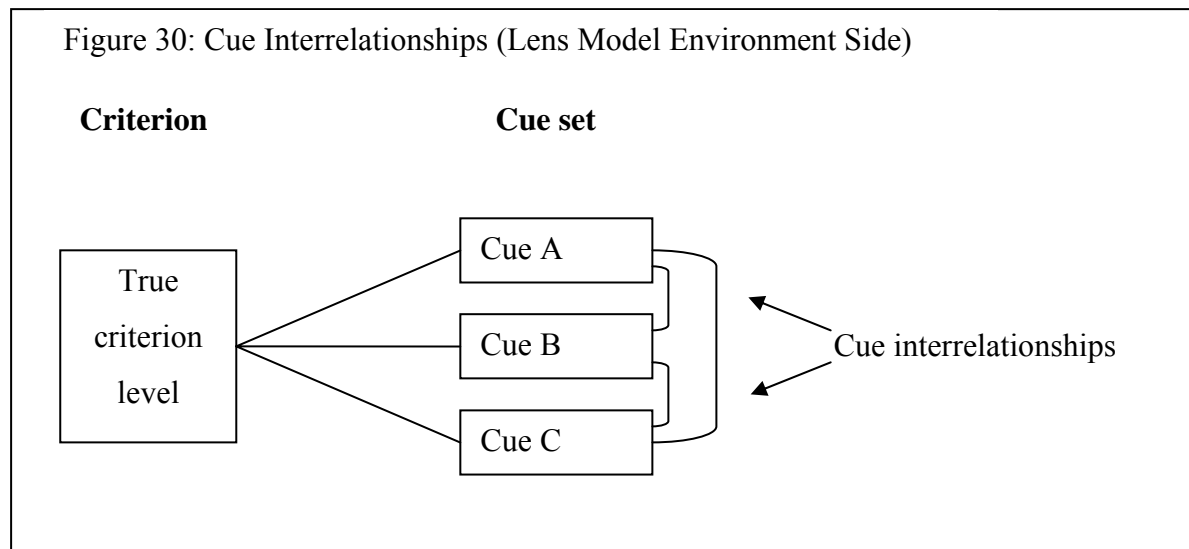
This provides the following control interrelationship functions (see table 17 below):

Table 18: Mathematical Representation of Control Interrelationship Functions

Control interrelationship	Mathematical Function	Description
1. Completely-dependent	$f(1,0) = f(0,1) = 0$, and $f(1,1) > 0$	Complete positive interrelationship
2. Amplifying	$0 < f(0,1) + f(1,0) < f(1,1)$ $0 < f(1,0), 0 < f(0,1)$	Partial positive interrelationship
3. Independent	$0 < f(0,1) + f(1,0) = f(1,1)$ $0 < f(1,0), 0 < f(0,1)$	No interrelationship
4. Compensating	$0 < f(1,1) < f(0,1) + f(1,0)$ $0 < f(1,0), 0 < f(0,1)$	Partial negative interrelationship
5. Substitutable	$0 < f(1,0) = f(0,1) = f(1,1)$	Complete negative interrelationship

RQ3: What is the nature and range of variation in the cue interrelationship variable?

This dissertation proposes a framework where cue interrelationships are a function of how cues interact in their relationship to a general criterion. While control interrelationships are tied to a risk scale, the general construct of cue interrelationships may be tied to any kind of scale (see figure 30 below).

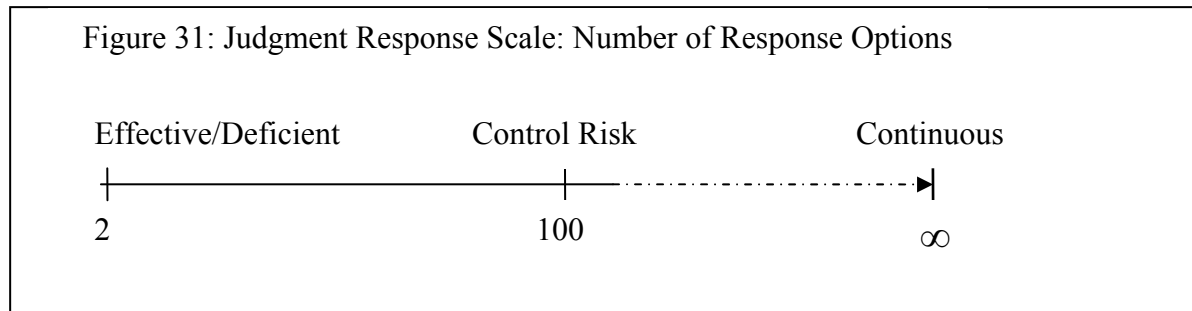


The range of variation in cue interrelationships is identical to what is described for control interrelationships under RQ2 above:

1. Completely-dependent
2. Amplifying
3. Independent
4. Compensating
5. Substitutable

RQ4a: What is the nature and range of variation in the judgment response scale variable?

This dissertation proposes that the judgment response scale as a task characteristic is the nature of the auditors' judgment response (e.g., risk, likelihood, impact) and the range of variation is the number of judgment response options the judge has available for a given judgment task. In the simplest form, a judgment response scale is binary (e.g., yes/no, effective/deficient, acceptable/unacceptable). As more response options become available, the scale approaches continuity (see figure 31 below). It is reasonable to assume that in many practical audit judgment situations the continuous response scale is approximated in the form of percentage point judgments (e.g., the audit risk model).



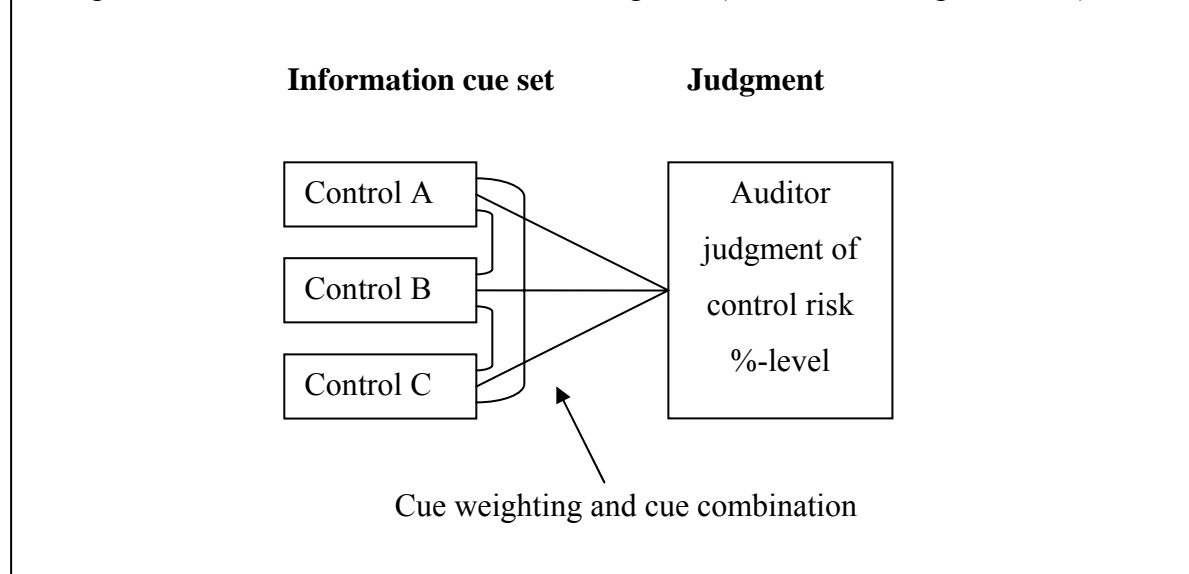
RQ4b: What is the nature and range of variation in the criterion scale variable?

The criterion is the true state of the reality that the judge is making a judgment about (typically an event or state). Although the level of the criterion is unknown to the judge, the nature and range of the criterion scale is known through knowledge of the judgment response scale (e.g., the judge knows that the nature of the criterion is control risk and that the range is a percentage scale since this is what he is asked to make a judgment about, but he does not know that the true level of control risk is e.g., 27%). Since the nature and range of the criterion are determined by the judgment response scale, the same scale also applies (i.e., the same scale as described under RQ4a above).

RQ5: What forms of judgment policies and models are relevant in auditors internal control judgments?

A judgment policy is the manner in which cues are weighted and combined by the judge when making a judgment. This can be illustrated by the judgment side of the Lens Model (see figure 32 below):

Figure 32: Lens Model for Control Risk Judgment (Lens Model Judgment Side)



A judgment model is a mathematical representation of a judgment policy. In policy capturing, a judgment model is a paramorphic (i.e., surface) representation of the judgment policy. The range of relevant judgment policies and models in internal control judgment tasks is proposed to include:

Compensatory, additive, linear models:

- Linear model

Compensatory, additive, nonlinear models:

- Compensatory form ordinal model
- Amplifying form ordinal model

Noncompensatory, nonadditive, nonlinear models:

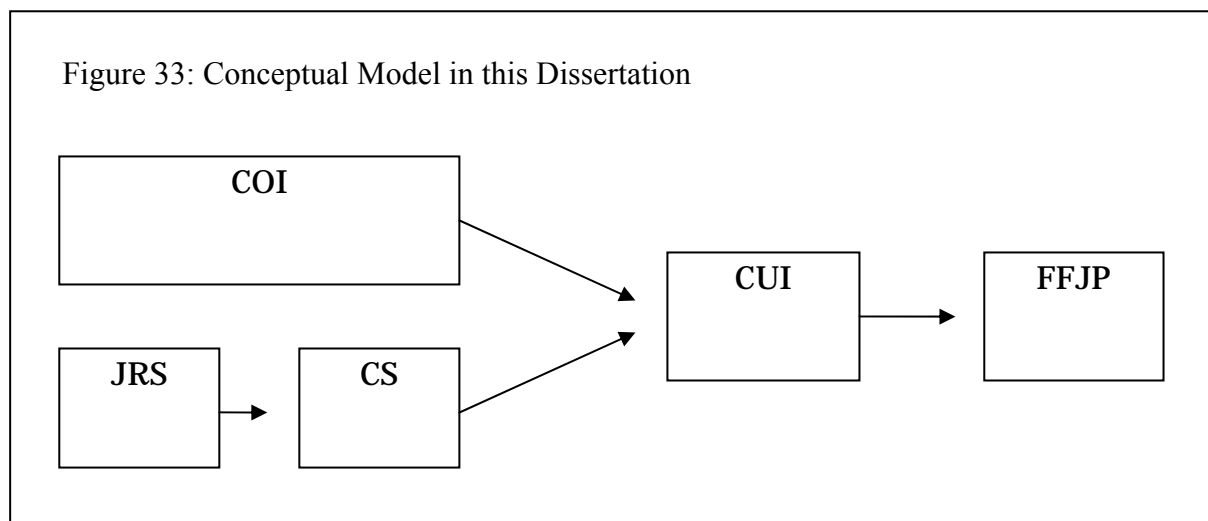
- Conjunctive model
- Disjunctive model

RQ6: How should control interrelationships and the judgment response scale affect cue interrelationships and judgment policies?

This dissertation suggests the following normative propositions:

- P1: A normatively appropriate functional form of the judgment policy (FFJP) can be derived from studying cue interrelationships (CUI).
- P2: Cue interrelationships (CUI) can be derived from studying control interrelationships (COI) and the judgment response scale (JRS).
- P3: The judgment response scale (JRS) determines the criterion scale (CS).

The propositions are illustrated in the conceptual model below (see figure 33 below):



The mechanism through which the judgment response scale alters cue interrelationships is proposed to be as follows:

- A change in the judgment response scale (JRS) implies a change in the criterion scale (CS)
- The criterion scale (CS) and control interrelationships (COI) determine cue interrelationships (CUI)
- Cue interrelationships (CUI) determine the functional form of the judgment policy (FFJP)

The following normative relationships are proposed:

Continuous judgment response scale

1. If controls are completely-dependent and the judgment response scale is continuous, then cues are completely dependent and a conjunctive judgment policy is appropriate.
2. If controls are compensating and the judgment response scale is continuous, then cues are compensating and a compensatory form ordinal judgment policy is appropriate.
3. If controls are independent and the judgment response scale is continuous, then cues are independent and a linear judgment policy is appropriate.
4. If controls are amplifying and the judgment response scale is continuous, then cues are amplifying and an amplifying form ordinal judgment policy is appropriate.
5. If controls are substitutable and the judgment response scale is continuous, then cues are substitutable and a disjunctive judgment policy is appropriate.

Binary judgment response scale:

1. If controls are substitutable and the judgment response scale is binary, then cues are substitutable (i.e., individually sufficient) and a disjunctive judgment policy is appropriate.
2. If controls are independent, amplifying, compensating or completely-dependent, and the judgment response scale is binary, then cues are completely-dependent (i.e., not individually sufficient) and a conjunctive judgment policy is appropriate

It can be noted that cues that are independent, compensating or amplifying in continuous judgment tasks are either (1) individually sufficient in binary judgment tasks (and thus become substitutable), or (2) not individually sufficient in binary judgment tasks (and thus become completely dependent). This can also be understood as sufficient (i.e., substitutable) or necessary (i.e., completely dependent) conditions for binary judgment outcomes. In general, independent, amplifying and compensating cues do therefore not, by definition, exist in binary judgment tasks.

RQ7: How do control interrelationships and the judgment response scale affect judgment policies?

This dissertation makes the following descriptive proposition:

P4: Auditors make judgments by using judgment policies that have normatively appropriate forms of cue integration given the cue interrelationships that result from control interrelationships and the judgment response scale.

RQ7 is the empirical equivalent to the normative research question in RQ6. It is hypothesized that auditors make judgments by using judgment policies that are consistent with what is normatively proposed in RQ6.

The conducted experiment provides the following evidence. For the continuous judgment response scale, findings are reported at the nomothetic level (i.e., aggregate group level). For the binary judgment response scale, findings are reported at the ideographic level (i.e., individual level):

Continuous judgment response scale

1. H1a: If controls are completely-dependent and the judgment response scale is continuous, then cues are completely dependent and a conjunctive judgment policy is appropriate. The aggregate group level judgment model is consistent with the use of such a judgment policy.
2. H2a: If controls are substitutable and the judgment response scale is continuous, then cues are substitutable and a disjunctive judgment policy is appropriate. The aggregate group level judgment model is consistent with the use of such a judgment policy.
3. H3a: If controls are independent and the judgment response scale is continuous, then cues are independent and a linear judgment policy is appropriate. The aggregate group level judgment model is consistent with the use of such a judgment policy.
4. H4a: If controls are compensating and the judgment response scale is continuous, then cues are compensating and a compensatory form ordinal judgment policy is

appropriate. The aggregate group level judgment model is consistent with the use of such a judgment policy.

5. H5a: If controls are amplifying and the judgment response scale is continuous, then cues are amplifying and an amplifying form ordinal judgment policy is appropriate. The aggregate group level judgment model is not consistent with the use of such a judgment policy. However, in this experiment a judgment model consistent with a linear judgment policy was, on average, used by participants. One potential explanation for this finding is that participants may not have perceived the cue interrelationship treatment as intended (see detailed discussion of H4a in the results section).

Binary judgment response scale:

1. H1b: If controls are completely-dependent and the judgment response scale is binary, then cues are completely-dependent (i.e., not individually sufficient) and a conjunctive judgment policy is appropriate and used. Analysis of individual judgment policies showed that most auditors (i.e., 20 out of 21) had a judgment model consistent with the hypothesized judgment policy. The likelihood of observing 20 out of 21 participants using the hypothesized judgment policy versus any other judgment policy being due to randomness is <0.001 . Results are therefore consistent with H1b.
2. H2b: If controls are substitutable and the judgment response scale is binary, then cues are substitutable (i.e., individually sufficient) and a disjunctive judgment policy is appropriate and used. Analysis of individual judgment policies showed that most auditors (i.e., 18 out of 21) had a judgment model consistent with the hypothesized judgment policy. The likelihood of observing 18 out of 21 participants using the hypothesized judgment policy versus any other judgment policy being due to randomness is <0.001 . Results are therefore consistent with H2b.
3. H3b: If controls are independent and the judgment response scale is binary, then cues are completely-dependent (i.e., not individually sufficient) and a conjunctive judgment policy is appropriate and used. Analysis of individual judgment policies

showed that most auditors (i.e., 18 out of 21) had a judgment model consistent with the hypothesized judgment policy. The likelihood of observing 18 out of 21 participants using the hypothesized judgment policy versus any other judgment policy being due to randomness is <0.001 . Results are therefore consistent with H3b.

4. H4b: If controls are compensating and the judgment response scale is binary, then cues are completely-dependent (i.e., not individually sufficient) and a conjunctive judgment policy is appropriate and used. Analysis of individual judgment policies showed that the majority of auditors (i.e., 13 out of 21, or 62%) had a judgment model consistent with the hypothesized judgment policy. The likelihood of observing 13 out of 21 participants using the hypothesized judgment policy versus any other judgment policy being due to randomness is <0.094 . Results are therefore marginally supportive of H4b. One potential explanation for the finding on compensating controls being only marginally supportive of H4b is that participants may have assessed the controls to be sufficiently compensating to be substitutable in the binary judgment, and thus may have not absorbed the intended cue interrelationship treatment (see further discussion in section 6.3.4).
5. H5b: If controls are amplifying and the judgment response scale is binary, then cues are completely-dependent (i.e., not individually sufficient) and a conjunctive judgment policy is appropriate and used. Analysis of individual judgment policies showed that most auditors (i.e., 15 out of 21) had a judgment model consistent with the hypothesized judgment policy. The likelihood of observing 15 out of 21 participants using the hypothesized judgment policy versus any other judgment policy being due to randomness is <0.006 . Results are therefore consistent with H5b.

Overall it seems that auditors generally make judgments by using normatively appropriate judgment policies given the cue interrelationships that result from control interrelationships and the judgment response scale. In this study, the task characteristics of amplifying cues and a continuous judgment response scale, was an exception to this general finding: On average auditors exhibited a linear judgment model instead of the hypothesized amplifying form ordinal model. It is, however, not unlikely that the exception was due to participants

not absorbing the cue interrelationship treatment as intended. Furthermore findings for compensating controls and a binary judgment response scale are only marginally supportive of the hypothesized judgment policy.

7.2 Contribution to Research

Brown and Solomon (1990) is the first and only study of auditor's control judgments finding evidence of configularity. Prior to this no consistent evidence of configularity had been found. The contribution of this dissertation to control judgment research can therefore be understood by asking what it adds to Brown and Solomon (1990).

First, control interrelationships and cue interrelationships are defined as two different constructs. Prior audit research did not distinguish between control interrelationships and cue interrelationships.

Second, prior research was limited to studying compensating, independent and amplifying controls. This dissertation extends the range of variation in control interrelationships so that it includes completely-dependent controls (i.e., multi-step controls). Furthermore, the importance of the degree of compensation between controls is clarified, and substitutable controls are defined. Overall, a framework defining the complete range of variation in control interrelationships is developed. Future research using control interrelationships as a variable should therefore have a better developed construct than what was previously available. In addition, the framework is applicable to cue interrelationships in general. It may therefore be useful for research using other judges and tasks, also outside of auditing.

Third, Prior research used percentage point control risk judgments (e.g., Brown and Solomon 1990) or categorical scales (e.g., Ashton (1974) used a seven point categorical risk scale). Furthermore, no assumptions were made regarding the judgment response scale and it was not included as a variable affecting cue combination. This study introduces and defines the judgment response scale as an important task characteristic when studying cue combination. In addition, a potential theoretical explanation is provided for why cue combination changes when the judgment response scale changes from continuous to binary (i.e., by changing the criterion scale and thus cue interrelationships).

Fourth, this study extends the range of variation in judgment policies studied to include disjunctive and conjunctive policies. Prior research on control judgments is limited to studying compensating, amplifying and linear policies.

Fifth, prior research made normative predictions for how independent, compensating and amplifying controls should determine judgment policies to be linear, compensating and amplifying (respectively). This study clarifies and extends prior predictions in the following way: (1) it is clarified that cue interrelationships, not control interrelationships, determine judgment policies. Independent, compensating and amplifying cues determine judgment policies to be linear, compensating and amplifying (respectively). (2) Normative predictions are extended to include completely-dependent cues and substitutable cues, which determine judgment policies to be conjunctive and disjunctive (respectively).

Sixth, theory is developed for how control interrelationships and the judgment response scale affect cue interrelationships. Prior research did not make any assumptions about this since cue- and control interrelationships were not viewed as two different constructs and since the judgment response scale was not included as a variable.

Seventh, prior research found that few auditors combined cues in an appropriate configural manner (see section 3.5.1). This study provides evidence indicating that on average, auditors combines cues appropriately, apart from for judgments of amplifying cues. The evidence on cue combination in control risk judgments for completely-dependent and substitutable controls is an extension of prior research. The evidence on cue combination in binary judgments is also new to control judgment research.

Audit research is thus extended by: (RQ1) clarifying the difference between control interrelationships and cue interrelationships, (RQ2) developing a framework defining the range of variation in control interrelationships, (RQ3) developing a framework defining the range of variation in cue interrelationships, introducing and defining the judgment response scale (RQ4a) and the criterion scale (RQ4b) as task characteristics in internal control judgments, including defining the range of variation, (RQ5) defining relevant functional forms of judgment policies in internal control judgments, (RQ6) developing normative theory about how control interrelationships and the judgment response scale should affect cue interrelationships and the functional form of judgment policies, and (RQ7) providing

empirical evidence on how control interrelationships and the judgment response scale affect the functional form of judgment policies.

The construct development and normative theory development of the dissertation should also be relevant for judgment research in other fields (i.e., other judges and/or other tasks) where similar task characteristics are relevant. The general contribution of this dissertation is thus to: (1) develop theory on how task characteristics determine cue interrelationships, (2) develop a more complete framework of cue interrelationships, (3) add to normative theory of how task characteristics should affect judgment policies, and (4) add to descriptive evidence on how task characteristics actually affect judgment policies.

7.3 Contribution to Audit Practice

The theoretical contribution of the dissertation should also benefit audit practice. First of all it is unclear whether today's auditors apply appropriate judgment policies in internal control judgments. Prior evidence revealed a relatively low extent of appropriate configural judgment policies even when the internal control task required it (Brown and Solomon 1990). Therefore, it should be of interest for practitioners to receive updated evidence on whether this is still a problem.

Second, if inappropriate judgment policies are applied, judgment quality, and thus audit quality may suffer (Brown and Solomon 1990, 1991; Hooper and Trotman 1996; Leung and Trotman 2008). Audit practice should therefore benefit from the development of normative benchmarks for evaluation of actual judgment policies. Such normative benchmarks may help in identifying differences between appropriate and actual judgment policies and thus shed light on where improvement is needed. Deviations from normative benchmarks may also provide an explanation for poor performance and disagreements between judges (Libby 1981, 31-32).

Third, knowledge about the relationship between task characteristics and the appropriate form of the judgment policy can help in training decision makers and in developing decision aids.

Finally, the construct development in this dissertation may provide useful frameworks and definitions for analyzing task characteristics both in real life audit settings and in the classroom.

7.4 Limitations

A general limitation with policy capturing is that the methodology does not allow observation of the cue processing that goes on within the mind. It only provides input-output correlations of cues and judgments. The resulting regression models are interpreted as representing various forms of judgment policies, but this can not be certain.

This causes a specific limitation for the hypotheses regarding the binary judgment response scale (i.e., the “b” hypotheses) since two alternative judgment policies may explain the observed regression model. It may be that the judgments are made in two steps; first a control risk score is judged, and then an accept/reject binary judgment is based on this score. This implies a different policy than the hypothesized conjunctive/disjunctive judgment policies, even though the judgments and cue levels are identical. This limitation can not be overcome with the applied methodology. However, in the post experimental survey, subjects are asked to describe their actual judgment policy, and the responses indicate that two thirds of participants (i.e., 14 out of 21) used the hypothesized judgment policy (the likelihood of this being due to randomness is <0.039). This is, however, not a strong test since subjects self insight may be limited (Bonner 2007). The issue may be interesting for future research.

Another limitation with the study is that seven out of twenty-eight participants were deleted from analysis since their judgments and responses to follow-up questions revealed that they may have perceived cue interrelationships differently from what was intended (see discussion under section 6.1.2). Future studies should provide clearer background information on inherent risk to avoid deviating cue interrelationship perceptions.

For the continuous judgment response scale hypotheses, the level of analysis is aggregated and not individual. The study does therefore not quantify the proportion of judges using appropriate judgment policies. This may be an issue for future research. See discussion in section 6.1.1 for why analysis is done at the aggregated level.

For the binary judgment response scale hypotheses, the level of analysis is individual. This is due to the statistical software package not being able to estimate regression parameters due to characteristics of the data set (see explanation in section 6.1.1). It is therefore not possible to perform a statistical test of the average judgment policies being as hypothesized. The descriptions of the individual judgment policies do, however, provide a basis for statistically testing the likelihood of the observed judgment policies versus any other judgment policy being due to randomness. The test provides a clear picture of the applied judgment policies generally being as hypothesized.

A final limitation is made regarding the audit setting: Internal controls only provide reasonable assurance, not absolute assurance (AS5.A7 PCAOB 2007). Controls that provide individually sufficient control risk reduction, given the reasonable assurance criterion, may therefore be considered as substitutable cues in a binary judgment task, even though they do not provide equal control risk reduction (i.e., they are not fully substitutable controls, but only sufficiently compensating controls). An example could be controls with a sufficiently high degree of compensation for both to be individually above the reasonable assurance criterion. In such a situation compensating controls would result in substitutable cues.

In addition, the usual experimental method limitations apply; e.g., see discussion of external validity of subjects under 5.1, discussion of operationalization of constructs in 5.2.2, discussion of representativeness of cases (i.e., external validity) in 5.2.2, and discussion of experimental procedures (i.e., internal validity) in 5.2.3.

7.5 Suggestions for Future Research

This dissertation has raised a number of issues that may provide interesting questions for future research. First, it seems that the judge's perception of cue interrelationships is important for the effect of cue interrelationships on the judgment response scale. It would therefore be useful to learn more about what affects the judge's ability to perceive cue interrelationships correctly. Research on this issue would need to be based on other methodologies than policy capturing.

Second, why is the evidence of auditor's ability to use judgment models appropriate to amplifying controls/cues so weak? This dissertation found that auditors at the aggregate

group level had judgment models consistent with a linear judgment policy. Similarly, Brown and Solomon (1990) found that only 9 out of 74 auditors (i.e., 12.2%) had the appropriate amplifying form interaction in their judgment models. A first step in resolving this question is to find out whether auditors have a problem with perceiving amplifying cue interrelationships correctly, or whether the problem lies with their ability to apply an amplifying form judgment policy conditional on a correct cue perception. For auditors, and others making control judgments, appropriate combination of amplifying cues may be important since entity level controls are generally thought to relate to the remaining control system in an amplifying manner.

Third, research aimed at understanding the thought process in binary judgments could help understand whether auditors make binary judgments by first judging control risk and then judging whether it is acceptable or not, or whether auditors directly judge acceptability of controls. For practitioners, a direct judgment would clearly be more efficient, since one with disjunctive and conjunctive models may be able to make the judgment based on only one control instead of testing and combining all controls to an overall risk judgment and then making the binary judgment.

Fourth, audit judgments are made on many discrete judgment response scales (e.g., classification of control deficiencies into three categories; deficiency, significant deficiency or material weakness, or classification of control maturity on a five point scale). Future research could therefore study the models application on other judgment response scales than binary and percentage point scales. This may also be useful for practitioners since a theoretical foundation for analyzing cue interrelationships conditional on varying judgment response scales may ease judgment complexity.

Finally, future research could study the usefulness of the proposed conceptual model and framework in settings with other tasks and judges. A first approach could be to cross validate the findings using other controls as cues, and other judges. A next step could be using audit tasks other than internal control judgments. At the more generic level, studies could use a setting with non-audit tasks and non- auditor judges.

References

- Abdel-khalik, A.R., D. Snowball and T.H. Wragge. 1983. The effects of certain internal audit variables on the planning of external audit programs. *The Accounting Review* (58): 215-227
- American Accounting Association, Committee on Basic Auditing Concepts. 1973. *A Statement of Basic Auditing Concepts*. AAA, Sarasota, FL
- American Institute of Certified Public Accountants (AICPA). 1979. *Report of the Special Advisory Committee on Internal Accounting Control*. New York: AICPA
- 1988. Consideration of Internal Control in a Financial Statement Audit. Statement on Auditing Standards No. 55. New York, NY: AICPA.
- . 2002. AU Section 319: Consideration of Internal Control in a Financial Statement Audit. AICPA.
- . 2002. AU Section 325: Communications about Control Deficiencies in an Audit of Financial Statements. AICPA.
- . 2006. Assessing and Responding to Audit Risk in a Financial Statement Audit. *AICPA Audit Guide*. New York, NY: American Institute of Certified Public Accountants, Inc.
- . 2008. AU Section 314: Understanding the Entity and its Environment and Assessing the Risk of Material Misstatement. AICPA.
- Aiman-Smith, L., S.E. Scullen, S.H. Barr. 2002. Conducting Studies of Decision Making in Organizational Contexts: A Tutorial for Policy-Capturing and Other Regression-Based Techniques. *Organizational Research Methods* 5 (October): 388-414
- Ashbaugh-Skaife, H., D.W. Collins, W.R. Kinney and R. LaFond. 2008. The Effect of Internal Control Deficiencies on Firm Risk and Cost of Capital. *The Accounting Review* 83: 217-250.
- Ashton, R. H. 1974. An Experimental Study of Internal Control Judgments. *Journal of Accounting Research* 12 (Spring):143-157.
- . 1982. *Human Information Processing in Accounting*. *Studies in Accounting Research* No. 17. Sarasota, FL: AAA.
- . 1983. *Research in audit decision making: Rationale, evidence and implications*. The Canadian Certified General Accountants' Research Foundation, Vancouver.
- Ashton, R. H. and A. H. Ashton. 1995. *Judgment and Decision Making Research in Accounting and Auditing*. New York, NY: Cambridge University Press.
- Ashton, R.H. and P.R. Brown. 1980. Descriptive modeling of auditors' internal control judgments: Replication and extension. *Journal of Accounting Research* 18: 269-277.

-
- Bedard, J. C. and S. F. Biggs. 1991. Pattern recognition, hypothesis generation, and auditor performance in an analytical task. *The Accounting Review* 66: 622–642.
- Bell, T.B. and A. Wright. 1995. Auditing practice, research and education: A productive collaboration. New York: American Institute of Certified Public Accountants.
- Bonner, S. E. 1994. A model of the effects of audit task complexity. *Accounting, Organizations and Society* 19 (3): 213–234.
- . 1999. Judgment and decision-making research in accounting. *Accounting Horizons* 13 (4): 385-398.
- . 2007. *Judgment and Decision Making in Accounting*. Prentice-Hall, Englewood Cliffs, NJ.
- Brazel, J.F., C.P. Agoglia and R.C. Hatfield. 2004. Electronic versus Face-to-Face Review: The Effects of Alternative Forms of Review on Auditors Performance. *Accounting Review* 79 (4): 949-966
- Brehmer, B. 1994. The psychology of linear judgment models. *Acta Psychologica* 87: 137-154.
- Brown, C. E., and I. Solomon. 1990. Auditor configural information processing in control risk assessment. *Auditing: A Journal of Practice & Theory* 9 (Fall): 17-38.
- . 1991. Configural information processing in auditing: the role of domain specific knowledge. *Accounting Review* 66 (January): 100-119.
- Brunswick, E. 1947. *Systematic and representative design of psychological experiments*. Berkely: University of California Press.
- Brunswick, E. 1952. *The conceptual framework of psychology*. Chicago: University of Chicago Press.
- Cooksey, R. W., 1996. *Judgment analysis: theory, methods and applications*. San Diego, CA: Academic Press.
- Colbert, J.L. 1988. Inherent risk: An investigation of auditors' judgments. *Accounting, Organizations and Society* 13: 111-121
- Committee of Sponsoring Organisations (COSO). 1992. *Internal Control-Integrated Framework*. New York, NY: COSO.
- . 2004. *Enterprise Risk Management-Integrated Framework (Executive Summary)*. New York, NY: COSO.
- . 2006. *Internal Control over Financial reporting – Guidance for Smaller Public Companies*. New York, NY: COSO.
- .2008. *Guidance on Monitoring Internal Control Systems (Exposure Draft, June 2008)*. New York, NY: COSO.

Doss, M., Jonas, G., 2004. Section 404 reports on internal control: Impact on ratings will depend on nature of material weaknesses reported. *Moodys Investors Service, Global Credit Research*. October.

Eilifsen, A. 1998. Auditing regulation and the statutory auditor's responsibilities in Norway. *The European Accounting Review* 7 (4): 709-722

Eilifsen, A. and W. F. Messier, S. M. Glover, and D. F. Prawitt. 2006. *Auditing & Assurance Services – International Edition*. McGraw-Hill, Berkshire, UK.

Einhorn, H. J. 1970. The use of nonlinear, noncompensatory models in decision making. *Psychological Bulletin* 73: 221–230.

Einhorn, H. J. 1971. Use of nonlinear, noncompensatory models as a function of task and amount of information. *Organizational Behavior and Human Performance* 6: 1–27.

Einhorn, H. J., D. N. Kleinmuntz, and B. Kleinmuntz. 1979. Linear Regression and Process Tracing Methods of Judgments. *Psychological Review* (1979), pp. 465-85.

Elrod, T., R.D. Johnson, and J. White. 2004. A new integrated model of noncompensatory and compensatory decision strategies. *Organizational Behavior and Human Decision Processes* 95: 1-19.

European Parliament and the Council of the European Union. 2006. Directive 2006/43/EC on Statutory Audits of Annual Accounts and Consolidated Accounts. *Official Journal of the European Union* (09.06.2006): 87-107.

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

Hammersley, J.S. 2006. Pattern identification and industry-specialist auditors. *The Accounting Review* 81 (2): 309-336.

Hammond, K.R. 1955. Probabilistic functionalism and the clinical method. *Psychological Review* 62: 255-262.

Hammond, K.R., T.R. Stewart, B. Brehmer, and D.O. Steinman. 1975. Social judgment theory. In M. Kaplan & S. Schwartz (Eds.), *Human judgment and decision processes*. New York: Academic Press: 271-312.

Heier, J.R., M.T. Dugan, D.L. Sayers. 2005. A century of debate for internal controls and their assessment: a study of reactive evolution. *Accounting History*; Nov 2005; 10, 3; *Accounting and Tax Periodicals* p39.

Hoffman, P.J. 1960. The paramorphic representation of clinical judgment. *Psychological Bulletin* 57: 116-131.

Hogarth, R.M., and H.J. Einhorn. 1992. Order effects in belief updating: the belief adjustment model. *Cognitive Psychology* 24: 1-55.

Hogarth, R.M., and N. Karelaia. 2007. Heuristic and Linear Models of Judgment: Matching Rules and Environments. *Psychological Review* 114 (3): 733–758

-
- Hoffman, P. J., P. Slovic, and L. G. Rorer. 1968. An Analysis-of-Variance Model for the Assessment of Configural Cue Utilization in Clinical Judgment. *Psychological Bulletin* (May): 338-49.
- Hooper, C., and K.T. Trotman. 1996. Configural information processing in auditing: further evidence. *Accounting and Finance* 26 (2): 125-136
- Humphrey, C., and A Loft. 2007. IFAC 2001-2007: A Regulatory Renaissance. IFAC: <www.ifac.org/history> (June 19, 2008).
- International Auditing and Assurance Standards Board (IAASB).2008. Project History: Audit Risk: IAASB: <<http://www.ifac.org/IAASB/ProjectHistory.php?ProjID=0004>> (August 19, 2008)
- International Federation of Accountants (IFAC). 2008. *Handbook of International Auditing, Assurance and Ethics Pronouncements (2008 Edition)*: IFAC: <http://www.ifac.org/Members/Downloads/2008_IAASB_Handbook_Part_I-Compilation.pdf> (August 19, 2008)
- Joyce, E. J. 1976. Expert Judgment in Audit Program Planning. *Journal of Accounting Research* (Supplement): 29-60.
- Justis- og Politidepartementet. 1997. Lov om Aksjeselskap (Aksjeloven). Avd I 1997 Nr. 14
- Karren, R.J., M.W. Barringer. 2002. A Review and Analysis of the Policy-Capturing Methodology in Organizational Research: Guidelines for Research and Practice. *Organizational Research Methods* 5 (October): 337-361
- Knechel, R.W.. 2007. The business risk audit: Origins, obstacles and opportunities. *Accounting, Organizations and Society* 32: 383-408
- Leon, M., G. C. Oden, and N. H. Anderson. 1973. Functional Measurement of Social Values. *Journal of Personality and Social Psychology*: 301-10.
- Leung, P.W., and K.T. Trotman. 2005. The effects of feedback type on auditor judgment performance for configural and non-configural tasks. *Accounting, Organizations and Society* 30: 537-553
- Libby, R. 1979. Bankers' and auditors' perceptions of the message communicated by the audit report. *Journal of Accounting Research* 17: 99-122
- Libby, R. 1981, *Accounting and human information processing: Theory and applications*. Prentice-Hall, Englewood Cliffs, NJ.
- Libby, R. and B. L. Lewis. 1977. Human information processing research in accounting: The state of the art. *Accounting, Organizations and Society* 2 (3): 245-268.
- Libby, R. and B. L. Lewis. 1982. Human information processing research in accounting: The state of the art in 1982. *Accounting, Organizations and Society* 7 (3): 231–286.
- Maletta, M. J., and T. Kida. 1993. The effect of risk factors on auditors' configural information processing. *The Accounting Review* 68 (July): 681–691.

Messier Jr, W.F. 1983. The effect of experience and firm type on materiality disclosure judgments. *Journal of Accounting Research* 21 (2): 618-622

Nelson, M. and H-T. Tan. 2005. Judgment and decision making research in auditing: A task, person, and interpersonal interaction perspective. *Auditing: A Journal of Practice & Theory* 24 (Supplement): 41-71.

Ogneva M., K.R. Subramanyam and K. Raghunandan. 2007. Internal Control Weakness and The Cost of Equity: Evidence from SOX Section 404 Disclosures. *The Accounting Review* 82 (5): 1255-1298.

Peecher, M.E., R. Schwartz, I. Solomon. 2007. Theory and experimentation in studies of audit judgments and decisions: avoiding common research traps. *International Journal of Auditing* 5: 193-203.

Peecher, M.E., I. Solomon. 2001. The business risk audit: Origins, obstacles and opportunities. *Accounting, Organizations and Society* 32: 463-485.

Public Company Accounting Oversight Board (PCAOB). 2004. Auditing Standard No. 2: An Audit of Internal Control Over Financial Reporting Performed in Conjunction With an Audit of Financial Statements. PCAOB Release No. 2004-001, March 2004. Washington D.C.

---. 2005. Standing Advisory Group Meeting: Risk Assessment in Financial Statement Audits. Washington, D.C.: PCAOB.

----. 2006. Proposed Auditing Standard: An Audit of Internal Control Over Financial Reporting That is Integrated with an Audit of Financial Statements and Related Other Proposals. PCAOB Release No. 2006-007 (Dec 19, 2006). Washington D.C.

---. 2007. Auditing Standard No.5: An Audit of Internal Control over Financial Reporting That is Integrated with an Audit of Financial Statements. PCAOB-2007-02; July 27, 2007. Washington D.C.

POB (Public Oversight Board). 2000. *Panel on Audit Effectiveness. Report and Recommendations*. Stamford. CT.

Rorer. L., P. J. Hoffman. H. R. Dickman, and P. Slovic, 1967. Configural Judgments Revealed. *Proceedings of the 75th Annual American Psychological Association Convention*: 195-96.

Sarbanes-Oxley Act. 2002. Corporate Responsibility, Public Law 107-204, 107th Congress of the United States. Washington, D.C.: Government Printing Office.

Schepanski, A., R.M. Tubbs, and R.A. Grimlund. 1992. Issues of Concern Regarding Between-Subjects Designs in Behavioral Accounting Research. *Journal of Accounting Literature* (11): 121

Securities and Exchange Commission. 2007. Management's Report on Internal Control over Financial Reporting. RELEASE NOS. 33-8762; 34-54976; File No. S7-24-06: p25. SEC.

-
- Shanteau, J., and T.R. Stewart. 1992. Why study expert decision making? Some historical perspectives and comments. *Organizational Behavior and Human Decision Processes* 53; 95-106.
- Simon, H. A. 1956. Rational choice and the structure of environments. *Psychological Review* 63: 129–138.
- Slovic, P. 1969. Analyzing the Expert Judge: A Descriptive Study of a Stockbroker's Decision Processes. *Journal of Applied Psychology* (August): 255-63.
- Slovic, P., and S. Lichtenstein. 1973. Comparison of Bayesian and regression approaches to the study of information processing in judgment. In L. Rappoport & D.A. Summers (Eds.), *Human Judgment and Social Interaction*. New York: Holt, Rinehart & Winston: 16-108.
- Slovic, P., B. Fischhoff, and S. Lichtenstein. 1977. Behavioral Decision Theory. *Annual Review of Psychology*: 1-39.
- Solomon, I. and M. D. Shields. 1995. Judgment and decision-making research in auditing. *Judgment and Decision-Making Research in Accounting and Auditing*, edited by R. H. Ashton and A. H. Ashton. Cambridge, MA: Cambridge University Press.
- Solomon, I. and K. T. Trotman. 2003. Experimental judgment and decision research in auditing: The first 25 years of AOS. *Accounting, Organizations and Society* 28 (4): 395–412.
- Stewart, T.R. 1988. Judgment analysis: procedures. In B. Brehmer, and C.R.B. Joyce (Eds.), *Human Judgment: the SJT view*. Amsterdam: North-Holland Elsevier: 41-74.
- Stewart, T.R., P.J. Roebber, and L.F. Bosart. 1997. The importance of the task in analyzing expert judgment. *Organizational Behavior and Human Decision Processes* 69 (March); 205-219.
- Trotman, K.T. and R. Wood. 1991. A meta-analysis of studies on internal control judgments. *Journal of Accounting Research* 29: 180-192.
- . 1996. *Research methods for judgment and decision making studies in auditing*. Coopers & Lybrand and Accounting Association of Australia and New Zealand: Melbourne.
- . 1998. Audit judgment research – issues addressed, research methods and future directions. *Accounting and Finance* 38: 115-156
- . 2005. DISCUSSION OF Judgment and decision making research in auditing: A task, person, and interpersonal interaction perspective. *Auditing: A Journal of Practice & Theory* 24 (Supplement): 73-87.

Appendix 1 – Cover Letter and Introduction

Dear Participant:

The study that you are about to participate in investigates how auditing professionals make judgments about internal controls. You are therefore asked to assume the role of an audit manager who is making judgments about internal controls.

Please respond to the questions as you would on an actual audit engagement. Please do not skip any of the questions. Since this research is focused on individual decision-making, please do not discuss with other participants before you complete the survey.

The information provided to you includes the following:

- Introduction: (1) background information about the hypothetical audit client, (2) a framework for understanding the interrelationship between controllers, and (3) a description of the task you are asked to perform. You may refer back to the introduction materials at any time.
- Envelope 1-5: Five envelopes, each containing one classification exercise and eight cases. For each envelope, please start by completing the classification exercise, and then continue with the eight cases. When the classification and cases in one envelope are completed, they should be sealed in the envelope before continuing to the next envelope.
- Envelope 6: You will be asked to provide some demographic data and respond to questions concerning the cases included in the study. These are to be sealed in the final envelope.

Please open and complete one envelope at a time, and in the provided order.

Please view the differences in information in cases as hypothetical variation, and solve each case independently (i.e., the information in a given case is not relevant for other cases).

Thank you very much for participating in this research study.

Jonas Gaudernack

PhD student, NHH

Senior Manager, PricewaterhouseCoopers

Background information about the audit client

General description:

The audit client is a large, three-location retail company with an average business risk profile. It can be considered as a generic retail company without any unusual/special risks. The three locations do not differ in any relevant way, and they are all individually material to the audit.

Risk and materiality:

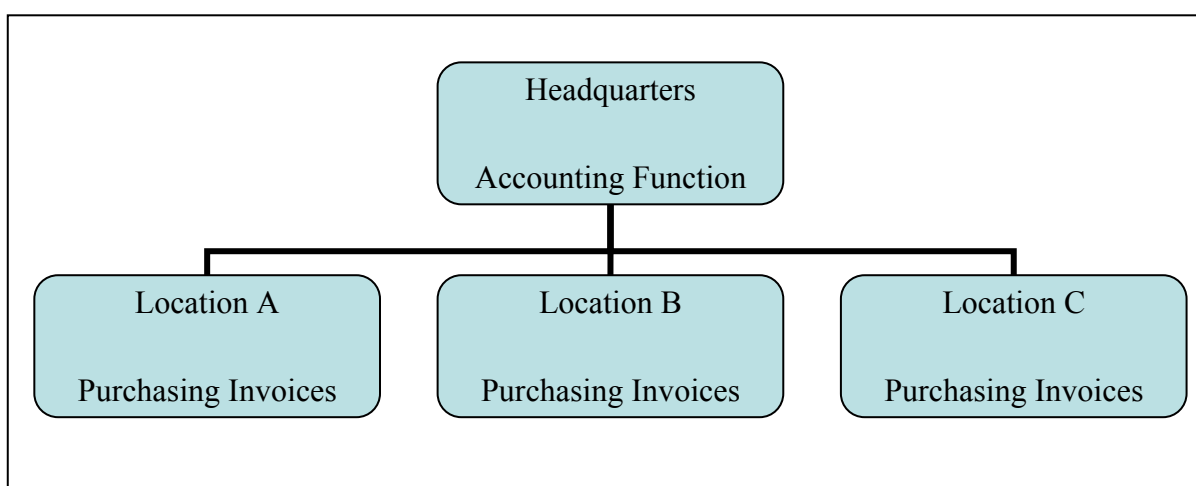
- The three locations are roughly identical in size and risk.
- The business at a location is conducted independent of business in other locations. Risks at a location are therefore independent of risks at other locations.
- Since each location is material to the audit, each location must have an acceptable level of effective internal control.
- Inherent risk, “IR”, has been set to 100% (i.e., maximum risk; there is 100% risk of error(s) if controls are deficient).

Number and value of transactions:

- Assume that the number of purchasing transactions is approximately the same at all three locations.
- Assume that the value of each individual purchasing transaction is approximately the same.

Internal control:

- The accounting function at the headquarters employs three controllers. Generally, each controller is dedicated to controlling transactions at a specific location, although this may vary in the case materials. When reading the case materials, please be sure to understand the control responsibility of each controller; especially if a controller’s work depends on, overlaps, or impacts, the work of other controllers (a framework is provided to help you with this – see next page).



Framework: Controller Interaction

Controller Interaction:	Explanation:
Independent	A controller checks a set of transactions. Other controller(s) double-check <u>none</u> of the same transactions (i.e., there is <u>no</u> overlap in control).
Compensating	A controller checks a set of transactions. Other controller(s) double-check <u>some</u> of the same transactions (i.e., there is <u>partial</u> overlap in control).
Substitutable	A controller checks a set of transactions. Other controller(s) double-check <u>all</u> of the same transactions (i.e., there is <u>complete</u> overlap in control).
Multi-Step	Each controller performs a separate step in a multi-step control. Individual steps are not separate controls, but together they make up one (multi-step) control. Unless <u>all</u> controllers perform their steps effectively, control is ineffective. All of the controller's work is therefore dependent on other controllers.
Amplifying	Interaction between controllers is named amplifying when: <ul style="list-style-type: none"> • A controller performs some work independently (i.e., the work is a separate control that is effective regardless of what other controllers do), <u>and</u> • A controller performs some work together with another controller (i.e., the work is not a separate control, but only a part of a control that is performed together with another controller who performs the other part of the control). The controller's work therefore results in a control being effective only if both controllers do their part. <p>Some, but not all, of a controller's work is therefore dependent on another controller.</p>

The following questions may help in understanding controller interactions:

- If two controllers check the same transactions, are they doing so because (1) it is a double check (i.e., double controls; the control is performed twice for those transactions) or (2) because the control consists of two parts where each controller performs one part (i.e., the control is only performed once, but by two controllers performing separate parts)?
- Effect on other controllers:
 - If a controller fails, does it affect other controller's ability to reduce control risk?
- Effect of other controllers:
 - If a controller fails, to what extent do other controllers compensate for that failure?
 - If other controllers fail, does it affect a controller's ability to reduce control risk?

Your task

You are the audit manager for a number of specific audit areas. For each audit area, you have been asked by the partner to classify the interaction between controllers by using the framework provided on the previous page, and to make the following two judgments:

- (1) **What is control risk in the audit area?** (i.e., “CR” in the audit risk model; $AR=IR*CR*DR$)? “CR” is loosely defined as the risk of error after the company has performed controls.
- (2) **Please assess the audit test results: Does the client (i.e., all locations) have sufficient operationally effective controls for the given audit area?**

Inherent risk, “IR”, has been set to 100% (i.e., maximum risk; there is 100% risk of error(s) if controls are deficient).

The scope of your responsibility is limited to judgments regarding the defined audit area. The audit area may vary, therefore, please be sure to understand the audit area.

You should not make any other assumptions than those provided in the introduction materials and in the specific case you are responding to (e.g., no other controls or control objectives should be relevant - you can assume these are handled by other audit managers).

Prior audit work

Control design: The internal controls you are assessing have been judged to be designed effectively (i.e., you can assume design effectiveness).

Control operation: The audit team has tested whether individual controllers perform the designed controls effectively (i.e., operational effectiveness), but no overall judgments have been made (this is your task). You can rely on the results from the auditor’s tests: If the auditor’s test is positive, you can assume that the tested controller performs all of his control(s) effectively. If the auditor’s test is negative, you can assume that the tested controller performs none of his control(s) effectively (i.e., you can assume that the controller doesn’t exist). A controller therefore performs all or none of his controls effectively.

Understanding the auditor's documentation template

Documentation of control design and audit test results is done in tables. Please be sure to understand the difference between (1) the descriptions of the clients control design, and (2) the results of the auditor's testing of whether the controllers perform their controls effectively.

Example:

Please be sure to understand the audit area you are asked to make judgments for



Audit Area: accuracy of booked incoming invoices			
Control Design:	Control Objective	Operationally Effective? (Audit test result)	
Controller "A" reviews <u>all</u> (i.e., 100% of) booked incoming invoices from <u>location "A"</u> to check that they are booked in the accounting system with the correct amount.	Accuracy	YES	
Controller "B" reviews <u>all</u> (i.e., 100% of) booked incoming invoices from <u>location "B"</u> to check that they are booked in the accounting system with the correct amount.	Accuracy		NO
Controller "C" reviews <u>all</u> (i.e., 100% of) booked incoming invoices from <u>location "C"</u> to check that they are booked in the accounting system with the correct amount.	Accuracy	YES	

Description of the design of control(s) performed by controller "C"

The result of the auditor's test of controller "C":

"YES" = Controller "C" performs all of his control(s) effectively

"NO" = Controller "C" performs none of his control(s) effectively

Appendix 2 – Examples of cases

Example Independent Controls:

Audit Area: accuracy of booked incoming invoices			
Control Design:	Control Objective	Operationally Effective? (Audit test result)	
<u>Controller “A”</u> reviews <u>all</u> (i.e., 100% of) booked incoming invoices from <u>location “A”</u> to check that they are booked in the accounting system with the correct amount.	Accuracy	YES	
<u>Controller “B”</u> reviews <u>all</u> (i.e., 100% of) booked incoming invoices from <u>location “B”</u> to check that they are booked in the accounting system with the correct amount.	Accuracy		NO
<u>Controller “C”</u> reviews <u>all</u> (i.e., 100% of) booked incoming invoices from <u>location “C”</u> to check that they are booked in the accounting system with the correct amount.	Accuracy	YES	

Example Compensating Controls:

Audit Area: accuracy of booked incoming invoices			
Control Design:	Control Objective	Operationally Effective? (Audit test result)	
<p><u>Controller "A"</u> reviews <u>all</u> (i.e., 100% of) booked incoming invoices from <u>location "A"</u> to check that they are booked in the accounting system with the correct amount.</p> <p>In addition, <u>Controller "A"</u> reviews <u>half of all</u> (i.e., 50% of) booked incoming invoices from <u>location "C"</u> to double check that they are booked in the accounting system with the correct amount.</p>	Accuracy	YES	
<p><u>Controller "B"</u> reviews <u>all</u> (i.e., 100% of) booked incoming invoices from <u>location "B"</u> to check that they are booked in the accounting system with the correct amount.</p>	Accuracy		NO
<p><u>Controller "C"</u> reviews <u>all</u> (i.e., 100% of) booked incoming invoices from <u>location "C"</u> to check that they are booked in the accounting system with the correct amount.</p> <p>In addition, <u>Controller "C"</u> reviews <u>half of all</u> (i.e., 50% of) booked incoming invoices from <u>location "A"</u> to double check that they are booked in the accounting system with the correct amount.</p>	Accuracy	YES	

Example Substitutable Controls:

<u>Audit Area:</u> accuracy of booked incoming invoices			
Control Design:	Control Objective	Operationally Effective? (Audit test result)	
<p><u>Controller "A"</u> reviews <u>all</u> (i.e., 100% of) booked incoming invoices from <u>location "A"</u> to check that they are booked in the accounting system with the correct amount.</p> <p>In addition <u>controller "A"</u> reviews <u>all</u> (i.e., 100% of) booked incoming invoices from <u>location "B"</u> to double check that they are booked in the accounting system with the correct amount.</p>	Accuracy	YES	
<p><u>Controller "B"</u> reviews <u>all</u> (i.e., 100% of) booked incoming invoices from <u>location "B"</u> to check that they are booked in the accounting system with the correct amount.</p> <p>In addition <u>controller "B"</u> reviews <u>all</u> (i.e., 100% of) booked incoming invoices from <u>location "A"</u> to double check that they are booked in the accounting system with the correct amount.</p>	Accuracy		NO
<p><u>Controller "C"</u> reviews <u>all</u> (i.e., 100% of) booked incoming invoices from <u>location "C"</u> to check that they are booked in the accounting system with the correct amount.</p>	Accuracy		NO

Example Multi-Step Controls:

<u>Audit Area:</u> Entity-level risk management.			
Entity-level risk management is a control process aiming to ensure that (1) all relevant risks are identified, <u>and</u> (2) all relevant risks are assessed for impact and likelihood, <u>and</u> (3) all relevant risks are appropriately responded to (i.e., no relevant risks are unidentified, <u>or</u> wrongly assessed, <u>or</u> lack appropriate responses).			
Control Design:	Objective of Control	Operationally Effective? (Audit test result)	
<u>Controller "A"</u> identifies all relevant risks for the entire entity (i.e., for all locations). Identified risks are documented in a "Risk Identification Report".	Entity-Level Risk Management		NO
<u>Controller "B"</u> assesses the potential impact and likelihood of all risks documented in the "Risk Identification Report" (if risks are not documented in the "Risk Identification Report", they are not included in the assessment). Assessments are documented in a "Risk Assessment Report".	Entity-Level Risk Management		NO
<u>Controller "C"</u> manages an entity wide process for deciding upon appropriate responses to all risks that are documented in the "Risk Assessment Report" (if risks are not documented in the "Risk Assessment Report", they do not receive a risk response). Risk responses are documented in a "Risk Response Report".	Entity-Level Risk Management	YES	

Example Amplifying Controls:

Audit Area: accuracy of booked incoming invoices			
Control Design:	Control Objective	Operationally Effective? (Audit test result)	
<u>Controller "A"</u> reviews <u>all</u> (i.e., 100% of) booked incoming invoices from <u>location "A"</u> to check that they are booked in the accounting system with the correct amount.	Accuracy		NO
<p><u>Controller "B"</u> reviews booked incoming invoices from <u>location "B"</u> to check that they are booked in the accounting system with the correct amount:</p> <ul style="list-style-type: none"> ○ Half (i.e., 50%) of the invoices from location "B" are sufficiently controlled by controller "B" <u>alone</u>. <p>The other half (i.e., the other 50%) of the invoices from location "B" require an accuracy control from <u>both</u> controller "B" <u>and</u> controller "C" <u>together</u> (due to special issues like language and GAAP). If only one, or neither, of the controllers perform their part of the control, these invoices cannot be considered booked in the accounting system with the correct amount.</p> <ul style="list-style-type: none"> ○ Controller "B" performs <u>his part</u> of the accuracy control over the other half (i.e., 50%) of the invoices from location "B". 	Accuracy	YES	
<p><u>Controller "C"</u> reviews <u>all</u> (i.e., 100% of) booked incoming invoices from <u>location "C"</u> to check that they are booked in the accounting system with the correct amount.</p> <p>In addition, controller "C" performs <u>his part</u> of the accuracy control over the other half (i.e., 50%) of the invoices from location "B".</p>	Accuracy	YES	

Below each of the 40 cases, the following text occurred:

Control risk is roughly defined as the risk of error after the company has performed controls. Inherent risk has been fixed at 100%, and should not impact your control risk judgment.

(Please note that 100% is maximum risk of error and 0% is minimum risk of error)

Control risk in the audit area is assessed to be: _____ %

Please assess the audit test results again: Does the client (i.e., all locations) have sufficient operationally effective controls for the given audit area?

YES

NO

Appendix 3 – Post Experimental Survey

Demographic Information

1. Name: _____
2. What is your age? _____ Years
3. How many years of experience do you have as an auditor? _____ Years
4. What is your gender? Female ___ Male ___
5. Did you have any difficulties due to these materials being in English? (circle one number)
(None) 1 2 3 4 5 6 7 (Very Difficult)

Internal control knowledge and practice:

6. How would you rate your knowledge of internal controls? (Please circle one number):
(Poor) 1 2 3 4 5 6 7 (Excellent)
7. How familiar are you with ISA/RS 315: Understanding the Entity and its Environment and Assessing the Risk of Material Misstatement (i.e., the “risk standard”)?
(Poor) 1 2 3 4 5 6 7 (Excellent)
8. How familiar are you with Section 404 of the Sarbanes-Oxley Act (SOX)?
(Poor) 1 2 3 4 5 6 7 (Excellent)
9. Have you attended SOX training? YES NO
10. How many hours of SOX 404 experience do you have (if any)? _____ Hours
11. On your engagements, who is usually involved in evaluating entity level controls? (Please circle one or more)

Associate	Senior Associate	Manager	Partner
-----------	------------------	---------	---------
12. On your engagements, who is usually responsible for judgments regarding entity level controls? (Please circle one)

Associate	Senior Associate	Manager	Partner
-----------	------------------	---------	---------

Post-Experimental Questionnaire

General questions:

1. Have you ever seen any of these materials prior to completing them today? Yes No
2. Did you discuss the materials or your answers with other participants? Yes No
3. About how many minutes did it take you to complete this study? _____
Minutes
4. Overall, how do you rate the effort needed to understand and complete the materials?

(Easy) 1 2 3 4 5 6 7 (Difficult)

Framework:

5. How do you rate your understanding of the framework presented in this study?

(Poor) 1 2 3 4 5 6 7 (Excellent)
6. Please circle any elements in the framework that you found difficult to understand (if any):
Amplifying Compensating Substitutable Multi-Step Independent
7. Please circle any elements in the framework that you found difficult to apply (if any):
Amplifying Compensating Substitutable Multi-Step Independent
8. How confident are you with your assessment of controller interaction? (Circle one)

1 2 3 4 5

no little neutral somewhat very
confidence confidence confidence confident confident

Judgments:

9. How confident are you with your assessment of control risk? (Circle one)

1	2	3	4	5
no confidence	little confidence	neutral confidence	somewhat confident	Very Confident

10. How confident are you with your assessment of overall control sufficiency? (Circle one)

1	2	3	4	5
no confidence	little confidence	neutral confidence	somewhat confident	Very Confident

11. If you had not been provided the framework and performed the classification exercises, would you have judged the cases differently?

(Not at all) 1 2 3 4 5 6 7 (Completely)

12. It is of special interest to this study to understand your thought process when you were judging the "YES/NO" question of whether sufficient controls were in place. Please read both options below and think carefully about the way you responded to that question in the survey. Circle the option (below) that best describes your thought process. Both options lead to appropriate judgments, so please think back to how you actually made the judgments in your mind:

- i. I simply looked at the control test results and considered whether there were sufficient controls in each location. That is, in your mind you were not just comparing the overall control risk score of the client to a threshold.
- ii. I simply looked at the percentage score in the client's overall control risk judgment and considered whether this was above or below my threshold. That is, in your mind you were not thinking in terms of sufficiency of controls in locations, but rather in terms of percentage scores. Please indicate the control risk threshold you applied: _____%

Please add any comments you may have regarding how you made the sufficiency of controls judgment:

Understandability:

13. Were the case materials easy to understand? (If no, please explain)

(Never) 1 2 3 4 5 6 7 (Always)

14. Were the case materials realistic? (If no, please explain)

(Never) 1 2 3 4 5 6 7 (Always)

Relevance:

15. On your engagements, do you usually consider control interrelationships?

(Never) 1 2 3 4 5 6 7 (Always)

16. Will your knowledge of the framework improve your future judgments in real world audits?

(Never) 1 2 3 4 5 6 7 (Always)

Thank you once again for participating in this research study.