# Multiple hypothesis evaluation in auditing

Rajendra P. Srivastava[a], Arnold Wright[b], Theodore J. Mock[c]

[a]*Ernst & Young Center for Auditing Research and Advanced Technology, The University of Kansas, Lawrence, KS 66045, USA*
[b]*Boston College, Carroll School of Management, Chestnut Hill, MA02467, USA*
[c]*Leventhal School of Accounting, University of Southern California, Los Angeles, California 90089, USA*

## Abstract

In many audit tasks, auditors evaluate multiple hypotheses to diagnose the situation. Research suggests this is a complex task that individuals have difficulty performing. Further, there is little guidance in professional standards or literature dealing with the many complexities present in the audit environment. Using probability theory, this study derives the appropriate revision of likelihoods for multiple hypotheses given different realistic audit conditions. The analysis shows that the relationships among the hypotheses dramatically impact the use of audit evidence and the resulting pattern of probability revisions. We also identify testable hypotheses to guide future research and discuss practice implications regarding ways to improve the effectiveness of analytical procedures.

*Key words*: Analytical procedures; Multiple hypothesis evaluation; Audit evidence aggregation

*JEL classification*: M 49

## 1. Introduction

Auditors implicitly or explicitly generate and evaluate multiple hypotheses on all audit engagements. They may generate several hypotheses concerning various possible misstatements in the financial statements during the planning stages of an audit as a result of their preliminary assessments of inherent and control risks. For example, the existence of a compensation scheme based on reported earnings might generate several hypotheses concerning possible causes

of potential misstatements in sales, cost of goods sold and other expenses. Auditors may also investigate potential causes (hypotheses) of a material unexpected financial statement fluctuation while conducting analytical procedures and often consider multiple reasons for discrepancies found in evidence obtained such as exceptions to receivable confirmations. In addition, a substantive test may be testing more than one hypothesis at a time because the test provides evidence that relates to more than one financial statement assertion. Hypothesis evaluation entails the examination of audit evidence to evaluate which of the potential hypotheses is (are) the most likely cause(s). This is a complex task for which there is limited guidance in professional standards and prior research.

Findings in psychology and auditing (Van Wallendael, 1989; Bonner and Pennington 1991) suggest that hypothesis evaluation is critical to successful performance of a diagnostic task such as business risk assessment and analytical procedures. Prior research suggests that multiple hypothesis evaluation is a task that individuals have difficulties performing.[1] Importantly, prior auditing studies examine only conditions where hypotheses are mutually exclusive and exhaustive. However, in the audit environment for complex enterprises, hypotheses may be interdependent and it is unlikely that all possible causes will be identified.

Prior auditing research (Asare and Wright 1997a, 1997b; Bedard and Biggs 1991a; Heiman, 1990) has not explicitly considered the impact of various relationships among hypotheses. Nonetheless, these studies explicitly recognise the potential for alternative conditions such as multiple causes (non-mutually exclusive hypotheses) and identify this as an important avenue for research. For instance, Anderson and Koonce (1998) examine auditor identification of multiple causes for a material unexpected fluctuation, a relatively common occurrence in practice. They find that auditors are more likely to detect such a problem if both the plausibility and sufficiency (i.e., whether a cause fully accounts for the fluctuations) of a cause are considered. Therefore, the prior literature suggests the following broad question:

How should alternative relationships among multiple hypotheses (causes) affect audit evidence evaluation?

The focus of this paper is on the crucial hypothesis evaluation stage of this process in assessing the most likely cause(s) for the fluctuations present. Specifically, we consider four scenarios: the hypotheses in the set are (1) mutually exclusive and collectively exhaustive; (2) mutually exclusive and non-exhaustive; (3) non-exclusive and non-exhaustive hypotheses with positive correlations; and (4) non-exhaustive with mixed correlations (positive, negative, and independent). If hypotheses are mutually exclusive, no two of them are viewed as likely to co-exist. The hypotheses set is considered exhaustive if all potential causes are

---

[1] See, e.g., in psychology – Van Wallendael and Hastie 1990; in auditing – Asare and Wright 1997a, 1997b; Jamal *et al.*, 1995, 1997; Hirst and Koonce 1996; Anderson and Koonce 1995; Bedard and Biggs 1991a. 1991b.

believed to have been identified. Hypotheses may also be independent and thereby not only are they considered to potentially co-exist, but also the existence of one does not preclude the existence of other hypotheses. Given the complex audit environment, all of these scenarios are relevant.

Our study provides two important contributions. First, using a probabilistic approach, we provide a comprehensive framework for auditors and audit researchers. Second, we analyse four decision scenarios and show how assumptions related to the hypotheses being considered affect evidence evaluation, hypothesis evaluation and hypothesis generation. The framework indicates that the pattern of probability revisions should vary substantially under alternative audit scenarios. A framework such as the one developed here has significant promise as a decision tool or to enhance auditor training.

The remainder of the paper is divided into three sections. The next section provides background information about hypothesis evaluation in auditing and contains a review of prior studies. The following section presents the general framework developed from probability theory. Section 4 discusses needed descriptive research and derives testable hypotheses based on our conceptual framework. The final section presents conclusions and implications and uses the modeling results to suggest other important avenues for future research.

## 2. Background and prior research

Auditing standards (e.g., AUS 210. AAuSB, 1999; SAS 53; AICPA 1988) and anecdotal evidence suggest that a broad array of scenarios are encountered in practice which are not limited to the mutually-exclusive and exhaustive conditions assumed in prior research. For example, in the Crazy Eddie, Inc. fraud case (Knapp, 1996), the gross margin increased despite slowing customer demand.[2] This increase was due to several simultaneous intentional and interrelated misstatements (an overstatement of the inventory count and also recording wholesale sales as retail and then reducing inventory by fewer units than actually sold). This case suggests a scenario in which multiple errors co-exist (a non-mutually exclusive hypothesis set) and where auditors may not have been considering an exhaustive set of feasible hypotheses, otherwise they would have investigated the fraud scenario.

As noted, hypotheses may be interdependent and thereby exist concurrently. For instance, when there are weak controls, several types of interrelated errors are possible. To illustrate, let us assume a client has weak segregation of duties where an accounting clerk makes daily deposits and also posts the accounting records. This weakness could lead to unintentional errors (e.g., mis-posting of accounts receivable) or fraud (e.g., lapping). Further, although prior research usually considers the hypotheses set to be exhaustive, Blocher (1991) and

---

[2] Crazy Eddie, Inc. was a widely publicised case, which led to significant litigation losses and US Securities and Exchange Commission sanctions for the auditing firms involved.

Blocher and Willingham (1993) have identified over 32 possible causes for a change in the gross margin of a manufacturing client. Thus, experienced auditors investigating a gross margin fluctuation may readily recognise that other potential, unidentified causes may have led to the fluctuation.

A widely accepted approach for modeling multiple hypotheses is to employ a probabilistic revision process (Winterfeldt and Edwards 1986; Winkler and Hays 1975; Luce and Raiffa, 1957). Probability theory prescribes that if there is an assumed mutually-exclusive and exhaustive scenario and an item of evidence indicates a particular hypothesis is more (less) likely, then the other hypotheses in the set are less (more) likely. This revision process is referred to as 'complementary' and leads to a fixed pool of probabilities that sum to one and thus are 'additive'.

Auditing research on diagnostic reasoning has focused on the generation of hypotheses, e.g., error frequency and recency (Libby 1985), experience and knowledge structure (Bedard and Biggs 1991a, 1991b; Libby and Frederick, 1990), and potential interference effects (Ng *et al.*, 1996; Church and Schneider 1993; Anderson *et al.*, 1992; Heiman, 1990). However, during an audit, the auditor also must obtain (search for) competent evidence to assess (evaluate) the likelihood or plausibility of the various possible explanations or hypotheses. These assessments can be thought of as leading to *priors* (given whatever knowledge the auditor currently possesses regarding the plausibility of a cause or of various causes) or as leading to *posteriors* (given incremental evidence, the plausibility of a cause or of various causes).

## 3. Hypothesis evaluation

As discussed earlier, prior research has found that decision-makers have difficulties performing hypothesis evaluation. Researchers conducting these studies hypothesise that these results are due to the cognitive difficulties required in updating probabilities. However, these studies were designed such that hypotheses were portrayed as being mutually exclusive and exhaustive, e.g., murder mysteries where there is a single culprit (Van Wallendael and Hastie, 1990) and, thus, did not consider the important issue of appropriate probability revisions under different conditions, which, as discussed earlier, may be present in auditing tasks.

Prior auditing studies have considered the impact of evidence on the probability assessments of an individual hypothesis or of a set of hypotheses. For example, similar to what has been observed in medical research, Biggs and Mock (1983) found evidence of 'early closure' on a particular hypothesis as auditors assessed internal control and detailed test results. Early closure refers to the tendency to discontinue hypothesis generation and evaluation prior to a thorough consideration of potential causes.

In hypothesis evaluation, an additional consideration is whether individuals appropriately encode evidence. One concern is the 'dilution effect' where non-

diagnostic evidence reduces posterior probabilities because it results in a lower proportion of direct evidence for a particular hypothesis (Nisbett *et al.*, 1981). The findings of auditing studies (e.g., Gaeth and Shanteau, 1984; Krishnamoorthy *et al.*, 1999) indicate that auditors correctly encode the direction of the evidence (positive, negative, and neutral), but there is conflicting evidence on the prevalence of a dilution effect (e.g., Hackenbrack, 1992, – confirming a dilution effect; Asare and Wright 1997b, – not finding an effect). Further, there have been numerous studies on potential order effects (e.g., recency) in evaluating evidence (e.g., Monroe and Ng, 2000; Tubbs *et al.*, 1990; Ashton and Ashton, 1988). However, recent studies suggest that task experience and response mode (Trotman and Wright, 1996), justification (Cushing and Ahlawat, 1997), and accountability (Kennedy, 1993) mitigate potential order effects.

Prior auditing research has examined the possibility that auditors may prematurely tend to accept a management explanation when encountering a problem or fluctuation (e.g., Heiman-Hoffman *et al.*, 1995; Bedard and Biggs, 1991a). Anderson and Koonce (1995) studied the effect of 'explanation' and 'quantification' (1998) on auditor's assessments. Their results indicate the possibility of two types of explanation failures or reasoning errors: plausibility and quantification. Plausibility refers to whether the proposed cause properly explains the fluctuations present. Quantification relates to whether the cause is sufficient to have accounted for the size of the fluctuation. However, it is somewhat problematic to assess the 'quality' of auditor reasoning without a precise analysis of the underlying hypothesis set and whether the hypotheses are independent, mutually exclusive or exhaustive.

In an experimental study, Heiman (1990) found that auditors decrease their probability assessment of a target hypothesis when directed to consider other competing hypotheses. These results suggest a 'discounting' strategy that is consistent with a complementary evaluation strategy whereby a fixed pool of probabilities is spread over the hypothesis set. However, since data were not gathered on the probability assessments of competing hypotheses, direct evidence was not provided to determine the precise nature of the probability revision process employed.

Bhattacharjee *et al.* (1999) examine the impact of hypothesis set size on hypothesis evaluation. They posit that there is an optimal number of hypotheses for the auditor to consider in balancing effectiveness and efficiency. A *large* set (e.g., five or more hypotheses) is likely to make hypothesis generation and evaluation cognitively complex and unwieldy. On the other hand, a very small set (e.g., one hypothesis) may be inefficient, having to go through multiple rounds in considering additional hypotheses, and ineffective in not examining a broad set of hypotheses. They find that a hypothesis set size of three is optimal, resulting in greater efficiency and equal decision accuracy as the five hypotheses set condition.

Three auditing studies directly test hypothesis evaluation approaches. In two of these, Asare and Wright (1995 and 1997a) asked auditors to assess the

probability of a set of inherited hypotheses after obtaining audit evidence. Their results suggest auditors update the probability of only the hypothesis implicated by the evidence received. From a probabilistic perspective and given a mutually exclusive and exhaustive scenario, audit evidence that suggests a particular hypothesis is more (less) likely also implies that the other hypotheses in the set are less (more) likely (i.e., a 'complementary' revision process).

Recently, Mock *et al.* (2001) conducted a verbal protocol study using a version of the Asare and Wright task. They found that most auditors framed the hypotheses as an exhaustive set of causes, however, their likelihood assessments were super-additive and reflected a multiple set of causes. When assessing audit evidence, these auditors considered a number of interdependencies among potential causes, but these interdependencies often were not evident in their likelihood revisions.

To summarise, prior auditing research has not explicitly considered the impact of various relationships among hypotheses. Nonetheless, these studies explicitly recognise the potential for alternative conditions such as multiple causes (non-mutually exclusive hypotheses) and identify this as an important avenue for research.

## 4. A framework for multiple hypothesis evaluation

In this section, we discuss a general framework based on probability theory for evaluating multiple hypotheses that are thought to be the primary causes of an effect, E. (See Srivastava *et al.*, 2001 for a comprehensive discussion of the development of this framework.) As mentioned earlier, four scenarios are analysed: (1) mutually exclusive and exhaustive hypotheses; (2) mutually exclusive and non-exhaustive hypotheses; (3) non-exclusive and non-exhaustive hypotheses with positive correlations; and (4) non-exclusive and non-exhaustive hypotheses with mixed correlations. To illustrate a mixed correlation situation, assume a hypothetical scenario where an auditor encounters an unexpected material increase in the gross margin of a manufacturing client while conducting analytical procedures. After some consideration the auditor identifies or inherits the following four hypotheses that may explain this fluctuation: $H_1$ – increase in sales volume; $H_2$ – increase in purchase discount; $H_3$ – increase in sales price; $H_4$ – an overhead allocation error between period and product costs. The first two hypotheses are positively correlated; an increase in sales volume would increase purchases and hence increase purchase discounts. But $H_1$ and $H_3$ are negatively correlated; an increase in sales price usually leads to a decrease in sales volume. Similarly, $H_2$ and $H_3$ are negatively correlated. However, one can argue that $H_4$ is independent of $H_1$, $H_2$, and $H_3$, i.e., whether there is an overhead allocation error or not does not tell us whether sales volume, purchase discounts, or sales prices have increased. To make the above hypotheses set non-exhaustive, we can further assume that the auditor believes there are likely other causes for the fluctuation that could be possible and have not yet been

identified. All such reasons are assumed to be subsumed in hypothesis $H_5$ – others, e.g., fictitious sales, cut-off errors in sales and/or inventory, errors in the inventory count, shift in product mix, or a decrease in labour or materials costs. $H_5$ may be independent or correlated with the other hypotheses.

The auditor is expected to obtain audit evidence to test the veracity of some or all of the potential causes (SAS 56, AUS 502). Based on the evidence gathered, initial (prior) probabilities would be revised in some fashion for the hypotheses and the most likely cause(s) would be identified. If an error is suspected, the amount of the error would be determined and the financial statements would be adjusted accordingly. Given the auditor's asymmetric loss function, the greatest concern is improperly concluding that the fluctuation was not due to an error or fraud when, in fact, there is a material misstatement present in the financial statements.

In Appendix A, a general model is derived for posterior probabilities of a set of five hypotheses that are assumed to have caused an effect E in terms of the strength of evidence (represented by the likelihood ratios, $\lambda$'s) and the interrelationships among the hypotheses. A positive piece of evidence ($\lambda > 1$) pertaining to a hypothesis increases its probability of being the potential cause of the fluctuation. A negative piece of evidence ($0 \geq \lambda < 1$) decreases the probability of the hypothesis being the potential cause. For $\lambda = 1$, the evidence has no information with respect to a particular hypothesis, *i.e.*, it is neutral and has no effect on the probability that a hypothesis is the underlying cause. The interrelationships among the hypotheses are represented in terms of the joint probability distribution over the space of the five hypotheses as given in (A-4) – (A-7) and also depicted in terms of the Venn diagrams in Figure 1, which illustrate alternative relationships between a set of hypotheses.

Panel A in Figure 1 depicts a situation where the hypotheses are mutually exclusive and exhaustive. The outer area in Panel A represents 'other potential causes' and is denoted by $H_0$. When the hypotheses are viewed as exhaustive, this area would be an empty set. However, in audit practice it is also possible that multiple causes may have led to the fluctuation (non-exclusive) (Panels C and D) and/or the auditor may recognise that there are other causes that have not been identified which jointly may have caused the fluctuation (non-exhaustive) (Panels B and D). As discussed earlier, prior auditing research has either stipulated or assumed a mutually exclusive, exhaustive decision setting. The analyses to follow consider various scenarios for hypothesis evaluation.

Case A: Mutually exclusive and exhaustive hypotheses set

A mutually exclusive and exhaustive set of hypotheses is plausible in practice. For instance, this scenario may occur for a client with low inherent risk and good controls where few, if any, errors are expected. Thus, error causes are expected to be isolated. Further, the fluctuation in question may be one in which the auditor feels confident that there are a limited number of potential
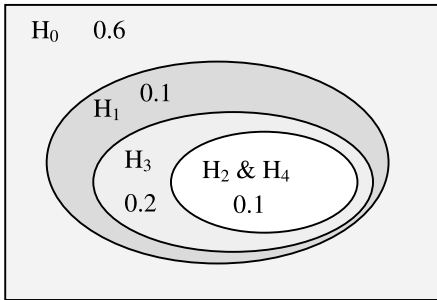
**Panel A:** Mutually exclusive and collectively exhaustive (Case A)
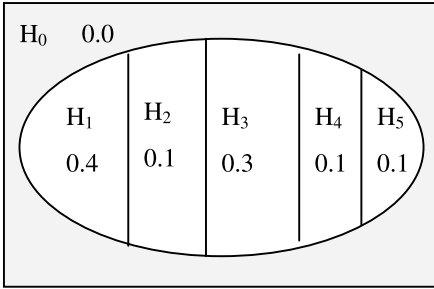
**Panel B:** Mutually exclusive and non-exhaustive (Case B)
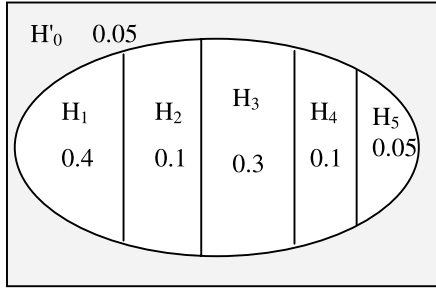
**Panel C:** Non-exclusive and non-exhaustive: positive correlation (Case C)

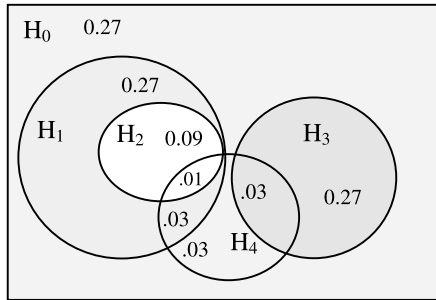**Panel D:** Non-exclusive and non-exhaustive: mixed correlation (Case D)



Figure 1

causes (exhaustive). Figure 1 Panel A depicts this case with the outer area representing a null set. Assume that there are five items of evidence available, $E_1$, $E_2$, $E_3$, $E_4$, and $E_5$, respectively, pertaining to $H_1$, $H_2$, $H_3$, $H_4$, and $H_5$, e.g., five substantive tests – one for each hypothesis. The nature and strength of evidence are defined by the likelihood ratios and depend on the auditor's judgment about the findings of the evidence.

Equation (A-7) in Appendix A gives a general expression for the posterior probability of hypothesis $H_1$ given that we have observed the effect E and the five items of evidence, $E_1 - E_5$ pertaining, respectively, to $H_1 - H_5$. The inter-relationships among the five hypotheses are determined by the joint probability distribution as given in (A-4) and (A-5). The case of a mutually exclusive and exhaustive set of five hypotheses is represented by the following non-zero probabilities in (A-4) and (A-5): $P(H_1{\sim}H_2{\sim}H_3{\sim}H_4{\sim}H_5|E)$, $P({\sim}H_1H_2{\sim}H_3{\sim}H_4{\sim}H_5|E)$, $P({\sim}H_1{\sim}H_2H_3{\sim}H_4{\sim}H_5|E)$, $P({\sim}H_1{\sim}H_2{\sim}H_3H_4{\sim}H_5|E)$, and $P({\sim}H_1{\sim}H_2{\sim}H_3{\sim}H_4H_5|E)$. Using these non-zero probabilities in (A-7), we obtain the following expression for the posterior probability that hypothesis $H_1$ is a potential cause given all the evidence:

$$P(H_1|EE_1E_2E_3E_4E_5) = \lambda_1 P(H_1 \sim H_2 \sim H_3 \sim H_4 \sim H_5|E)/[\lambda_1 P(H_1 \sim H_2 \sim H_3 \sim H_4 \sim H_5|E)$$

$$+ \lambda_2 P(\sim H_1 H_2 \sim H_3 \sim H_4 \sim H_5|E) + \lambda_3 P(\sim H_1 \sim H_2 H_3 \sim H_4 \sim H_5|E)$$

$$+ \lambda_4 P(\sim H_1 \sim H_2 \sim H_3 H_4 \sim H_5|E) + \lambda_5 P(\sim H_1 \sim H_2 \sim H_3 \sim H_4 H_5|E)]$$

$$(1)$$

where $\lambda$'s represent the strength of various items of evidence. Similar expressions can be written for the other hypotheses. The above expression is a general result and special cases can be obtained by setting different values for the likelihood ratios. For example, if we had no evidence at all then all the likelihood ratios would be one (i.e., $\lambda_1 = \lambda_2 = \lambda_3 = \lambda_4 = \lambda_5 = 1$) and the right hand side of (1) will reduce to $P(H_1 \sim H_2 \sim H_3 \sim H_4 \sim H_5|E)$ which is the initial estimate of probability that $H_1$ is present given that effect E is observed.

Let us consider another special case. Suppose we have infinitely strong evidence in support of hypothesis $H_1$ that it is present, *i.e.*, $\lambda_1 \rightarrow \infty$, and no other evidence (i.e., $\lambda_2 = \lambda_3 = \lambda_4 = \lambda_5 = 1$). For this case, from equation (1) we obtain $P(H_1|EE_1) = 1$, and similarly one can show that $P(H_2|EE_1) = P(H_3|EE_1) = P(H_4|EE_1) = P(H_5|EE_1) = 0$. This result is as expected in a mutually exclusive and exhaustive case. If the evidence suggests that only one given hypothesis is present, then all other hypotheses are excluded.

Consider another case for illustration. Assume the following initial assessment of probabilities (priors) for five mutually exclusive and exhaustive hypotheses after having observed effect E: $P(H_1|E) = 0.4$, $P(H_2|E) = 0.1$, $P(H_3|E) = 0.3$, $P(H_4|E) = 0.1$, $P(H_5|E) = 0.1$. Since these hypotheses are mutually exclusive and exhaustive, the above probabilities can be written as:

$P(H_1|E) = P(H_1 \sim H_2 \sim H_3 \sim H_4 \sim H_5|E) = 0.4,$

$P(H_2|E) = P(\sim H_1 H_2 \sim H_3 \sim H_4 \sim H_5|E) = 0.1,$

$P(H_3|E) = P(\sim H_1 \sim H_2 H_3 \sim H_4 \sim H_5|E) = 0.3,$

$P(H_4|E) = P(\sim H_1 \sim H_2 \sim H_3 H_4 \sim H_5|E) = 0.1,$ and

$P(H_5|E) = P(\sim H_1 \sim H_2 \sim H_3 \sim H_4 H_5|E) = 0.1.$

These values are substituted in equation (1) to obtain a general expression for the posterior probability of hypothesis $H_1$ for this case as a function of various likelihood ratios. Similar expressions of posterior probabilities for the other hypotheses can be derived.

Next, we analyse the impact of successive pieces of audit evidence on various hypotheses. For illustration, we assume a mixed set of evidence[3]: evidence

---

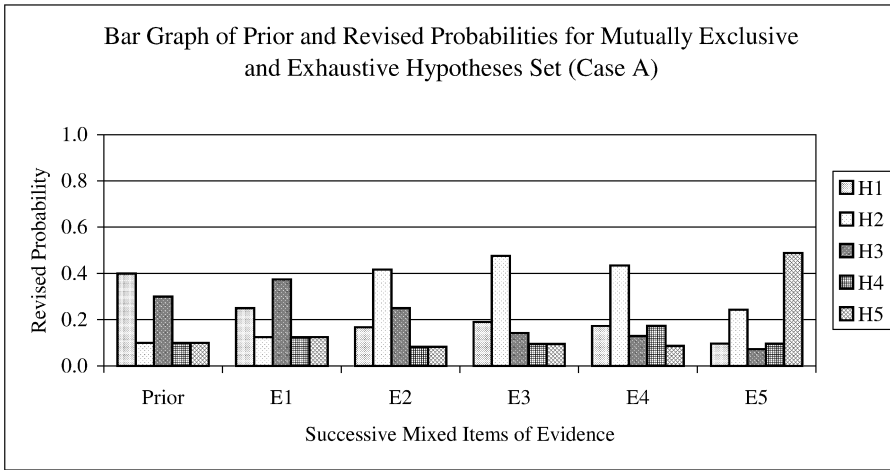[3] The likelihood ratios are chosen to illustrate the impact of strength of evidence on various hypotheses.

Figure 2

$E_1$ and $E_3$ are moderately negative ($\lambda_1 = \lambda_3 = 0.5$), $E_2$ and $E_4$ are moderately positive ($\lambda_2 = 5$, $\lambda_4 = 2$), and $E_5$ is strongly positive ($\lambda_5 = 10$). Figure 2 shows that as $E_1$ is observed, the posterior probability of $H_1$ decreases from 0.4 to 0.25 whereas the posterior probabilities of the other hypotheses increase in such a way that the sum of all the posteriors is one. When $E_2$ is observed (moderately positive evidence) the posterior probability of $H_2$ increases to 0.4167 and other posterior probabilities are adjusted downward to make the sum one. When $E_3$ (moderately negative evidence for $H_3$) is observed, the posterior probability of $H_3$ decreases and the posterior probabilities of the remaining hypotheses increase in such a way that the sum remains one. These results are what one expects under an exhaustive and mutually exclusive set of hypotheses.

Case B: Mutually exclusive and non-exhaustive

A mutually exclusive and non-exhaustive hypothesis set occurs when the auditor believes that there is one or more cause(s) of an unusual material fluctuation, but cannot retrieve from memory or identify all possible causes. Assume that the auditor generates four hypotheses that are mutually exclusive: $H_1$, $H_2$, $H_3$, and $H_4$. However, there could be other potential causes contained in the other hypothesis set $H_0$.

Under a probability treatment, one can put all non-identified causes or hypotheses in one group $H_0$ ('Other hypotheses'). This process makes the current case similar to Case A involving a mutually exclusive and exhaustive set discussed earlier with $H_1$, $H_2$, $H_3$, $H_4$, and $H_0$ as the set of hypotheses. Suppose during the audit process, the auditor has collected four independent items of evidence, $E_1$, $E_2$, $E_3$, and $E_4$, pertaining to the four hypotheses: $H_1$, $H_2$, $H_3$, and

$H_4$, respectively. In the auditor's judgment these items of evidence negate the presence of the corresponding hypotheses, some negate strongly and some negate moderately. In terms of likelihood ratios, these strengths are assumed to take the following values as judged by the auditor: $\lambda_1 = 0.1$, $\lambda_2 = 0.5$, $\lambda_3 = 0.2$, $\lambda_4 = 0.5$. Since these items of evidence provide no support for any of the initially considered hypotheses, the auditor generates a fifth potential hypothesis, $H_5$, an additional mutually exclusive hypothesis subsumed in $H_0$, as the cause of the unusual fluctuation. Now, suppose the auditor obtains a significantly strong and positive piece of evidence ($E_5$ with say, $\lambda_5 = 15$) that implicates $H_5$. The last piece of evidence suggests that $H_5$ is the most likely cause of the fluctuation as discussed below.

The above situation can be treated in our framework developed in Appendix A by considering $H_0$ to consist of two mutually exclusive hypotheses, $H_5$ and $H_0'$ (other potential hypotheses). We use equations (A-5) – (A-7), as used in Case A, to determine the posterior probability of hypothesis $H_1$ by establishing that the only non-zero probabilities are the following: $P(H_1{\sim}H_2{\sim}H_3{\sim}H_4{\sim}H_0|E)$, $P({\sim}H_1H_2{\sim}H_3{\sim}H_4{\sim}H_0|E)$, $P({\sim}H_1{\sim}H_2H_3{\sim}H_4{\sim}H_0|E)$, $P({\sim}H_1{\sim}H_2{\sim}H_3H_4{\sim}H_0|E)$, and $P({\sim}H_1{\sim}H_2{\sim}H_3{\sim}H_4H_0|E)$. Similar expressions can be obtained for hypotheses $H_2$ – $H_4$ by extending (A-4) – (A-7). Finally, when the auditor generates a fifth hypothesis, $H_5$, subsumed in $H_0$, by splitting $H_0$ into two mutually exclusive hypotheses, the posterior probability of $H_5$ after all the items of evidence, $E_1$ – $E_5$, have been collected would be similarly obtained from (A-5) – (A-7) by using the property that hypotheses in the set are mutually exclusive.

Figure 3 presents a bar graph of posterior probabilities of the hypotheses as each item of evidence is obtained. We have assumed the following initial
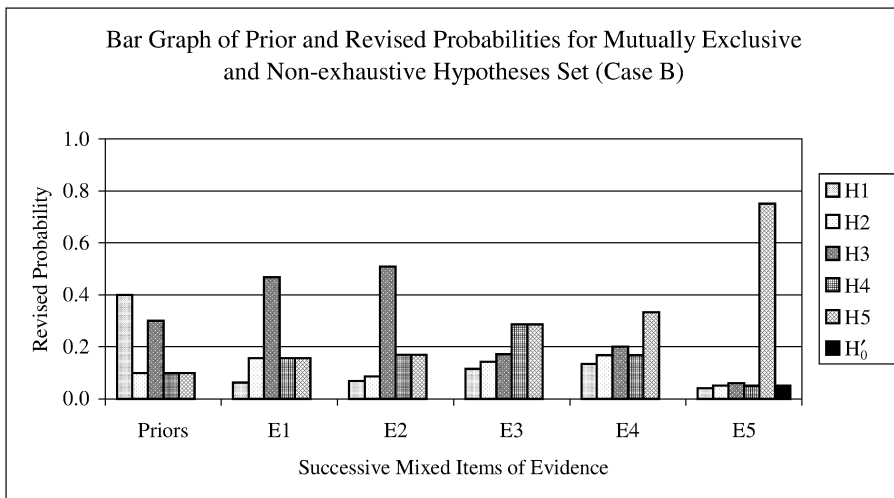


Figure 3

probabilities: $P(H_1|E) = 0.4$, $P(H_2|E) = 0.1$, $P(H_3|E) = 0.3$, and $P(H_4|E) = 0.1$, for the first four hypotheses based on observing an unusual fluctuation in the gross margin of a client as mentioned earlier. The remaining prior is assigned to the other hypotheses contained in $H_0$, i.e., $P(H_0) = 0.1$. As mentioned earlier, we assume that the auditor obtains only negative items of evidence, some strongly negative and some moderately negative ($\lambda_1 = 0.1$, $\lambda_2 = 0.5$, $\lambda_3 = 0.2$, $\lambda_4 = 0.5$) pertinent, respectively, to hypotheses $H_1 - H_4$. The posterior probabilities of these hypotheses and $H_0$, after considering all the above items of evidence are calculated to be: $P(H_1|EE_1E_2E_3E_4) = 0.133$, $P(H_2|EE_1E_2E_3E_4) = 0.167$, $P(H_3|EE_1E_2E_3E_4) = 0.200$, $P(H_4|EE_1E_2E_3E_4) = 0.167$, and $P(H_0|EE_1E_2E_3E_4) = 0.333$.

Realising the fact that there is no overwhelming evidence about any of the initially considered hypotheses, the auditor generates a fifth hypothesis, $H_5$, by splitting $H_0$, as mentioned earlier. In the present discussion we assumed $H_5$ and $H_0'$ to be equally likely in the absence of other information; and therefore $P(H_0|EE_1E_2E_3E_4) = 0.333$ is distributed between $H_5$ and $H_0'$ equally, i.e., $P(H_5|EE_1E_2E_3E_4) = 0.1665$ and $P(H_0'|EE_1E_2E_3E_4) = 0.1665$, which is the same as assuming from the beginning that $P(H_5|E) = 0.05$ and $P(H_0'|E) = 0.05$, as shown in Panel B of Figure 1. Since evidence $E_5$ pertinent to $H_5$ is positive and strong ($\lambda_5 = 15$), hypothesis $H_5$ becomes the dominant hypothesis as the cause of the unusual fluctuation as evidenced from the posterior probabilities (see Figure 3): $P(H_1|EE_1E_2E_3E_4E_5) = 0.04$, $P(H_2|EE_1E_2E_3E_4E_5) = 0.05$, $P(H_3|EE_1E_2E_3E_4E_5) = 0.06$, $P(H_4|EE_1E_2E_3E_4E_5) = 0.05$, $P(H_5|EE_1E_2E_3E_4E_5) = 0.75$, and $P(H'_0|EE_1E_2E_3E_4E_5) = 0.05$.

Case C: Non exclusive and non-exhaustive hypotheses set with positive correlation

Panel C in Figure 1 provides a Venn diagram for the case where all the non-exclusive and non-exhaustive hypotheses are positively correlated. This scenario appears likely in certain client situations. For example, assume a scenario where managers receive a bonus on reported earnings and controls over sales and inventory are weak. These risks increase the chances that bonuses are overstated because profits are inflated as a result of several potential causes such as fictitious sales, improper cut-off of sales and/or purchases, and improper inventory count or valuation. Additionally, the nature of a particular fluctuation may result in a low level of confidence by the auditor that all likely causes are considered (non-exhaustive). For example, the following hypotheses may all be potential causes of a material increase in the gross margin: $H_1$ – understatement of overhead costs due to an accrual error; $H_2$ – increase in sales volume; $H_3$ – decrease in unit overhead costs; and $H_4$ – increase in purchase discounts. However, there are other hypotheses that are not known to the auditor at the moment, which can be assumed to be contained in $H_0$. As mentioned earlier, Panel C in Figure 1 depicts a Venn diagram for this case. The inner ellipse represents the conjunction of all the four hypotheses ($H_1$, $H_2$, $H_3$, and $H_4$). The next
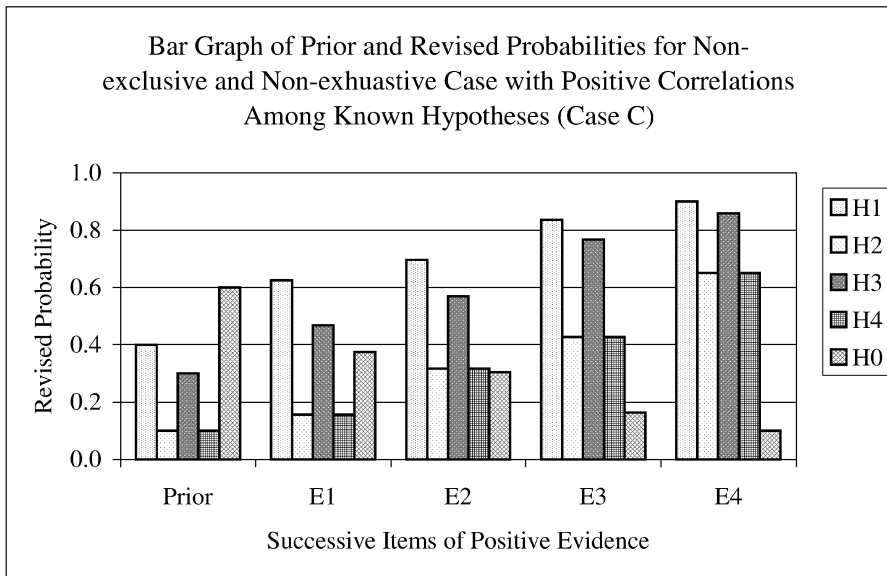
Figure 4

ellipse represents the overlapping of $H_1$ and $H_3$. This overlapping of Venn diagrams represents positive correlations[4] among the hypotheses. The area outside of the outer ellipse represents the unknown hypotheses.

The posterior probability expressions in terms of the strength of evidence for various hypotheses can again be obtained from the general results derived in Appendix A by substituting the corresponding joint probabilities from Panel C in Figure 1 and assuming that $H_5 = H_0$. Figure 4 represents the impact of successive positive ($\lambda_i = 2.5$, $i = 1, 2, 3$, and 4) items of evidence on posterior probabilities of various hypotheses for the present case. We find that evidence $E_1$ has not only increased the posterior probability of $H_1$ from 0.4 to 0.625 but also has increased the posterior probabilities of the rest of the hypotheses. This is a logical result. From the assumed strong correlation between the hypotheses, one would expect that as the posterior probability of one hypothesis increases, the posterior probabilities of the remaining hypotheses would also increase. Also, we observed that if the evidence suggests that a given hypothesis is less likely then the same evidence implies that the remaining hypotheses are also less likely. This is again a logical result because of the assumed strong positive correlation between the hypotheses.

---

[4] The correlation coefficient between hypotheses $H_i$ and $H_j$ is defined (Morris and Larick 1995) as: $\phi_{ij} = [P(H_iH_j) - P(H_i)P(H_j)]/[P(H_i)(1-P(H_i))P(H_j)(1-P(H_j))]^{1/2}$. Based on the Venn diagram in Figure 1, Panel C, we obtain the following correlation coefficients among the four hypotheses: $\phi_{12} = 0.41$, $\phi_{13} = 0.80$, $\phi_{14} = 0.41$, $\phi_{23} = 0.51$, and $\phi_{24} = 1.0$.

Unlike Cases A and B, the revised probabilities do not necessarily add up to one (See Figure 4). In fact, the sum could be less than one or greater than one depending upon the prior marginal probabilities and the nature of evidence. For example, in the above illustration, the sum of all the posteriors of the four hypotheses after the first piece of evidence is 1.41 and after all the four positive items of evidence it is 3.05. Once again, these results make logical sense. Since the four hypotheses are not mutually exclusive, they can occur at the same time. However, if some items of evidence were negative then the sum would become less than one. Also, when all the evidence pertaining to $H_1 - H_4$ is strongly negative, the 'other hypotheses' ($H_0$) becomes more likely although there is no direct evidence pertaining to $H_0$. In an audit context, this suggests that all of the hypotheses the auditor initially thought might explain the fluctuation have been eliminated[5]. This situation indicates the need for the auditor to generate and test additional causes.

Case D: Non-exclusive and non-exhaustive hypotheses set with mixed correlation

In this section we consider a more complex situation. Consider the following hypotheses to be the potential causes of a material increase in the gross margin: $H_1$ – increase in sales volume; $H_2$ – increase in purchase discount; $H_3$ – increase in sales price; and $H_4$ – overhead allocation error, and $H_0$ – all other unknown causes. As mentioned earlier, the first two hypotheses are positively correlated; an increase in sales volume would increase purchases and hence increase purchase discounts. But $H_1$ and $H_3$ are negatively correlated; an increase in sales price usually leads to a decrease in sales volume. Similarly, $H_2$ and $H_3$ are negatively correlated. However, one can argue that $H_4$ is independent of $H_1$, $H_2$, and $H_3$, i.e., whether there is an overhead allocation error or not does not tell us whether the other hypotheses are present.

Panel D in Figure 1 depicts a Venn diagram for the present case with an assumed relationship as represented by the joint probabilities given in the Venn diagram in Panel D of Figure 1. The correlation coefficients (see footnote 4 for definition) for various pairs of hypotheses based on the joint probabilities given in Figure 1, Panel D, are: $\phi_{12} = 0.41$, $\phi_{13} = -0.53$, $\phi_{14} = 0$, $\phi_{23} = -0.22$, $\phi_{24} = 0$, $\phi_{34} = 0$. These correlation coefficients represent the following relationships: (1) hypotheses $H_1$ and $H_2$ are positively correlated, (2) $H_1$ and $H_3$ are negatively correlated, (3) $H_2$, and $H_3$ are negatively correlated, and (4) hypotheses $H_1$, $H_2$ and $H_3$ are independent of hypothesis $H_4$.

Again the posterior probability expressions in terms of the strength of evidence for various hypotheses can be obtained from the general results derived in Appendix A by substituting the corresponding joint probabilities from

---

[5] Actually, the probability of these hypotheses would only approach, but not equal, zero. But they can be assumed close enough to zero to be immaterial.
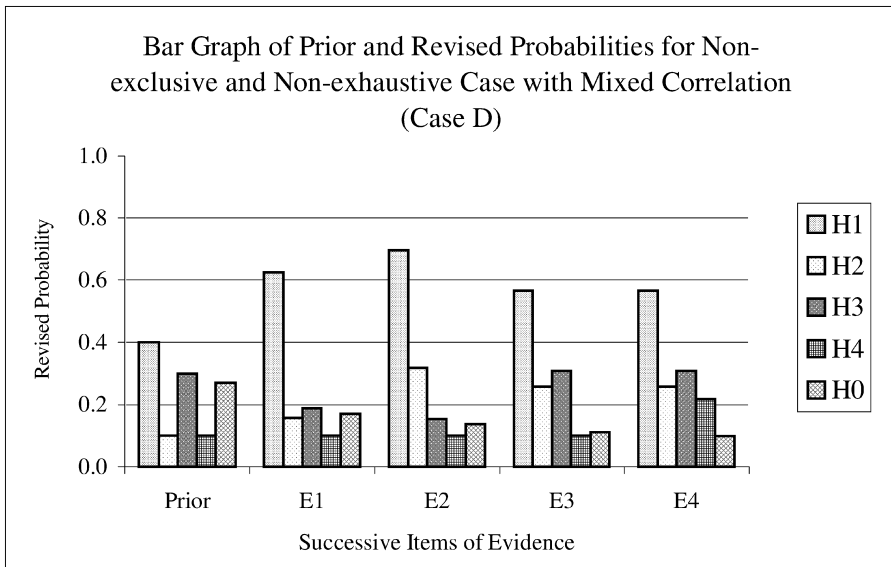
Figure 5

Panel D in Figure 1 and assuming that $H_5 = H_0$. Figure 5 represents the impact of successive positive ($\lambda_i = 2.5$, i = 1, 2, 3, and 4) items of evidence on posterior probabilities of various hypotheses for the present case. We find that evidence $E_1$ has not only increased the posterior probability of $H_1$ from 0.4 to 0.625 but also has increased the posterior probability of $H_2$ from 0.1 to 0.156, decreased the posterior probability of $H_3$ from 0.3 to 0.188, and has no effect on the probability of $H_4$. This result is in accordance with the assumed relationships; $H_1$ and $H_2$ are positively correlated, $H_1$ and $H_3$ are negatively correlated, and $H_1$ and $H_4$ are independent. Similarly when $E_2$ is observed, the probability of $H_2$ goes up while the probabilities of $H_1$ and $H_3$ go down and the probability of $H_4$ remains unchanged. Again this is a logical result given the assumed relationships. Evidence $E_3$ has a similar effect as that of $E_1$. When Evidence $E_4$ is observed, the probability of $H_4$ goes up while the probabilities of $H_1$, $H_2$, and $H_3$ remain unchanged as expected because of the independent relationship between $H_4$ and the other hypotheses, $H_1$, $H_2$, and $H_3$. The probability that other unknown hypotheses subsumed in $H_0$ are the potential causes decreases as additional positive items of evidence are obtained for the known hypotheses. However, as one can easily show using the general expression, when we obtain evidence negating the presence of the known hypotheses, the probability that hypotheses subsumed in $H_0$ are the potential cause increases. If there is over-whelming evidence that none of the initially considered hypotheses are present then the probability that a hypothesis in $H_0$ is the cause becomes almost one.

## 5. Future research and testable hypotheses

To be of greatest value, the analytical model discussed above should lead to empirical research and have implications for practice. In this section we first discuss descriptive research that is needed to help us better understand the nature of the multiple hypothesis tasks that auditors face. Then testable hypotheses are discussed that illustrate some important empirical issues in multiple hypothesis settings that need to be resolved.

### 5.1. Descriptive research

It is first necessary to conduct exploratory descriptive research in important audit settings to understand how auditors actually frame the hypotheses they consider for a particular task. For example in some settings they may assume a small set of exhaustive causes that are mutually exclusive whereas in others they may consider a large set of interrelated causes. Prior psychological research (Pennington, 1987; Bauer and Eddy, 1986; Newell and Simon, 1972) has found that problem framing (representation) affects hypothesis generation, information search, and decision performance. As revealed by our modelling results, probability theory calls for a different pattern of belief revisions, given the same piece of evidence, under different multiple hypotheses schemas. For example, in a non-exhaustive and non-exclusive hypotheses set situation where two hypotheses are independent, the posterior probability of a hypothesis will not change and the summed probabilities may exceed one (super-additive) when a piece of evidence pertaining to another hypothesis is obtained.

However, probability assessments may also be affected by a number of behavioural factors including cognitive limitations or information processing effects (e.g., order effects). For instance, Asare and Wright (1997a, 1995) and Van Wallendael (1989) have observed independent probability revisions that they attribute to cognitive strain in revising probabilities for multiple hypotheses. Whether these revisions are coherent or not depends on how the decision makers are framing the task and the types of considerations discussed in our models.

Descriptive evidence about the decision process (e.g., Mock *et al.*, 2001) also would be valuable in assessing the manner in which auditors adopt and evaluate hypotheses under different situations. Subsequent experimental studies may then be guided by this evidence in evaluating whether decision-makers revise probabilities consistently with the way they frame the task.

### 5.2. Testable research hypotheses derived from the analytical models

In considering the kinds of hypotheses that may be derived from the models that have been discussed, one might initially wish to consider some of the direct implications of the analyses. For example, researchers may wish to consider

the explicit issues the models address: evidence strength/direction; evidence implicating multiple causes; or the implied need to generate new hypotheses. Such hypotheses could be framed directly from one or more of the main analytical derivations, for instance, 'evidence strength will significantly impact the magnitude of probability revisions'. This would test whether auditors are sufficiently sensitive to source diagnosticity and reliability. Although this general type of test has been conducted (Hirst, 1994; Bamber, 1983; Krishnamoorthy *et al.*, 1999), our model provides rigorous predictions that apply to more realistic cases where multiple hypotheses are important.

Similarly, one could test the hypothesis that 'probability revisions will be made for multiple causes when evidence that jointly implicates them is present' or that 'new hypotheses will be generated when evidence reduces the probability of hypotheses in the current set to a low level.' This would test whether auditors are good at identifying decision settings where generating new hypotheses is implicated. If they are not, decision aids may be introduced to improve performance. Asare and Wright (2000) provide experimental evidence indicating auditors infrequently generate additional hypotheses once audit testing commences, but researchers have not studied contexts where the probabilities of the set of considered hypotheses is low.

However, *ideally* in the kinds of settings discussed in this paper, deriving research hypotheses should be jointly dependent on three factors: a theory of behaviour; the task; and the analytical implications of our framework. In the following paragraphs, three examples of such testable hypotheses are discussed. Obviously there are many other hypotheses that may be specified if one considers other important behavioural theories or the possible effects and interactive effects of 'mediating' variables such as the use of decision-aids or training.

For our first example we consider bounded rationality, which suggests that departures from optimal behaviour are dependent upon the cognitive complexities at hand. For instance, it may be hypothesised that in a simple task setting with only two mutually exclusive and exhaustive causes, individuals will revise probabilities in a manner that is consistent with probability theory. However, as one encounters increasingly complex causal schemas such as multiple related, non-exhaustive hypotheses, greater departures from the predictions of our framework would be expected. Thus, the above discussion suggests the following testable hypothesis:

*H1: As a result of cognitive complexities and strain, departures of probability revisions from the normative framework advanced will increase with greater complexities in the causal schema (see Cases C and D in Section 3).*

Second, consider an audit context where a significant change in inventory turnover is observed. Further assume we are interested in the probability revision behaviour of novice versus expert auditors and the auditors are asked to provide an assessment of various possible explanations of the change in the ratio. One

theory that might be considered would be schema theory, which postulates that experts have developed more sophisticated schemas and thus are able to retrieve from memory more complex models of possible relationships between causes, evidence and effects (a significant change in the inventory turnover ratio). The hypotheses derived would then depend on the specifics of the task setting. Assuming a simple task such as the Morris and Larick case of only two possible causes and based on the analytical models presented earlier, one would hypothesise no difference in probability revision behaviour between the expert and novice auditors. However, if the task setting is made more complex, it is reasonable to expect that the experts' assessments of the likelihood of the possible causes will be more coherent (closer to those implied by our analytical models) than the assessments of the novices.[6] However, at some point the complexities of the situation may make hypothesis evaluation overly demanding even for experienced individuals (e.g., the presence of inherent risks such as questionable management integrity which can lead to pervasive effects or the receipt of evidence implicating more than one hypothesis). The analytical modelling presented in this paper provides both rigorous benchmarks as to the judgments that should be made and a gauge of the complexity inherent in the task. A hypothesis that captures this setting can be expressed as:

*H2: Experts' probability revisions will exhibit a greater correspondence with the assessments predicted by our models in more complex causal settings than novices. However, due to cognitive difficulties, in the most complex settings both experts and novices will exhibit similarly wide departures from the framework (See Cases B and D in Section 3).*

Third, consider a task where auditors are assessing unexpected earnings and are given traditional accounting information, industry data and information obtained from discussions with management. Such information is likely to imply many possible causes for unexpected earnings. Such explanations (causes) many be interrelated in complex ways and probably would not be mutually exclusive. In such a setting, one may wish to test the implications of risk aversion and prospect theory on the auditors' assessment of the likelihoods of the proposed causes. Although a more detailed task specification and analysis would be needed to derive explicit testable hypotheses, in general we would hypothesise that the auditors would overweight unfavourable evidence and thus generate inappropriately conservative probability assessments. The opposite behaviour would be expected for favourable evidence. Again the modelling developed earlier provides rigorous benchmarks as to what appropriate probability assessments should be. The third testable hypothesis can be expressed as:

———————

[6] Biggs *et al.*, 1988 obtained results consistent with this hypothesis in a study of auditor analytical review.

*H3: Revisions in probabilities for unfavourable evidence will be excessively conservative when compared to the predictions of our analytical models, i.e., overweight negative evidence and underweight positive evidence.*

## 6. Conclusions, implications and future research

Using a probabilistic approach, we provide a comprehensive analytical framework for auditors to aggregate evidence for hypothesis evaluation under four realistic scenarios. As discussed in Section 3, the nature of the hypothesis set dramatically impacts the pattern of belief revisions prescribed under this framework. Given the multiple possibilities described and the inherent complexities of hypothesis evaluation, an analytical framework such as the one developed here has promise both as a decision tool and as a means to enhance auditor training.

The framework also identifies conditions (Cases B, C and D) where there is a need for hypothesis generation. This helps formally link the hypothesis evaluation and generation activities and potentially provides a more rigorous understanding of the role of hypothesis generation in auditing. Lastly, the framework provides a rigorous foundation from which to guide both future analytical and empirical research.

For example, in a professional environment such as auditing, little is known regarding the hypothesis conditions normally encountered in practice and, thus, the problem representation that may be or should be utilised. Prior psychological research (Pennington, 1987; Bauer and Eddy 1986; Newell and Simon, 1972) has found that problem representation affects hypothesis generation, information search, and decision performance. As revealed by our modelling results, probability theory calls for a different pattern of belief revisions, given the same piece of evidence, under different multiple hypothesis contexts. Thus, there exist several reasons why behaviour might differ from theory including cognitive limitations, information processing effects (e.g., order effects), and alternative hypotheses set representations. Each of these possibilities needs to be investigated in future empirical research.

The framework suggests a number of other avenues for future research. As noted, the pattern of expected probability revisions, as suggested by our analyses, would be dependent upon the perceived dependencies between the hypotheses. Such dependencies would be learned through experience from the normal audit environment for the context examined. Future research should explicitly consider the actual and perceived dependencies between the hypotheses.

Decision process evidence (e.g., verbal protocols) would be valuable in assessing the manner in which auditors naturally adopt and evaluate hypotheses under different client situations. Subsequent experimental studies may then be guided by this evidence in evaluating whether auditors revise beliefs consistent with the way they frame the task.

Another area of research relates to the question of which of the four scenarios discussed is prevalent in audit practice. For instance, in interpreting the

example used in this study of an increase in gross margin, we found it difficult to envision client scenarios where the causes would be mutually exclusive. In fact, it appears that all of the causes discussed in our examples could co-exst. If mutually exclusive causes are, in fact, rare in audit practice, prior audit or psychology research, which assumes such a relationship, may offer limited insights into the evaluation of multiple hypotheses in auditing.

Further analytical modelling that considers the limitations in our analyses also appears promising. For example, the current study assumes that evidence directly implicates a *single* hypothesis. In practice, audit tests often provide differing degrees of assurance relating to multiple assertions. For example, an accounts receivable confirmation provides evidence concerning both existence and valuation. Future modelling may incorporate evidence that is viewed as impacting multiple hypotheses.

In addition, alternative probability paradigms such as belief functions[7] may be investigated to examine the extent to which they assist in dealing with contexts encountered in the audit environment. Finally, the costs and benefits of such paradigms need to be assessed through analytical, field, and experimental work.

Last, research into multiple-hypothesis settings has a number of implications for audit practice. Additional research is expected to provide insights into whether auditor behaviour appears appropriate under various hypothesis contexts as well as providing insights into the costs and benefits of alternative modelling approaches. This research has promise for improving audit effectiveness and efficiency. For instance, inappropriate assessment of multiple hypotheses may lead to incorrect audit conclusions in accepting an erroneous cause or failure to consider multiple causes.

If deficiencies in auditor behaviour are identified, the models may then serve as a basis for training or decision tools to assist auditors in dealing with the complex decision process entailed in multiple hypothesis evaluation. Early research has been completed within psychology that shows positive training effects (McKenzie, 1998). Finally, there is little guidance provided in professional standards and auditor training in this area. Analytical modelling of multiple hypothesis evaluation is important in addressing this gap and ultimately in contributing toward a theory of hypothesis reasoning in auditing.

## References

American Institute of Certified Public Accountants, 1988, Analytical Procedures. Statement on Auditing Standards No. 56, New York: AICPA.
Australian Auditing and Assurance Standards Board, 1995, Audit Evidence, Auditing Standard 502, AAuSB, Melbourne.
Australian Auditing and Assurance Standards Board, 1999, Irregularities, Including Fraud, Other Illegal Acts and Errors, Auditing Standard 210, AAuSB, Melbourne.

---

[7] See Shafer 1976, Srivastava and Shafer 1992 and Srivastava and Mock 2002.

Anderson, J. C., S. Kaplan, and P. Reckers, 1992, The effects of output interference on analytical procedures judgments, *Auditing: A Journal of Practice & Theory* 11 (Fall), 1–13.

Anderson, U. and L. Koonce, 1995, Explanation as a method for evaluating client-suggested causes in analytical procedures, *Auditing: A Journal of Practice & Theory* 14 (Fall), 124–132.

Anderson, U. and L. Koonce, 1998, Evaluating the sufficiency of causes in audit analytical procedures, *Auditing: A Journal of Practice & Theory* 17 (Fall), 1–12.

Asare, S. and A. Wright, 1995, Normative and substantive expertise in multiple hypotheses evaluation, *Organizational Behavior and Human Decision Process* 66 (December), 171–84.

Asare, S. and A. Wright, 1997a, Evaluation of competing hypotheses in auditing, *Auditing: A Journal of Practice & Theory* (Spring), 1–13.

Asare, S. and A. Wright, 1997b, Hypothesis revision strategies in conducting analytical procedures, *Accounting, Organizations and Society* (November), 737–756.

Asare, S. and A. Wright, 2000, The interdependence of hypothesis generation and information search on performance in conducting analytical procedures, Working paper (Boston College).

Ashton, A. and R. Ashton, 1988, Sequential belief revision in auditing, *The Accounting Review* 63 (October), 623–641.

Bauer, D. and J. Eddy, 1986, The representation of command language syntax, *Human Factors*, 1–10.

Bedard, J. C. and S. Biggs, 1991a, The effect of domain-specific experience on evaluation of management representations in analytical procedures, *Auditing: A Journal of Practice & Theory* 10, (Supplement); 77–90.

Bedard, J. C. and S. Biggs, 1991b, Pattern recognition, hypotheses generation, and auditor performance in an analytical task, *The Accounting Review* 66 (July), 622–42.

Bhattacharjee, S., T. Kida and D. Hanno, 1999, The impact of hypothesis set size on the time efficiency and accuracy of diagnostic decision makers, *Journal of Accounting Research* (Spring), 83–100.

Biggs, S. 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, 234–55.

Biggs, S. F., T. J. Mock and P. R. Watkins, 1988, Auditor's use of analytical review in audit program design, *The Accounting Review* 63 (January), 148–161.

Blocher, E., 1991, *Answers: An Audit Planning Software System*, (Prentice Hall Professional Software: Atlanta, GA).

Blocher, E. and J. Willingham, 1993, *Analytical Review: A Guide to Performing Analytical Procedures* (New York: McGraw-Hill).

Bonner, S. E. and N. Pennington, 1991, Cognitive processes and knowledge as determinants of auditor expertise, *Journal of Accounting Literature*, 1–50.

Church, B. K. and A. Schneider, 1993, Auditors' generation of diagnostic hypotheses in response to a superiors' suggestion: Interference effects, *Contemporary Accounting Research* (Fall), 333–50.

Cushing, B. E. and S. S. Ahlawat, 1997, Mitigation of recency bias in audit judgment: The effect of documentation, *Auditing: A Journal of Practice & Theory*, 15, 110–122.

Dutta, S. K. and R. P. Srivastava, 1993, Aggregation of evidence in auditing: A likelihood perspective, *Auditing: A Journal of Practice &Theory* 12 (Supplement), 137–160.

Dutta, S. K. and R. P. Srivastava, 1996, A Bayesian perspective on the strength of evidence in auditing, *Indian Journal of Accounting*, XXVII June, 4–18.

Edwards, A. W. F., 1984, *Likelihood: An account of the statistical concept of likelihood and its application to scientific inferences* (Cambridge: Cambridge University Press).

Gaeth, G. and J. Shanteau, 1984, Reducing the influence of irrelevant information on experienced decision makers, *Organizational Behavior and Human Performance*, 263–282.

Hackenbrack, K., 1992, Implications of seemingly irrelevant evidence in audit judgment, *Journal of Accounting Research* (Spring), 125–135.

Heiman, V., 1990, Auditors' assessments of the likelihood explanations in analytical review, *The Accounting Review* 65, 875–90.

Heiman-Hoffman, V., D. Moser and J. Joseph, 1995, The impact of an auditor's initial hypotheses on subsequent performance at identifying actual errors, *Contemporary Accounting Research* (Spring), 763–779.

Hirst, D. E., 1994, Auditor's sensitivity to source reliability, *Journal of Accounting Research* (Spring), 113–126.

Hirst, E. and L. Koonce, 1996, Audit analytical procedures: A field investigation, *Contemporary Accounting Research* (Fall), 457–486.

Jamal, K., P. E. Johnson and R. G. Berryman, 1995, Detecting framing effects in financial statements, *Contemporary Accounting Research*, 85–105.

Jamal, K., P. E. Johnson and R. G. Berryman, 1997, Improving the effectiveness of the auditor in fraud detection. Working paper, University of Alberta.

Kennedy, J., 1993, Debiasing audit judgment with accountability: A framework and experimental results, *Journal of Accounting Research* (Autumn), 231–245.

Knapp, M., 1996, *Contemporary Auditing: Issues and Cases* (West Publishing Co: New York).

Krishnamoorthy, G., T. Mock and M. Washington, 1999, A comparative evaluation of belief revision models in auditing, *Auditing: A Journal of Practice & Theory*, 105–127.

Libby, R., 1985, Availability and the generation of hypotheses in analytical review, *Journal of Accounting Research* 23 (Autumn), 648–67.

Libby, R. and D. M. Frederick, 1990, Expertise and the ability to explain audit findings, *Journal of Accounting Research* 28 (Autumn), 348–67.

Luce, D. and H. Raiffa, 1957, *Games and decision: Introduction and critical survey* (New York: John Wiley & Sons, Inc).

McKenzie, C. R. M., 1998, Taking into account the strength of an alternative hypothesis, *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 24, 771–792.

Mock, T. J., A. Wright, R. Srivastava and H. Lu, 2001, The framing and evaluation of multiple hypotheses, Working paper (University of Southern California).

Monroe, G. S. and J. Ng, 2000, An examination of order effects in auditors' inherent risk assessments, *Accounting and Finance* 40(2), 153–68.

Morris, M. W. and R. P. Larick, 1995, When one cause casts doubt on another: A normative analysis of discounting in causal attribution, *Psychological Review*, 102, No. 2, 331–355.

Newell, A. and H. Simon, 1972, *Human problem solving* (Englewood Cliffs, N.J.: Prentice-Hall).

Ng, T. B. P., W. Green and R. Simnett, 1996, The effects of inherent risk and management representations on hypothesis generation, Paper presented at the International Symposium on Audit Research, Maastricht (June).

Nisbett, R., Zukier H. and R. Lemley, 1981, The dilution effect: Nondiagnostic information weakens the implications of diagnostic information, *Cognitive Psychology*, 248–277.

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

Shafer, G., 1976, *A mathematical theory of evidence* (Princeton University Press, Princeton, NJ).

Srivastava, R. P. and T. Mock, 2002, Introduction to belief functions', in, R. P. Srivastava and T. Mock, eds., *Belief functions in business decisions* (Physica-Verlag, Heidelberg, Springer-Verlag Company).

Srivastava, R. P. and G. Shafer, 1992, Belief-function formulas for audit risk, *The Accounting Review* (April), 249–283.

Srivastava, R. P., A. Wright, and T. Mock, 2001, Causal schemas and multiple hypotheses evaluation: A general framework, Working paper (School of Business, The University of Kansas).

Trotman, K. and A. Wright, 1996, Recency effects: Task complexity, decision mode, and task special experience, *Behavioral Research in Accounting*, 175–193.

Tubbs, R., W. Messier and W. Knechel, 1990, Recency Effects in the Auditor's Belief-Revision Process', *The Accounting Review* (April), 452–460.

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

Van Wallendael, L. and R. Hastie, 1990, Tracing the footsteps of Sherlock Holmes: Cognitive representations of hypothesis testing, *Memory and Cognition* 18 (1990), 240–50.

Winkler, R. and W., Hays, 1975, *Statistics: Probability, inference, and decision* (New York: Holt, Rinehart and Winston).

Winterfeldt, D.V., and W. Edwards, 1986, *Decision analysis and behavioral research*. (Cambridge University Press, New York, NY).

## Appendix A

In this appendix, we derive analytical formulas for posterior probabilities of various hypotheses (causes) that may be thought to be the reason for the observed effect, E, using Bayes rule. These probabilities represent the revised probabilities after having observed and analysed the impact of various items of evidence gathered in the process for determining the most likely hypothesis or hypotheses (i.e., cause or causes). Although Srivastava *et al.*, (2001) have developed a general model for posterior probabilities, we derive the current model with the following modifications for the present paper. First, we assume that the auditor has observed the effect E, i.e., there is no uncertainty in his/her finding; there is a material difference between this year's value and previous year's values. Second, we extend their formulation from a set of four hypotheses to a set of five hypotheses, $H_1$, $H_2$, $H_3$, $H_4$, and $H_5$, to compare our results with the empirical study of Asare and Wright (1995, 1997b). Third, we assume that an item of evidence, $E_i$, pertains to a single hypothesis, $H_i$ where i = 1, 2, 3, 4, and 5. Srivastava *et al.*, (2001) have discussed situations where one item of evidence pertains to more than one hypothesis.

Using Bayes rule, we can write the posterior odds[8] for $H_1$ after having gathered all the additional five items of evidence as:

$$\frac{P(H_1|EE_1E_2E_3E_4E_5)}{P(\sim H_1|EE_1E_2E_3E_4E_5)} = \frac{P(H_1EE_1E_2E_3E_4E_5)}{P(\sim H_1EE_1E_2E_3E_4E_5)} \tag{A-1}$$

where the symbol '~' in front of a hypothesis represents negation of the hypothesis. We want to express the right hand side of equation (A-1) in terms of the strength of evidence collected for each hypothesis. Using the conditional probability definition[9], we can express $P(H_1EE_1E_2E_3E_4E_5)$ as:

$$P(H_1EE_1E_2E_3E_4E_5) = P(E_1|H_1EE_2E_3E_4E_5)P(H_1EE_2E_3E_4E_5).$$

Also, we know that once the hypothesis $H_1$ is known to be true then knowing E, $E_2$, $E_3$, $E_4$, or $E_5$ does not tell us anything more about $E_1$. In other words, $E_1$ is conditionally independent of all other items of evidence, E, $E_2$, $E_3$, $E_4$, and $E_5$, given that we know $H_1$. Thus, we can write:

---

[8] Posterior odds is equal to the ratio of the posterior probability of the hypothesis H given all the evidence (P(H|all the evidence)) to the probability of ~H given all the evidence (P(~H|all the evidence)). We know from Bayes rule that $P(H|EE_1E_2E_3E_4E_5) = P(HEE_1E_2E_3E_4E_5)/P(EE_1E_2E_3E_4E_5)$, and $P(\sim H|EE_1E_2E_3E_4E_5) = P(\sim HEE_1E_2E_3E_4E_5)/P(EE_1E_2E_3E_4E_5)$. These values yield equation (A-1). For simplicity, we have suppressed the conjunction symbol ∩ in between the above variables. For example, $P(H|EE_1E_2E_3E_4E_5)$ stands for $P(H|E{\cap}E_1{\cap}E_2{\cap}E_3{\cap}E_4{\cap}E_5)$.

[9] In general, we have P(DF) = P(D|F)P(F).

$$P(E_1|H_1EE_2E_3E_4E_5) = P(E_1|H_1),$$

and hence,

$$P(H_1EE_1E_2E_3E_4E_5) = P(E_1|H_1)P(H_1EE_2E_3E_4E_5).$$

or

$$P(H_1EE_1E_2E_3E_4E_5) = P(E_1|{\sim}H_1)\lambda_1P(H_1EE_2E_3E_4E_5) \qquad (A\text{-}2)$$

where $\lambda_1$ is the likelihood ratio ($\lambda_1 = P(E_1|H_1)/P(E_1|{\sim}H_1)$) that represents the strength of evidence $E_1$ pertaining to $H_1$. In general, the likelihood ratio ($\lambda$) determines the strength of the evidence pertaining to a hypothesis (for details, see, e.g., Edwards 1984, and Dutta and Srivastava 1993, 1996). For a positive item of evidence, the likelihood ratio lies between one and infinity ($\infty \geq \lambda > 1$). A value of infinity represents an infinitely strong piece of evidence. If such an item of evidence is observed, the hypothesis is present with certainty. For a negative item of evidence, the likelihood ratio lies between zero and one (i.e., $1 > \lambda \geq 0$). A value of zero represents an infinitely strong negative piece of evidence, which implies that if such an item of evidence is observed, the hypothesis is definitely not present. A value of one for the likelihood ratio (i.e., $\lambda = 1$) represents a neutral (non-diagnostic) item of evidence. That is, the evidence has no information about the hypothesis. We want to express (A-2) in terms of the strength of evidence, $\lambda$, for each item of evidence. In order to achieve this we proceed in several steps as described below.

First, we introduce a vacuous state, $H_2 \cup {\sim}H_2$, in '$H_1EE_2E_3E_4E_5$', and use the property that two sets ($H_1H_2EE_2E_3E_4E_5$) and ($H_1{\sim}H_2EE_2E_3E_4E_5$) are mutually exclusive. Thus the term $P(H_1EE_2E_3E_4E_5)$ in (A-2) can be written as:

$$P(H_1EE_2E_3E_4E_5) = P(H_1H_2EE_2E_3E_4E_5) + P(H_1{\sim}H_2EE_2E_3E_4E_5)$$
$$= P(E_2|H_1H_2EE_3E_4E_5)P(H_1H_2EE_3E_4E_5)$$
$$+ P(E_2|H_1{\sim}H_2EE_3E_4E_5)P(H_1{\sim}H_2EE_3E_4E_5).$$

Again, in our situation since $E_2$ pertains to only $H_2$, once we know that $H_2$ is true then knowing $H_1$, $E$, $E_3$, $E_4$, and $E_5$ does not tell us anything new about $E_2$. In other words, $E_2$ is conditionally independent of $H_1$, $E$, $E_3$, $E_4$, and $E_5$ given $H_2$. Thus, we can write the above expression as:

$$P(H_1EE_2E_3E_4E_5) = P(E_2|H_2)P(H_1H_2EE_3E_4E_5) + P(E_2|{\sim}H_2)P(H_1{\sim}H_2EE_3E_4E_5),$$

or

$$P(H_1EE_2E_3E_4E_5) = P(E_2|{\sim}H_2)[\lambda_2P(H_1H_2EE_3E_4E_5) + P(H_1{\sim}H_2EE_3E_4E_5)], \quad (A\text{-}3)$$

where $\lambda_2$ is the likelihood ratio ($\lambda_2 = P(E_2|H_2)/P(E_2|\sim H_2)$) that represents the strength of evidence $E_2$ pertaining to $H_2$.

Next, the two terms, $P(H_1H_2EE_3E_4E_5)$ and $P(H_1\sim H_2EE_3E_4E_5)$, in (A-3) can be further expanded by introducing a vacuous state $H_3\cup\sim H_3$ in each of the parenthesis above. Following the steps similar to the ones used in arriving from (A-2) to (A-3) we can obtain:

$$P(H_1H_2EE_3E_4E_5) = P(E_3|\sim H_3)[\lambda_3P(H_1H_2H_3EE_4E_5) + P(H_1H_2\sim H_3EE_4E_5)],$$

and

$$P(H_1\sim H_2EE_3E_4E_5) = P(E_3|\sim H_3)[\lambda_3P(H_1\sim H_2H_3EE_4E_5) + P(H_1\sim H_2\sim H_3EE_4E_5)].$$

Continuing this process until all the evidence is exhausted, we obtain the following expression for $P(H_1EE_1E_2E_3E_4E_5)$:

$$P(H_1EE_1E_2E_3E_4E_5) = P(E_1|\sim H_1)P(E_2|\sim H_2)P(E_3|\sim H_3)P(E_4|\sim H_4)P(E_5|\sim H_5)P(E).N_1$$

where $N_1$ represents partly the impact of various items of evidence on hypothesis $H_1$ and partly the impact of the relationships among the hypotheses on $H_1$. The following expression defines $N_1$ where various joint probabilities together define the relationship among the hypotheses.

$$\begin{aligned}
N_1 = \lambda_1[&\lambda_2\lambda_3\lambda_4\lambda_5P(H_1H_2H_3H_4H_5|E) + \lambda_2\lambda_3\lambda_4P(H_1H_2H_3H_4\sim H_5|E) \\
&+ \lambda_2\lambda_3\lambda_5P(H_1H_2H_3\sim H_4H_5|E) + \lambda_2\lambda_4\lambda_5P(H_1H_2\sim H_3H_4H_5|E) \\
&+ \lambda_3\lambda_4\lambda_5P(H_1\sim H_2H_3H_4H_5|E) + \lambda_2\lambda_3P(H_1H_2H_3\sim H_4\sim H_5|E) \\
&+ \lambda_2\lambda_4P(H_1H_2\sim H_3H_4\sim H_5|E) + \lambda_2\lambda_5P(H_1H_2\sim H_3\sim H_4H_5|E) \\
&+ \lambda_3\lambda_4P(H_1\sim H_2H_3H_4\sim H_5|E) + \lambda_3\lambda_5P(H_1\sim H_2H_3\sim H_4H_5|E) \\
&+ \lambda_4\lambda_5P(H_1\sim H_2\sim H_3H_4H_5|E) + \lambda_2P(H_1H_2\sim H_3\sim H_4\sim H_5|E) \\
&+ \lambda_3P(H_1\sim H_2H_3\sim H_4\sim H_5|E) + \lambda_4P(H_1\sim H_2\sim H_3H_4\sim H_5|E) \\
&+ \lambda_5P(H_1\sim H_2\sim H_3\sim H_4H_5|E) + P(H_1\sim H_2\sim H_3\sim H_4\sim H_5|E)].
\end{aligned} \tag{A-4}$$

Based on the steps similar to the ones used in $P(H_1EE_1E_2E_3E_4E_5)$ in deriving (A-4), we express $P(\sim H_1EE_1E_2E_3E_4E_5)$ as:

$$P(\sim H_1EE_1E_2E_3E_4E_5) = P(E_1|\sim H_1)P(E_2|\sim H_2)P(E_3|\sim H_3)P(E_4|\sim H_4)P(E_5|\sim H_5)P(E).D_1$$

where $D_1$ represents partly the impact of various items of evidence on hypothesis $\sim H_1$ and partly the impact of the relationships among the hypotheses on $\sim H_1$. The following expression defines $D_1$ where various joint probabilities together define the relationship among the hypotheses as mentioned earlier.

$$D_1 = [\lambda_2\lambda_3\lambda_4\lambda_5 P(\sim H_1 H_2 H_3 H_4 H_5|E) + \lambda_2\lambda_3\lambda_4 P(\sim H_1 H_2 H_3 H_4 \sim H_5|E)$$

$$+ \lambda_2\lambda_3\lambda_5 P(\sim H_1 H_2 H_3 \sim H_4 H_5|E) + \lambda_2\lambda_4\lambda_5 P(\sim H_1 H_2 \sim H_3 H_4 H_5|E)$$

$$+ \lambda_3\lambda_4\lambda_5 P(\sim H_1 \sim H_2 H_3 H_4 H_5|E) + \lambda_2\lambda_3 P(\sim H_1 H_2 H_3 \sim H_4 \sim H_5|E)$$

$$+ \lambda_2\lambda_4 P(\sim H_1 H_2 \sim H_3 H_4 \sim H_5|E) + \lambda_2\lambda_5 P(\sim H_1 H_2 \sim H_3 \sim H_4 H_5|E)$$

$$+ \lambda_3\lambda_4 P(\sim H_1 \sim H_2 H_3 H_4 \sim H_5|E) + \lambda_3\lambda_5 P(\sim H_1 \sim H_2 H_3 \sim H_4 H_5|E)$$

$$+ \lambda_4\lambda_5 P(\sim H_1 \sim H_2 \sim H_3 H_4 H_5|E) + \lambda_2 P(\sim H_1 H_2 \sim H_3 \sim H_4 \sim H_5|E)$$

$$+ \lambda_3 P(\sim H_1 \sim H_2 H_3 \sim H_4 \sim H_5|E) + \lambda_4 P(\sim H_1 \sim H_2 \sim H_3 H_4 \sim H_5|E)$$

$$+ \lambda_5 P(\sim H_1 \sim H_2 \sim H_3 \sim H_4 H_5|E) + P(\sim H_1 \sim H_2 \sim H_3 \sim H_4 \sim H_5|E)]. \qquad \text{(A-5)}$$

Equation (A-1) provides the posterior odds in terms of $N_1$ and $D_1$ as:

$$\frac{P(H_1|EE_1E_2E_3E_4E_5)}{P(\sim H_1|EE_1E_2E_3E_4E_5)} = \frac{N_1}{D_1} \qquad \text{(A-6)}$$

The posterior odds given in (A-6) directly yields[10] the following expression for the posterior probability in terms of the strength of all the items of evidence and the joint probabilities of all the five causes (hypotheses) given that effect E has been observed.

$$P(H_1|EE_1E_2E_3E_4E_5) = N_1/(N_1 + D_1) \qquad \text{(A-7)}$$

Equation (A-7) that determines the posterior probability of $H_1$ after having observed and evaluated all the evidence is obtained from successive application of Bayes rule. In fact, the numerator $N_1$ in (A-7) represents the impact of all the evidence on hypotheses $H_1$ in conjunction with the assumed relationships among the hypotheses whereas the denominator $(N_1 + D_1)$ represents the impact of all the evidence on $H_1$ and $\sim H_1$ in conjunction with the assumed relationships among the hypotheses. Similar expressions[11] can be derived for the other four causes.

Equation (A-7) gives a general model for any type of perceived relationship among the five assumed potential causes (hypotheses) with the constraint that each additional item of evidence pertains to only one cause or hypothesis.

---

[10] Posterior probability = (posterior odds)/(1 + posterior odds) = (N/D)/(1 + N/D) = N/(N + D).

[11] These models are available from the authors on request.