

[This document contains the author's accepted manuscript. For the publisher's version, see the link in the header of this document.]

The Framing and Evaluation of Multiple Hypotheses

By Theodore J. Mock, Arnold Wright, Rajendra P. Srivastava, Hai Lu

University of Southern California, Northeastern University, University of Kansas,
University of Toronto

Paper citation:

Srivastava, Rajendra. (2008) The Framing and Evaluation of Multiple Hypotheses. *Asia-Pacific Journal of Accounting and Economics*, 15, 123-140.

Keywords:

Audit, Judgement, Framing, Multiple Hypothesis, Evidence

Abstract:

This study provides exploratory evidence on auditors' framing and evaluation of hypotheses, identifies implications for improving audit decision-making and facilitates the interpretation of prior research. Prior studies usually assume hypotheses to be framed as mutually exclusive and exhaustive. However, both verbal protocol evidence and probability assessments reveal that in a realistic case most auditors frame the hypotheses as a non-mutually exclusive and exhaustive set of causes. Further, auditor probability assessments tend to reflect multiple causes. Finally, exploratory analyses indicate auditors have difficulty in updating assessments consistent with the perceived interrelationships between hypotheses.

Asia-Pacific Journal of Accounting and Economics, Vol. 15, 2008, pp. 123–140.

The Framing and Evaluation of Multiple Hypotheses*

Theodore J. Mock, University of Southern California, USA

Arnold Wright, Northeastern University, USA

Rajendra P. Srivastava#, University of Kansas, USA

Hai Lu, University of Toronto, Canada

Abstract

This study provides exploratory evidence on auditors' framing and evaluation of hypotheses, identifies implications for improving audit decision-making and facilitates the interpretation of prior research. Prior studies usually assume hypotheses to be framed as mutually exclusive and exhaustive. However, both verbal protocol evidence and probability assessments reveal that in a realistic case most auditors frame the hypotheses as a non-mutually exclusive and exhaustive set of causes. Further, auditor probability assessments tend to reflect multiple causes. Finally, exploratory analyses indicate auditors have difficulty in updating assessments consistent with the perceived interrelationships between hypotheses.

JEL Classification: M41

Keywords: Audit, judgment, framing, multiple hypothesis, evidence

#Corresponding author: Rajendra P. Srivastava, 1300 Sunnyside Avenue, University of Kansas, Lawrence, KS 66045. Tel: 785-864-7590. Fax: 785-864-5328. Email: rsrivastava@ku.edu.

*We wish to acknowledge the study's participants from a Big Four CPA firm and comments received from Jean Bedard, Stan Biggs, and Gary Monroe, and from participants at workshops at the University of Maastricht, Boston College and the 5th International Symposium on Audit Research. In addition, we express our sincere thanks to Rong-Ruey Duh for providing insightful comments as a discussant of the paper at the 2007 APJAE Symposium on Accounting in Honor of Professor A. Rashad Abdel-Khalik. Finally, we greatly appreciate the comments and suggestions of the editor, and anonymous reviewers.

1. Introduction

The primary purpose of this study is to determine how auditors frame and evaluate a set of multiple hypotheses while assessing the cause(s) of a material financial statement fluctuation in performing analytical procedures. More specifically, we explore whether auditors frame or represent the hypotheses set as mutually exclusive and exhaustive, non-mutually exclusive and non-exhaustive, independent or interdependent or a combination of these relationships. Prior research, to be reviewed in the next section, has assumed participants frame the hypotheses set as mutually exclusive and collectively exhaustive.

A proper understanding of the framing of multiple hypotheses is important, since it can significantly impact decision efficiency and effectiveness. For instance, in an auditing setting failure to consider the full inferential value of the evidence while conducting analytical procedures may result in having to perform additional, unnecessary tests to arrive at a decision as to the cause of the fluctuation (inefficiency) with sufficient comfort. Further, with time constraints the auditor may feel compelled to prematurely terminate testing and arrive at an incorrect conclusion as to the actual cause of the fluctuation (ineffectiveness). Of greatest concern would be an inference that the fluctuation is due to a non-error, when, in fact, a material misstatement due to error or fraud is present and goes undetected.

Given its importance in everyday life and professional settings, there is a growing body of research that seeks to examine if decision-makers are proficient in hypothesis evaluation and whether performance can be enhanced, for instance, through training or decision aids (e.g., Asare and Wright, 1997a). Prior experimental research (e.g., Asare and Wright, 1997a, b) concludes that decision-makers do not update probabilities in accordance with Probability Theory. This

literature suggests that such behavior is primarily the result of cognitive difficulties and thus can lead to suboptimal decision performance. However, an alternative explanation is that the framing of the hypothesis set led to the observed behaviors. We do not wish to impugn these prior studies but rather to highlight that they implicitly assume participants frame hypotheses in a particular manner (mutually exclusive and exhaustive) without explicitly testing this assumption. Our paper shows that auditors do not necessarily frame hypotheses in this way, and analytical research based on Probability Theory has shown that the assumed underlying structure or framing of the hypotheses set significantly affects the manner in which the probabilities of the hypotheses should be updated given evidence (e.g., Srivastava, Wright & Mock, 2002).

The context of the current study entails professional auditors who are investigating the cause or causes of an unexpected change in a company's financial performance. Concurrent protocols provide evidence to explore how auditors evaluate the perceived relationships among hypotheses and then revise probabilities as evidence is obtained. Thus, the framing of hypotheses is revealed. Our findings show that auditors tend to frame the hypotheses as entailing multiple potential causes that are exhaustive and interrelated. These results conflict with the assumptions of prior research (i.e., mutually exclusive) and demonstrate the need in future research to determine the participants' framing of the hypotheses and then look for a pattern of responses that are logically consistent with that framing. Further, auditors are found to have difficulties in revising probability assessments when considering their framing of the hypothesis set, suggesting the need for training or decision aids.

The remainder of the paper is divided into four sections. The next two sections contain a discussion of relevant literature and the research questions, followed by a description of the

method. The empirical results are then presented. The final section is devoted to a discussion of the implications of the findings and future research issues.

2. Relevant Literature and Research Questions

Prior empirical research in both psychology and accounting (e.g., Pennington, 1987; and Chang, Yen and Duh, 2002) has found that framing or problem representation affects hypothesis generation, information search and decision performance. Prior experimental findings also indicate that during hypothesis evaluation individuals predominantly employ a non-complementary revision process where only beliefs for the hypothesis directly implicated by the evidence are revised (referred to as “independent” revision) rather than the beliefs for the other identified hypotheses (e.g., Asare and Wright, 1997a, b).

For instance, Asare and Wright (1997b) provide auditors with a case based on an actual engagement in which a material fluctuation in the gross margin was the result of the client using outdated labor and overhead standard costs. Participants were given five potential causes for the fluctuation as determined by a competent audit team. The task was to perform the hypothesis evaluation phase of analytical procedures by sequentially evaluating 12 pieces of audit evidence that related to potential causes and revising beliefs as to the probability of the various causes after each piece of evidence. The study experimentally framed the hypotheses as mutually exclusive and exhaustive by asking participants to assume the fluctuation was the result of a single cause (error or non-error) and by indicating that the audit team that identified the five causes was highly competent. The findings indicated that auditors largely used a non-complementary, independent revision process. However, complementary behavior was found to increase somewhat as more evidence was obtained.

Asare and Wright (1997a) extend their previous study by examining hypothesis generation, testing, and evaluation. The same case was used. Auditors were again asked to assume the fluctuation was due to a single error or non-error. Participants first generated likely causes for the fluctuation and then selected tests from a list of 17 audit procedures. As in practice, they were provided a time budget to conduct tests but were free to terminate testing when it was considered appropriate and could exceed the budget. The results indicated that the probability associated with the most likely hypothesis initially generated was not related to the number of hypotheses identified; suggesting hypotheses were viewed independently rather than in a complementary (discounting) fashion. During hypothesis evaluation the summed probability revisions were greater than zero, and the more hypotheses that were considered the greater was the summed posterior probabilities, again suggesting non-complementary revision. These findings are consistent with those in the psychology literature employing non-professional participants (e.g., Van Wallendael and Hastie, 1990). Asare and Wright (1997a) suggest an independent hypothesis evaluation appears to reflect a balance between efficiency, defensibility (relying only on direct evidence) and cognitive strain.

Given the experimental framing of the hypotheses as mutually exclusive and exhaustive, independent revision appears contrary to prescriptions of Probability Theory. However, independent revisions are hypothesized to be the result of cognitive difficulties (see, e.g., Asare and Wright 1997a, b). At the same time, these studies do not capture the underlying assumed framing of the hypotheses, which, as discussed in the section to follow, affects the appropriate manner in which probabilities should be revised.

2.1 Theoretical Framework

To analyze probability judgments regarding the presence of multiple hypotheses (causes) when an effect (e.g., a material fluctuation in an audit client's financial statements) is observed, let us consider a simple case where two hypotheses ($H1$ and $H2$) are the sole causes of an effect, E . According to Morris and Larrick (1995) this means that the conditional probabilities $P(E|H1) = 1$ and $P(E|H2) = 1$ given that $H1$ and $H2$ are the sole causes of the effect. This condition yields: $E = H1 \cup H2$. This case can be expressed as follows:

$$P(E) = P(H1) + P(H2) - P(H1 \cap H2)$$

and the posterior probabilities after observing the effect can be written by using Bayes' rule as:

$$P(Hi | E) = \frac{P(Hi \cap E)}{P(E)} = \frac{P(Hi)}{P(H1) + P(H2) - P(H1 \cap H2)}, \quad i = 1, 2$$

The above values yield:

$$P(H1 | E) + P(H2 | E) = \frac{P(H1) + P(H2)}{P(H1) + P(H2) - P(H1 \cap H2)} \quad (1)$$

As one can see from equation (1), the sum of the posterior probabilities for the two hypotheses is always greater than one ("super-additive") unless $P(H1 \cap H2) = 0$ (i.e., $H1 \cap H2 = \emptyset$, the two hypotheses are mutually exclusive) in which case the sum is one ("additive").

The above results make intuitive sense. For instance, suppose the auditor determines that an increase in the gross margin for the current year is definitely (conclusive evidence) the cumulative result of two causes: overstated sales as well as overstated inventory. In this case, the probability that sales are overstated given the evidence is equal to or close to 1, and the probability that inventory is overstated is also equal to or close to 1. Since both of these causes

have co-occurred ($P(H1 \cap H2) = 1$), equation (1) indicates the summed probabilities are equal to or close to two. This result is not contrary to Probability Theory. It simply means that the two hypotheses can co-exist. Such situations, for instance, are common when fraud is suspected, as discussed later. In a general situation where more than two causes of an effect are expected and not all the causes are known, one can show that the sum of posterior probabilities of the known hypotheses can be equal to 1, less than 1 (sub-additive), or greater than 1 (super-additive), depending on whether the hypotheses are mutually exclusive or interrelated, and whether the evidence is strong or weak. However, if the hypotheses are exhaustive, the sum of the posterior probabilities must be equal to or greater than one (see Srivastava et al. (2002) for details).

2.2 Research Questions

The above discussion shows that in some cases probability revisions logically can be non-complementary and also additive, super-additive, or sub-additive. Therefore, it is not clear whether the non-complementary belief revisions found in prior studies are the result of cognitive difficulties, as posited, or are the result of revisions that are based on a different representation or framing of the hypothesis set than that assumed by researchers. This latter possibility is noted by Van Wallendael and Hastie (1990, p. 249):

"It might be, of course, that the 'real world' encourages such a view; most hypothesis testing situations do not involve clear sets of mutually exclusive and exhaustive possibilities...Treating the hypotheses as independent is not necessarily suboptimal if one does not have an exhaustive set, and if participants are accustomed to representing real-world hypotheses as independent entities, they may continue to use such a representation even when other representations are appropriate and more efficient".

In actual client settings, auditors may perceive situations in which there is a single cause of a fluctuation or there are multiple causes. A single cause indicates a mutually exclusive setting, since by definition more causes cannot co-occur. A belief that there may be multiple

hypotheses (more than one cause), on the other hand, suggests either non-mutually exclusive (joint causes) or independent framing. Prior experimental research has directed participants to assume there is only a single cause (e.g., Van Wallendael and Hastie 1990). However, how often is there a single cause in a realistic setting such as in auditing? Further, there may be situations where the decision-maker is confident that all of the likely causes are identified (exhaustive) or other situations where unidentified causes may be present (non-exhaustive). The current study provides exploratory evidence on the prevalence of each of these framing situations.

Evidence of non-mutually exclusive framing would be indicated by unidirectional and independent revisions in probabilities. For instance, if probabilities are increased for two hypotheses based on confirming evidence for one of them, this implies the two are viewed as positively correlated and thus likely to be co-occurring. In contrast, as discussed, if hypotheses are framed as mutually exclusive, one would expect complementary revisions. To date, little empirical evidence exists as to how hypotheses are actually framed by decision-makers, which leads to the following research question:

RQ 1: How do auditors frame hypotheses in considering the cause(s) of a material, unexpected financial statement fluctuation? More specifically, are these assessments:

- **Based on an assumed exhaustive set of hypotheses?**
- **Based on an assumed single or multiple set of hypotheses?**
- **Are the hypotheses thought to be independent, complementary or interdependent? If hypotheses are thought to be interdependent, how are they thought to be interrelated (e.g. negatively or positively correlated)?**

2.3 Consistency between Hypothesis Framing and Evaluation

Given, as discussed earlier, that the way a decision-maker should assess evidence depends on how the task is framed, it is important to determine how auditors actually represent

hypotheses under various realistic decision contexts. Research cannot evaluate whether hypothesis evaluation is performed in a manner consistent with an individual's representation if that representation is unknown. If deficiencies in the manner or magnitude of evaluations are found given the representation, training (or decision aids) may be utilized to improve performance.

For example, McKenzie (1998) reports that participants who are trained to represent causes as mutually exclusive in evaluating evidence (“contrastive” learning) demonstrate a complementary revision of beliefs. In contrast, participants trained to use “non-contrastive” learning exhibit non-complementary revisions and super-additivity. The consistency between decision-makers' hypothesis framing and evaluation provides evidence of whether cognitive difficulties in revisions are present, as suggested by prior studies, and is examined in the second research question:

RQ 2: Are auditors' probability revisions logically consistent with their perceived framing of the hypotheses?

3. Research Method

We used three data sources to address the research questions: process-tracing (verbal protocol analysis) evidence, probability revisions for the hypotheses under consideration and perceived interrelationships in a post-case questionnaire. The verbal protocol method provides rich decision-process evidence as auditors go through the task of considering multiple hypotheses, providing an unobtrusive means to examine how they frame and evaluate the hypotheses (see e.g., Ericsson and Simon 1993). In contrast, prior experimental studies provide little information on the underlying judgment processes regarding the assumed nature and relationships between the hypotheses and the evaluation of evidence. An inherent trade-off of

using the verbal protocol approach, however, is that it entails very detailed, costly data collection and analysis; thus, the sample size is usually small.

3.1 Case-study Task and Elicitation Procedures

The audit task entailed a follow-up investigation of a material, unexpected fluctuation while performing analytical procedures. The setting is one of increasing importance as auditing firms place greater reliance on knowledge of the client and analytical procedures. Since prior research (Abdolmohammadi, 1999) indicates that audit seniors and managers typically perform this task, we sought the participation of experienced auditors at these levels. A Big Four audit firm identified nine auditors (five managers and four experienced seniors) to participate. Due to the failure of recording equipment during participant 2's session, we lost part of the protocol data and thus considered only eight participants' results in the present study. Participants averaged 5.6 years of experience with some background in manufacturing industry.

Given that one of the objectives of this research is to build upon prior studies, the experimental case was a modified version of the case used by Asare and Wright (1997a, b). The main differences are that we employ a verbal protocol approach rather than an experiment; we do not require participant assessments to be mutually exclusive and exhaustive; we ask the participants to assess fewer items of evidence; and we use the version where there is an increase in the gross margin. While Asare and Wright's earlier work looks at the actual case from practice where there was a decrease in the gross margin, they (2003) later modify the case and evidence to be reflective of an increase in the margin, which is a riskier, more interesting setting.

Both the listed differences from Asare and Wright's earlier work and the inherent complexity of studies based on real audit cases preclude the derivation of a clear normative solution as to the appropriate framing of the hypothesis set. Thus, we cannot determine which

form of hypothesis framing is “correct”. However, the primary objective of this study is to provide exploratory evidence on the nature of and diversity in which auditors frame hypotheses in a realistic setting.

The main tasks require auditors first to assess the probability that each of the five inherited hypotheses led to the material, unexpected increase in the gross margin and then revise these probabilities based on additional items of evidence from five audit tests. The verbal protocol elicitation procedures involve four parts: instructions; a practice exercise; the Precision Measurement, Inc. case; and a debriefing questionnaire (the research instrument is available from the authors).

The session began with one of the researchers reviewing the page of instructions that provided an overview of the task. The instructions stated that the study's purpose was “to examine how auditors evaluate evidence in investigating material, unexpected fluctuations encountered while conducting analytical procedures”; stated that we were “asking individuals to ‘think aloud’ as they perform a task”; and asked them to “*verbalize whatever* you are thinking about *no matter how trivial, indiscreet, or harsh*”. The instructions concluded by stating that the responses were anonymous and that although the researcher would remain in the room to operate the recording equipment, he could not answer any substantive questions once the main exercise began.

The main task was an investigation of a material, unexpected fluctuation in the financial statements of Precision Measurement, Inc., an audit client. The case materials (13 pages in total) included: general information about the client, the engagement, and the prior year’s audit; a summary assessment of the control environment; an assessment of materiality; and comparative

financial information for the most recent two years including key financial statement ratios. The participants were then informed that an audit team had identified a material unexpected increase in the gross margin ratio.

Participants were then given the results of the audit staff's investigation of the fluctuation presented as follows (emphasis and hypothesis numbering added):

Assume that the staff that did the preliminary investigation is quite competent. Based on knowledge of the client and after consulting among themselves and with the client's personnel, the audit staff has identified the following potential causes:

- 1. Purchases cutoff errors have led to current year's purchases (correctly included in ending inventory) being excluded from the purchase records. As a result, purchases/cost of goods sold is understated (**H1: purchase cutoff errors**).*
- 2. There has been a shift in sales mix toward high margin items. As a result, gross margin for the current year has increased (**H2: shift in sales mix**).*
- 3. Standard costs in ending inventory have not been updated to reflect decreased actual costs. As a result, ending inventory is overstated (**H3: standard costs not updated**).*
- 4. Price competition among vendors has led to a decrease in the price of raw materials and other items used in production. As a result, cost of goods sold has decreased (**H4: decrease in raw material prices**).*
- 5. A significant portion of administrative costs has not been properly allocated to inventory costs. As a result, cost of goods sold is understated (**H5: misallocating administrative costs**).*

Participants were then asked to "assess the probabilities of the five potential causes in having led to the material unexpected increase (fluctuation) in the gross margin" using a scale of 0 to 100% where 0% probability means "certain the cause did not lead to the fluctuation" and 100% probability means "certain the cause led to the fluctuation". The instructions and response scale were thus designed to evoke a probability assessment and only indirectly entail judgments as to the most likely cause(s), i.e. probability assessments are the underlying means in which an auditor eventually decides on the most likely cause(s).

In addition, auditors were provided with the option of assigning some probability to “another cause(s) that is (are) not yet identified”. Some prior studies restrict the summed probability assessments to be equal to 1.00 (e.g., Heiman, 1990). However, as discussed previously, there are likely to be many cases where summed probabilities may normatively be posited to be super-additive or sub-additive. Thus, our use of a non-restrictive probability assessment is more appropriate to address the research questions in an unconstrained, natural setting.

Participants’ revision of probabilities is used to examine RQ 1 (framing of the hypothesis set). The issue is whether the participant frames the setting as one with a single or multiple causes. While all evidence evaluations are at least partly hypothesis evaluation, it is appropriate to consider single vs multiple hypothesis revision as evidence of a type of framing since the manner in which one uses evidence would depend on the type of framing.

The final part of the main task involved reviewing audit evidence that was obtained to test each of the potential causes (hypotheses). Following each test result, the participants once again assessed the probability that each cause resulted in the fluctuation of the gross margin. For instance, Test 1, relating to the shift in sales mix hypothesis (H2), was reported as follows:

Test 1: The audit staff reports completion of the following audit procedure: Consult with the marketing manager about the extent that a shift in sales mix and marketing strategies may have led to the fluctuation in gross margin.

Result: The marketing manager indicates there has been a shift to selling more products with relatively high gross margin rates, especially in the European market. She indicates that “As you know, this is an industry-wide phenomenon. This shift was especially noticeable in the last quarter of the year and we expect to reap its full impact in the coming year.”

Five audit test results were given in the same order to each participant. Based on pilot testing, Asare and Wright (1997a) determined that experienced auditors assessed the inferential

value of the audit evidence as follows: T3 disconfirms H1, T5 disconfirms H4, T4 disconfirms H5, T1 confirms H2 and T2 confirms H3. Confirming (disconfirming) evidence is expected to lead a decision-maker to increase (decrease) the assessed probability of a particular hypothesis. If the auditor believed that the evidence was also relevant to other hypotheses, adjustments in the probabilities of other hypotheses would be made.

3.2 The Practice Exercise and Debriefing Questionnaire

Following the initial instructions, auditors were asked to complete a two-page practice exercise designed to give them experience with verbal protocol methods. The practice task was unrelated to the experimental task and asked participants to evaluate the propriety of classifications of leased machinery and land acquisition costs and determine the need for audit adjustments or reclassifications. Once auditors seemed comfortable with the elicitation process, they were asked to complete the main task.

Following the main task, each individual was given a debriefing questionnaire where they were asked whether it was likely that there were multiple causes for the fluctuation and the level and direction of interrelationships between the hypotheses (positive correlation, negative correlation, or no correlation). Auditors also assessed the materiality of the gross margin fluctuation, as intended, to be from somewhat material to highly material with no one evaluating it as being immaterial.

4. Results

4.1 Protocol Coding and Inter-Coder Agreement

The protocols were coded twice. The objective of the initial coding was to provide descriptive information about the nature of the decision process, while the second coding was accomplished to address the research questions. In the initial coding one of the researchers parsed the protocols to identify basic information operations (moving from one knowledge state

to another, in Ericsson and Simon's terminology) and each of these was coded by two coders and reconciled. This coding used essentially the same operators as Biggs and Mock (1983). The level of inter-coder agreement was 76.3%.

For the second coding two researchers with experience in verbal protocol analysis independently coded the verbalizations from the auditor participants to identify “segments”. A segment is defined as one or more information operations that relate to a research question of interest (e.g. verbalizations that indicated whether the hypotheses were, at that point in the decision process, viewed as mutually independent or were exhaustive). To accomplish this, a coding sheet was developed that listed each research question as well as exemplars of verbalizations that would provide direct evidence on the various research questions (the coding scheme is available on request from the authors). Each coder read a single or a series of verbalizations that reflected the respondent's stream of thoughts indicating the nature of the framing of hypotheses. Given the complexity of framing, a single phrase / single operator was usually not sufficient to identify the nature of the framing of hypotheses as one can see from the example relating to RQ 1 (see the coding scheme). The level of agreement for the second coding was 82.5% (Cohen's Kappa coefficient .669; $p < .01$).¹ The two coders then reconciled all disagreements; the reported results reflect the reconciled coding.

4.2 Findings Regarding the Decision Process

Detailed coding of the operations being performed within the assessment process provides evidence regarding the participant's underlying decision process based on Newell and Simon's theory of human problem solving; provides process information that facilitates

¹ Note that this analysis goes beyond what is normally done in coding operators in that it was necessary to identify a series of information operations for coding. Cohen's Kappa coefficient is not relevant for the initial coding, since it

comparison of this task to other accounting and auditing studies and enables an assessment of the details of the captured decision process. The observed behavior at the operator level indicates a complex decision task involving up to 366 operations to complete. The basic nature of this process is consistent with prior auditing studies (e.g. Biggs and Mock, 1983) with major proportions of the decision process involving overall information acquisition (24%), evaluation of information (29%), generation of queries (13%) and audit decisions (15%).

A review of the detailed coding of the protocols revealed that the task involved three basic phases: 1) Client information acquisition; 2) Analytical review of financial statement data; and 3) Completion of the main decision task, which involved providing a series of assessments of the probability that the five hypotheses suggested by the audit staff caused the observed change in a client's gross margin. These decision phases are typical of those observed in prior audit research. Phases 1 and 2 are essentially "pre-decisional" and provide the client background information needed to complete the task. The main research questions focus on behaviors during the third phase, the focus of the following discussion.

4.3 Summary of Probability Assessments

The descriptive results (data available upon request) regarding the initial, *prior* probability and subsequent *changes* in probability judgments after receiving the results of five audit tests obtained from each subject reveal interesting findings. Basically, subjects provided their probability assessments after evaluating each test results for the five given hypotheses and for the sixth hypothesis labeled as "other hypothesis", in case they thought it to be possible. The aggregate change over all subjects in the probability of each hypothesis shows a substantial amount of disagreement in the net effect of audit tests 1 and 2, but general agreement in the net

is only intended to provide descriptive information concerning the nature of the decision process for comparison to

effects of tests 3, 4 and 5. Overall, the net change in probabilities indicates that the summed probabilities are rarely complementary (i.e., equaling or approaching 0), corroborating the findings of prior experimental studies (e.g., Asare and Wright, 1997a, b; Van Wallendael and Hastie, 1990; Van Wallendael, 1989). The key issue, as discussed at length earlier, is the proper interpretation of this finding, which on the surface appears contrary to Probability Theory. The objective of the research questions is to more fully understand the underlying decision process surrounding the framing and evaluation of the hypothesis set. In addressing these questions three forms of evidence are considered: segments within the verbal protocols; probability assessments; and probability revisions.

4.4 Results Concerning the Research Questions

RQ 1: Framing of the Hypothesis Set As Exhaustive (Non-Exhaustive)

Table 1 (Panel A) compiles the evidence provided by the verbal protocols on RQ 1. As indicated in row one of this table, there were relatively few protocol segments (only 10 of the 82 coded segments that evidenced the framing being utilized) indicating whether auditors considered the hypothesis set as exhaustive (2) or non-exhaustive (8). To illustrate, the following are protocol segments that were coded as indicating an exhaustive or non-exhaustive set of hypotheses:

Exhaustive

Participant 6: *Other causes; We'd have to think about; Again we said that we had competent staff that went out and found our potential causes; So, we'll rely on that...*

Non-exhaustive

Participant 3: *Other causes; I'd say it's possible that there could be something else*

prior protocol studies and does not examine the research questions.

Participant 7: “ ... *certainly there is always the possibility of other causes and I think it is our best interest to be cognizant of other causes*”

[Insert Table 1 around here]

Evidence of exhaustive (non-exhaustive) hypothesis framing is also found when there is a zero (non-zero) probability or weight attached to the category of “other causes”. The results indicate that only 3 of the 8 participants assigned a non-zero probability value for H6, ‘Other Hypotheses,’ and in only 4 of the 48 assessments (8.3%) were the 5 inherited hypotheses assumed to be non-exhaustive. These results, thus, suggest the predominant framing condition entailed an exhaustive set. On the surface this finding seems to be inconsistent with the protocols, which suggest non-exhaustive framing. However, there were relatively few protocols that expressed views regarding whether the hypothesis set was considered exhaustive or non-exhaustive. Further, the auditors’ protocols that were present suggest that this difference in the findings may be because while the possibility of other causes is recognized through training and skepticism and thus verbalized within the protocols, they are willing to rely on a competent audit team in identifying likely causes. However, as expressed by participant 6 (as quoted above), if subsequent evidence then suggests another cause, this would be followed up.

RQ 1: Framing of the Hypothesis Set As Multiple or Single Probability Revisions

Tables 1 and 2 provide evidence related to this part of Research Question 1. For example, Table 1 (Panel A) compiles the frequency of single versus multiple probability revisions evident in the verbal protocols. Examples within the verbal protocols of these types of behavior are as follows:

Complementary or Partially Complementary

Participant 6: *so those are kind of interrelated, so maybe I'll put a 50 there and 25 on the sales mix.*

Independent

Participant 3: *So I guess at this point I'd say that shift is probably 10%. The rest I'm probably going to leave the same until I get more information*

Table 1 (Panel A) shows that multiple revisions were dominant within 24 protocol segments (all 8 participants), indicating complementary or partially complementary revisions. Within only 3 protocol segments was there evidence of single (independent) revisions. The results in Table 1 are also consistent with this interpretation with frequent revisions of multiple hypotheses (ranging from 25% to 75% of the assessments).

Table 1 (Panel B) classifies the 40 decisions (the 8 auditors' assessments after they examined the 5 audit test results) into 4 possibilities: no change from their prior assessments; revision of a single hypothesis; complementary or partially complementary multiple revisions; and unidirectional multiple revisions. As indicated, in many cases (45%) multiple revisions occur; either complementary or partially complementary revisions are present 30% of the time with unidirectional revisions 15% of the time. In all, both the verbal protocol and probability revision findings suggest auditors often revise probabilities for multiple hypotheses, albeit not in the majority of cases.

RQ 1: Framing of the Hypothesis Set As Mutually Exclusive (Non-Mutually Exclusive)

Table 1 (Panel A), row 3, provides the findings on the frequency of verbalizations suggesting mutually exclusive or non-mutually exclusive framing. There is strong evidence that the pervasive framing condition was non-mutually exclusive. Specifically, within 44 protocol segments (all 8 participants) there was evidence of verbalizations that indicated the auditors considered the hypotheses as not mutually exclusive. In contrast to the predominant framing

assumed in previous experimental studies, we found only 1 protocol indicating mutually exclusive framing. Examples of such protocol segments are noted below.

Not Mutually Exclusive

Participant 1: *I still think there's going to be something else contributing to this fluctuation in addition to just the shift in the sales mix.*

Participant 5: *I think it's probably going to be things that are changes in the business. Probably either a shift in the sales mix. Or a decrease in the raw materials prices. And probably not an error from either the allocation of administrative; Costs...or purchase cutoff errors.*

As shown in Table 1 (Panel B), the predominant pattern of probability revisions also suggests non-mutually exclusive framing. That is, 52.5% of the revisions are either independent (37.5%) or unidirectional (15%). In contrast, only 30% of the revisions entailed complementary revisions suggestive of mutually exclusive framing. This finding is consistent with that found in the verbal protocols reported previously.

Recall that Table 2 provides detailed information of changes in the assessed probabilities of each hypothesis as an additional audit test result was being considered. The shaded cells indicate the expected direct effects of the audit evidence. In most of the shaded cells, the effects are either as expected or there was no change.

The un-shaded cells show indirect effects of the evidence. For example for the hypothesis concerning a shift in sales mix ($H2$), there are 8 times where the probability assessments were changed when test results directly relating to other hypotheses were being considered. These changes reflect either complementary behavior, or, more likely, other perceived interdependencies among the hypotheses and the evidence being assessed. We observe similar interdependent assessments for $H1$ (4 times), $H3$ (7 times), $H4$ (12 times) and $H5$ (8 times).

RQ 2: Consistency between Perceived Interrelationships and Probability Assessments

The second research question examines whether the interdependent probability assessments found are logically consistent with the interrelationships between hypotheses perceived by auditors. Following the main task, each individual was given a debriefing questionnaire where they were asked to indicate which hypotheses were interrelated, and if so, whether they were related in a positive or negative manner. This evidence is considered to be exploratory in nature, since it was gathered after the task and is, thus, subject to potential concerns about the accuracy of self-insight and the difficulties of elicitation for the complex judgments studied.

Table 2 (Panel A) reports the perceived interrelationships noted by participants in the post-case questionnaire. As shown, a number of interrelationships were noted. In particular, 6 of the participants viewed *H3* (standard costs not updated) and *H4* (decrease in raw material prices) as interrelated, and 4 individuals identified *H2* (shift in sales mix) and *H4* as well as *H3* and *H5* (misallocating administrative costs) as interrelated. In all, these observations suggest that the majority of participants framed the hypotheses as non-mutually exclusive and, as discussed in the prior section, exhaustive.

Table 2 (Panel B) provides a summary of the consistency between the interrelationships identified in the debriefing questionnaire and the pattern of actual probability assessments. The first issue is what behavior represents a consistency or an inconsistency. For example, if a participant indicates an interrelationship between *H1* and *H2*, consistency would mean the probability of *H1* should be updated when evidence concerning *H2* is received *and vice versa*. In addition, behavior could be judged consistent for each case where no relationship was specified

in the questionnaire and in fact the individual appropriately made no change as evidence was obtained. Table 2 (Panel B) documents both types of consistency.

Table 2 (Panel B), column 2 (3) lists the number of interrelationships (no interrelationship) that were identified by each auditor for the 20 pairs of hypotheses. The next 4 columns show the frequencies that the auditor decisions were consistent or inconsistent with each category. The results suggest low consistency between the actual revisions and the identified interrelationships (only 16.7% or 7 of the 42 possible cases) where an interrelationship was indicated in the debriefing questionnaire. However, the consistency measure is much higher (58.5% or 69 of 118 cases) for the situations where the auditors felt there were no interrelations between the various hypotheses. This finding is consistent with the notion that cognitive complexity affects decision performance. For the more complex, interrelated hypotheses, auditors apparently had more difficulty evaluating evidence in a manner consistent with this framing of the hypotheses.

[Insert Table 2 about here]

5. Summary and Conclusions

An important feature of the evidence gathered from the protocols obtained in this exploratory study is that it provides a detailed trace of how auditors frame the hypotheses and of the thought process occurring as probabilities are generated and then revised given audit test results. Prior experimental studies do not capture such data and have assumed, perhaps inappropriately, that hypotheses are represented as mutually exclusive and exhaustive when, in fact, they may not be. It is essential to obtain decision process data to properly interpret the findings of prior studies, which suggest auditors have difficulties in completing this task, and to

then move forward in investigating the need for and efficacy of promising ways to improve performance.

The findings from verbal protocols, actual probability assessments, and the post-test questionnaire indicate that auditors completing the task often viewed the hypotheses as interrelated. Also, as audit evidence was being reviewed, participants predominantly adjusted their assessments for multiple hypotheses. For these auditors, the findings are contrary to prior experimental results (e.g., Asare and Wright, 1997a, b), which show a largely independent, single-hypothesis, revision process where interrelations do not affect the observed audit judgments. Interestingly, auditors believe multiple causes are possible even after disconfirming evidence on three of the causes has been obtained and only weak confirming evidence of a fourth hypothesis is presented. The findings may mirror the realistic nature of the modified case where no explicit guidance is provided as to the nature of the hypothesis set vis-à-vis prior studies that suggest the hypotheses are likely to be mutually exclusive and exhaustive.

The majority of the participants were willing to assume that the set of inherited hypotheses were exhaustive. Thus, they evidently relied on their subordinates' efforts to identify an exhaustive set of possible causes to consider or they felt it was cost-effective to assume there were no other, unidentified, causes. However, many auditors expressed the thought that other causes were possible within their protocols. Finally, exploratory evidence suggests auditors have difficulties in revising probabilities in a manner that is logically consistent with their framing of the hypothesis set, especially when this framing is complex (e.g., perceived interrelationships).

In all, the findings provide initial evidence in a realistic setting indicating auditors often frame hypotheses as non-mutually exclusive and exhaustive. This result is contrary to prior

experimental research that explains non-complementary probability revisions as evidence of behavior inconsistent with Probability Theory, based on an explicit assumption that hypotheses are framed as mutually exclusive and exhaustive. The results underscore the importance of determining the framing adopted by the decision-maker in interpreting prior research.

The findings of the current study have significant implications in interpreting prior studies and demonstrate the importance for future studies to explicitly control for the hypothesis framing adopted by participants. One way in which this can be accomplished is by making the framing explicit in an experiment and then employing manipulation check questions to ensure this framing is accurately encoded by participants. Another approach would be to provide training that would suggest a certain framing and again verify this via manipulation checks. Once the framing condition is determined, performance can then be compared to normative benchmarks.

The findings also have important implications for auditing practice, given the potential significant impact on decision effectiveness and efficiency of such common tasks. Exploratory findings in this study suggest auditors may have difficulties revising probabilities in a manner consistent with the manner in which the hypotheses are framed. If deficiencies in performance are corroborated in follow-up research, the efficacy of methods to assist individuals such as training or decision tools can be explored.

As with all empirical studies, this exploratory research is constrained by limitations and raises many research issues that need to be addressed. While use of a verbal protocol approach provides rich decision-process evidence, it also inevitably results in a relatively small sample size. Further, protocols may not capture all elements of the decision process. For instance, some

cues and decisions may be so familiar and automatic that they may not be verbally expressed. Nonetheless, verbal protocol analysis provides much more detailed evidence of the decision process than other methods such as experiments or archival data. In addition, the order of evidence presentation may have influenced behavior. Finally, we implicitly assume that the framing of the hypothesis set is stable. Yet, at the beginning, the auditors may frame the task as an independent and exhaustive scenario, but after audit tests have been performed, their perceptions may change to assuming a non-exhaustive scenario or vice versa. Future research is needed to address such decision process dynamics.

We use three data sources to address our research questions. We observe few inconsistencies between the protocols and the probability assessment data. As noted, this is an initial descriptive study of how auditors frame and evaluate hypotheses for a particular realistic task. To reconcile the inconsistency between the probability assessment and the data from post-experiment questionnaires, we need further corroborating work. For example, in some settings auditors may assume a small set of exhaustive causes that are mutually exclusive whereas in others they may consider a large set of interrelated causes.

Probability assessments may also be affected by a number of behavioral factors including effects of cognitive style (e.g. Fuller and Kaplan, 2004), cognitive limitations or information processing differences such as order effects. Researchers may also wish to examine other factors affecting hypothesis evaluation including: evidence strength; evidence implicating multiple causes; or the need to generate new hypotheses.

Lastly, further analytical research is needed to determine how probability assessments should vary depending on the framing of the hypothesis set. The work of Morris and Larrick (1995) and Srivastava et al. (2002) are a first step in this regard.

REFERENCES

- Abdolmohammadi, M., 1999, "A Comprehensive Taxonomy of Audit Task Structure, Professional Rank, and Decision Aids for Behavioral Research", *Behavioral Research in Accounting* 11, 51–92.
- Asare, S. and A. Wright, 1997a, "Evaluation of competing hypotheses in auditing", *Auditing: A Journal of Practice & Theory* 16, 1, 1–13.
- Asare, S. and A. Wright, 1997b, "Hypothesis revision strategies in conducting analytical procedures", *Accounting, Organizations & Society* 22, 8, 737–755.
- Asare, S. and A. Wright, 2003, "A note on the interdependence of hypothesis generation and information search in conducting analytical procedures", *Contemporary Accounting Research* 20, 2, 235–251.
- Biggs, S.F. and T.J. Mock, 1983, "An investigation of auditor decision processes in the evaluation of internal controls and audit scope decisions", *Journal of Accounting Research* 21, 1, 234–55.
- Chang, C.J., S.-H. Yen and R.-R. Duh, 2002, "An empirical examination of competing theories to explain the framing effect in accounting-related decisions", *Behavioral Research in Accounting* 14, 1, 35–64.
- Ericsson, K.A. and H.A. Simon, 1993, *Protocol analysis: Verbal reports as data*, rev. ed., (Cambridge, MA: MIT Press).
- Fuller, L.R. and S.E. Kaplan, 2004, "A Note about the Effect of Auditor Cognitive Style on Task Performance", *Behavioral Research in Accounting* 16, 1, 131-143.

Heiman, V., 1990, "Auditors' assessments of the likelihood of error explanations in analytical review", *The Accounting Review* 65, 4, 875–890.

McKenzie, C., 1998, "Taking into account the strength of an alternative hypothesis", *Journal of Experimental Psychology* 24, 771–792.

Morris, M.W. and R.P. Larrick, 1995, "When one cause casts doubt on another: A normative analysis of discounting in causal attribution", *Psychological Review* 102, 2, 331–355.

Pennington, N., 1987, "Stimulus structures and mental representation in expert comprehension of computer programs", *Cognitive Psychology*, 295–341.

Srivastava, R., A. Wright and T.J. Mock, 2002, "Multiple hypothesis evaluation in auditing", *Accounting and Finance* 42, 3, 251–278.

Van Wallendael, L., 1989, "The quest for limits on noncomplementarity in opinion revision", *Organizational Behavior and Human Decision Processes* 44, 3, 385–405.

Van Wallendael, L. and R. Hastie, 1990, "Tracing the footsteps of Sherlock Holmes: Cognitive representations of hypothesis testing", *Memory and Cognition* 18, 3, 240–250.

Table 1: Verbal Protocol Results

Panel A: Evidence from Coded Protocol Segments on How the Hypotheses Were Framed (RQ 1)

Protocol segments indicating Exhaustive or Non-Exhaustive Framing (10 segments in total)	<u>Exhaustive</u> 2 coded segments (2 participants*)	<u>Non-Exhaustive</u> 8 coded segments (5 participants*)
Protocol segments indicating Revisions for Single or Multiple Causes (27 segments in total)	<u>Multiple Revisions</u> 24 coded segments (8 participants*)	<u>Single Revision</u> 3 coded segments (2 participants*)
Protocol segments indicating Mutually Exclusive (Single Cause) or Not (45 segments in total)	<u>Mutually Exclusive Revision</u> 1 coded segment (1 participant*)	<u>Non-Mutually Exclusive Revisions</u> 44 coded segments (8 participants*)

* Number of participants that exhibited indicated behavior.

Panel B: Frequency of Types of Revisions of Probability Assessments (RQ 1)

Evidence of Frequency of Type of Revision Based on 40 Revisions (100%) and Number of Participants Making that Revision at Least Once	
<u>Single (Independent) Revision</u> 15 revisions (37.5%) (7 participants)	<u>No Revision (No Change)</u> 7 revisions (17.5%) (4 participants)
<u>Multiple Revisions</u> Complementary or Partially Complementary 12 revisions (30%) (6 participants)	<u>Multiple Revisions</u> Unidirectional 6 revisions (15%) (5 participants)

Table 2: Interrelationships Identified

Panel A: Number of Participants Indicating an Interrelationship Among the Hypotheses

	<i>H1</i> : Purchase cutoff error	<i>H2</i> : Shift in sales mix	<i>H3</i> : Standard costs not updated	<i>H4</i> : Decrease in raw material prices	<i>H5</i> : Misallocating administrative costs
<i>H1</i> : Purchase cutoff errors					
<i>H2</i> : Shift in sales mix					
<i>H3</i> : Standard costs not updated	2	1			
<i>H4</i> : Decrease in raw material prices	2	4	6		
<i>H5</i> : Misallocating administrative costs		1	4		

Panel B: Number of Interrelationships Identified in the Questionnaire and Consistency of Probability Assessments with These Interrelationships

	Total number of pairs of causes that were considered interrelated and not interrelated in the questionnaire		Judgments <u>consistent</u> with identified interrelationships		Judgments <u>inconsistent</u> with identified interrelationships	
	Interrelated	Not Interrelated	Interrelated	Not Interrelated	Interrelated	Not Interrelated
<i>P1</i>	4	16	1	12	3	4
<i>P2</i>	4	16	2	16	2	0
<i>P3</i>	2	18	0	15	2	3
<i>P4</i>	6	14	0	11	6	3
<i>P5</i>	4	16	3	8	1	8
<i>P6</i>	6	14	0	0	6	14
<i>P7</i>	10	10	1	0	9	10
<i>P8</i>	6	14	0	7	6	7
Totals	42	118	7	69	35	49