Belief Functions in Business Decisions, edited by R. P. Srivastava and T. Mock, Physica-Verlag, Heidelberg, Springer-Verlag Company (forthcoming).

Auditors' Evaluations of Uncertain Audit Evidence: Belief Functions versus Probabilities

Keith E. Harrison, Rajendra P. Srivastava, R. David Plumlee

Truman State University, University of Kansas, University of Utah[1]

Abstract

Recently, Shafer and Srivastava [1], Srivastava and Shafer [2], Srivastava [3]-[4], and Van den Acker [5] have identified appealing features of belief function evidential networks. These networks can express the support that audit evidence provides for assertions, accounts and financial statements. These networks can also aggregate many pieces of evidence into an overall level of support for a particular assertion, account or an entire set of financial statements.

There is little empirical evidence about the ability of practicing auditors to express their evaluations of the strength of audit evidence in terms of belief functions. Many traditional models assume the use of probabilities. These might be called the traditional type of subjective probabilities. They are additive by definition, i.e. **P**(a) + **P**(~a) = 1. Throughout the remainder of this paper they will simply be referred to as probabilities. This study examines the question of expressing the support provided by audit evidence empirically. Auditors are asked to express the level of support that evidence provides for or against an assertion or account and the ignorance that remains about the assertion or account after considering the evidence.

Many auditors who use probabilities to measure risk express ignorance by giving equal weight to support for and support against the objective. Belief functions express ignorance by allocating mass to all elements of the frame (all possible

1 The authors would like to acknowledge the generous support of the Deloitte & Touche Foundation through their Doctoral Fellowship program that supported the research in this project.

outcomes), an alternative that is distinct from assigning equal mass to each element of the frame. Thus, belief functions show ignorance as an amount separate from the amounts of support for and against the objective.

Forty-nine experienced auditors were asked to estimate the strength of the evidence provided in twenty-eight different audit situations. These auditors were given the opportunity to model their estimates of support and ignorance in ways that were consistent either with belief functions or with Probabilities. In this study, a statistically significant percentage of the auditors represented their estimates in ways that were consistent with belief functions and were inconsistent with probabilities. This suggests that future practitioner decision aids may include belief functions as a way of naturally expressing the ignorance and risk that persist in many audit engagements.

1. Introduction

Evidence is a key component of any audit process. Professional standards require that audit conclusions be supported by enough persuasive evidence to survive a challenge from skeptical critics. It seems natural to consider Dempster-Shafer belief functions as a way to describe the support provided by evidence as well as the ignorance that one feels even after gathering that evidence. Beginning with Shafer and Srivastava [1] and Srivastava and Shafer [2], Srivastava and his colleagues have explored the theoretical possibilities of designing belief function models for expressing audit evidence. Some of these papers include Srivastava [3-4], Van den Acker [5], Srivastava and Lu [6], Dutta, Harrison and Srivastava [7], Gillett [8] and Srivastava, Dutta and Johns [9]. This paper will begin to examine the empirical issues of whether auditors express their evaluations about the strength of evidence and the existence of ignorance in ways that are similar to belief functions.

First, the paper provides some background information on financial statement audits. Next, there is a section discussing the semantics of belief functions, The following section reviews the literature about empirical investigations of the use of belief functions including the application of belief functions to auditing situations. Then there is a section that develops the hypothesis of this paper, a section that describes the methods used in the experiment, a section that presents results and briefly discusses them and the paper ends with a brief conclusion.

2. Financial Statement Audits

Financial statement audits are a key component of the American and world economies. Arens and Loebbecke [10] provide a useful definition of a financial statement audit in a leading auditing textbook:

> Auditing is the accumulation and evaluation of evidence about information to determine and report on the degree of correspondence between the information and established criteria. Auditing should be done by a competent independent person.

Auditors examine financial statements and issue an auditor's report on them. To issue such a report, the auditors must be reasonably sure that the financial statements are correct. To do this, they must gather, evaluate and aggregate evidence about the financial aspects of the company. Gathering evidence about the financial statements as a whole is important and necessary, but more specific information about the company's financial aspects requires examination of individual accounts and various dimensions of those accounts.

An account is a record of the business transactions that affect one specific asset, liability, equity, revenue or expense. The cash account is an account that records the affects of all of the business transactions that involve a company's cash. Most companies report sales and this appears as a revenue account. Effective evidence gathering procedures require that auditors subdivide accounts. Each account has various dimensions, each of which must be tested by the auditors. These dimensions are called management assertions because management asserts that all dimensions of the account are correct when they give the financial statements to the auditors to be audited.

Auditors create a set of objectives to test each of these management assertions called audit objectives. There may be one or more than one audit objective for each management assertion.

The audit process consists of gathering enough evidence to support each one of these audit objectives for each account in the financial statements. Different types of evidence provide different levels of support for audit objectives.

The support that audit evidence provides is almost never certain. Some doubt remains about all evidence because people make mistakes or intentionally mislead others for their own advantage. The process of combining pieces of evidence also contributes to uncertainty because there are not uniform procedures for combining different items of evidence that relate to a single audit objective or a single account.

Auditors are to consider audit risk in every audit. Professional auditing standards [11, paragraph .02] define audit risk as "the risk that the auditor may unknowingly fail to appropriately modify his opinion on financial statements that are materially misstated." One aspect of audit risk is inherent risk. Professional auditing standards [11, paragraph .20] also define this risk, "inherent risk is the susceptibility of an assertion to a material misstatement, assuming that there are no related internal control structure policies or procedures." Assessing inherent risk for a particular assertion may involve many factors about the company, its operating environment and the overall economy. Making this assessment for every assertion in the financial statements may be very costly and time consuming. In such cases, the professional standards suggest [11, paragraph .22]:

> If an auditor concludes that the effort required to assess inherent risk for an assertion would exceed the potential reduction in the extent of his auditing procedures derived from such an assessment, he should assess inherent risk as being at the maximum when designing auditing procedures.

If one is using probabilities to represent audit risks, this leads to the inherent risk component of audit risk being set at 1.0 or 100%. Srivastava and Shafer [2, p. 252] use this situation to point out one of the difficulties in using probabilities to express audit risk:

> We advance belief functions … because we believe that the usefulness of the Bayesian approach is limited by divergences between the intuitive and Bayesian interpretations of audit risk. For example, according to SAS No. 47, if an auditor decides not to consider inherent factors, then the inherent risk is set equal to 1. Since a probability measure of 1 means certainty, this seems to be saying that it is certain that the account is materially in error. But this is not what the auditor has in mind when deciding not to depend on inherent factors. The auditor's intention is represented better by a belief-function plausibility of 1 for material error, which says only that the auditor lacks evidence based on inherent factors.

Professional standards ask auditors to proceed with an open mind. They should neither assume that an amount is correct nor assume that an amount is incorrect before gathering evidence. The auditor simply has a high level of ignorance about all types of risk, including inherent risk, at the beginning of the audit. If this problem is viewed from a belief function perspective, Inherent Risk is the plausibility of material misstatement due to inherent factors. Under this consideration, IR = 1 implies that the plausibility of material missttatement due to inherent factors is 1, i.e. $\mathbf{PL}_{\mathbf{IF}}$(Material mistatement) = 1, where IF stands for inherent factors.

With the probability framework, most decision-makers will use the convention of assigning equal probability to each alternative to express ignorance. For example, if there are only two alternatives, 50% is assigned to each alternative. However, several interpretations of this information are possible. It is not possible to distinguish between 50% support for each alternative, complete ignorance, or some situation in-between. These different interpretations lead auditors to different conclusions. Clarity about which of them exists improves an auditor's understanding of audit risk. Belief functions can provide this clarity because belief functions express ignorance separately from the support for any of the alternatives in a situation. With belief functions, the inherent risk example can be modeled by setting ignorance of inherent risk at 100% and then seeking evidence to provide support for lowering audit risk in other areas such as internal controls.

Auditors also face difficult issues in knowing how to model the relationship between evidence and the audit objective or account that it supports. For example, analytical procedures are audit tests that examine the relationships between various pieces of financial (and non-financial) information about a company. Some relationships can be predicted because of industry trends and other internal and external factors. Analytical procedure tests provide evidence not only about audit objectives but also about accounts and financial statements as a whole. This means that an auditor must be able to consider evidence that supports different types of variables (financial statements, accounts and audit objectives) simultaneously. Again Srivastava and Shafer [2] demonstrate that a belief function model arranged in tree-type structures can handle evidence that supports multiple levels of financial variables.

Another issue that auditors frequently face is the problem of one piece of evidence that supports more than one audit objective simultaneously, often with differing levels of support for each. One type of evidence about accounts receivable, a confirmation letter, is usually thought to provide evidence for both the existence audit objective and the valuation audit objective. Many auditors think that it provides a higher level of support for existence than for valuation. In such a case, the tree-type structures developed in Srivastava and Shafer [2] are not theoretically applicable because this one piece of evidence would need to be treated as though it were two independent pieces of evidence. In practice, the independence assumption cannot always be made. In this case, an evidential network can be used. Such a network allows an item of evidence to support multiple audit objectives with a different level of support for each. Srivastava [3] develops such an evidential network for the belief function framework.

Almost all of the work in belief function evidential networks (for example, Srivastava and Shafer [2], Srivastava [3], [4] and Mock, Wright and Srivastava [12]) has been theoretical and not empirical. The theoretical arguments for using belief function evidential networks have been strong and intuitively appealing, but no empirical testing of the use of belief functions in an audit setting has been performed. This paper addresses that need for empirical testing.

3. The Semantics of Belief Functions

Background on the basic elements of belief functions can be found in the introductory chapters of this volume. In this section, one perspective for interpreting and understanding belief functions from a semantic perspective will be presented. The experiment described in this paper is related to this understanding and interpretation of belief functions.

Shafer and Tversky [13] and Shafer [14] clearly describe belief functions as an alternate language of probability, not as something distinct from probability. Each language permits the user to construct different types of analogies between evidence and mental models that the user may be familiar with from past experience.

Shafer and Tversky [13] also redirect the traditional argument between the subjective and objective understandings of probability into a middle-ground interpretation of probability where the meaning of the probability data is constructed each time the decision-maker encounters a situation requiring reasoning with uncertain information. The meaning is constructed from canonical examples that the decision-maker has learned from his or her experience or has been taught in the past. These examples can be frequency events for physical objects such as tosses of fair coins. They can also be subjective judgments about mental experiments such as the relationship between management displaying certain attitudes and the increased risk of fraud. Shafer and Tversky [11, p. 336] describe probability judgment in this way:

> Probability judgment is a process of construction rather than elicitation. People may begin a task of probability judgment with some beliefs already formulated. But the process of judgment, when successful, gives greater content and structure to these beliefs and tends to render initial beliefs obsolete. It is useful, in this respect, to draw an analogy between probability and affective notions such as love and loyalty. A declaration of love is not simply a report on a person's emotions. It is also a

> part of a process whereby an intellectual and emotional commitment is created; so too with probability.

Shafer and Srivastava [1, p. 119] state this point somewhat more directly:

> Numerical probabilities usually do not have objective reality independent of human judgment. Nor do they exist in people's minds prior to deliberation. But people can construct numerical probabilities on the basis of objective evidence.

The constructive interpretation gains depth from the idea that probability judgment involves matching practical problems to abstract canonical examples. Shafer and Tversky describe canonical examples in the following way [11, 310-311]:

> We also need canonical examples for each degree of probability on this scale–examples where it is agreed what degree of probability is appropriate.

They continue [11, p. 311]:

> Using a theory of subjective probability means comparing the evidence in a problem with the theory's scale of canonical examples and picking out the canonical example that matches it best.

Belief functions differ from probabilities in the type of canonical examples they use, but not in the process of constructing analogies. Belief functions use indirect canonical examples because they establish a relationship between a situation with known probabilities and a situation with unknown probabilities. The situation with the unknown probabilities is the situation of interest. Belief functions become useful when the decision-maker cannot estimate the probabilities to the question of interest, but can estimate probabilities for some related question. The example below illustrates this point.

Company X has a control procedure where a computer program verifies the general ledger account coding for transactions prepared in the field by sales representatives. Different types of sales receive different types of commissions. Auditor testing shows that on days with a lot of sales activity across the country, the director of computing can disable the general ledger coding control procedure in order to get customer orders processed and shipped in a timely fashion. This means that the general ledger code supplied by the sales person will automatically be used without being checked by the computer. The auditor estimates that the control operates for 85% of all sales transactions. The auditor is not primarily interested in the control procedure itself. The auditor wishes to give an opinion on the financial statements and whether or not the amount of commission expense is correctly stated. The auditor believes that if the control is operating, the amount of commission is calculated correctly. However, it may also be calculated correctly if the control is not functioning. No individual sales representative will know when the system has been turned off since the transactions are coded during the day on a laptop computer and transmitted to headquarters in the evening. The sales representative codes the transaction assuming that the control will be in effect. If the control is in effect, attempts to create overstated commissions will be detected and stopped.

Belief functions are useful in this example because the auditor is unable to gather direct evidence about the fact that really concerns the auditor, the existence of errors in the sales commission amount shown in the financial statements. Evidence can be gathered on the related question of whether the control was operating at the time of the event. The auditor has developed the relationship that if the control was turned on, the amount in the financial statement is correct. If the control was not turned on, the data may either be correct or incorrect. Suppose that the auditor estimates that the control is in operation 85% of the time. Combining the facts and the relationship between sets of facts gives the auditor a belief that at least 85% of the commissions are processed correctly and no belief that the commissions are processed incorrectly. Up to 15% of the transactions may be processed either correctly or incorrectly when the controls are not in operation. This judgment can be expressed in terms of ***m***-values as $\boldsymbol{m}(a) = 0.85$, $\boldsymbol{m}(\sim a) = 0$, and $\boldsymbol{m}(\{a,\sim a\}) = 0.15$, or in terms of beliefs as $\boldsymbol{Bel}(a) = 0.85$ and $\boldsymbol{Bel}(\sim a) = 0$, where "*a*" represents the situation that sales commissions are fairly stated and "~*a*" represents the situation that sales commissions are not fairly stated.

This example can be described using a belief function canonical example because belief functions do not require additivity. In other words, $\text{Bel}[A] + \text{Bel}[\sim A] \leq 1$. This is one of the properties that makes belief functions particularly useful in audit settings. Many of the situations that auditors encounter can best be modeled by this sort of canonical example. A belief function canonical example is designed to take evidence that is known to the decision-maker but is only indirectly relevant and relate it to the question of interest for which the decision-maker has less ability to gather information. Shafer describes one type of canonical example that can be used for belief functions [12, 338]:

> In my mind, the simplest and most effective metaphor for belief functions is the metaphor of the witness who may or may not be reliable. In many cases, the example of such a witness can serve as a standard of comparison for the strength of evidence. We can assess given evidence by saying that it is comparable in strength to the evidence of a witness who has a certain chance of being reliable.

> A witness testifying to a specific proposition leads to a relatively simple belief function—one that gives a specified degree of belief to that proposition and its consequences and zero degree of belief to all other propositions.

Shafer and Tversky [13] suggest that probability is a useful canonical example for expressing frequencies. To illustrate this, a decision-maker might say that the probability of finishing the current project in less than one week is 25% and the probability of finishing in one week or more is 75%. She bases this evaluation on similar projects with similar amounts of work to complete that she has led. She is using past frequencies to model this uncertainty and expressing this as a probability.

As Shafer and Srivastava [1] note, some situations that decision-makers face can best be handled by probabilities. If this is the case, they should be used. Probabilities and belief functions can be used together in the same model of evidence because probabilities are a subset of belief functions. Probabilities are belief functions that follow the rule of additivity (they can be called Bayesian belief functions). Shafer and Srivastava [1, p. 130] provide some additional insight about how probabilities are related to belief functions. For example, they state "The Bayesian formalism has two elements—the idea of a probability measure and the rule of conditioning. Both of these elements have their place in the belief-function formalism." They also discuss additivity, "The general point is that both (Bayesian) additivity and (non-Bayesian) nonadditivity are permitted in the belief function formalism. A belief function does not have to be a probability measure, but it can be one."

4. Empirical Investigations into the Use of Belief Functions

Experiments that explore the empirical use of belief functions are very limited at this time. Several bear mentioning, Curley and Golden [15], Krishnamoorthy, Mock and Washington [16] Dusenbury, Reimers and Wheeler [17] and Monroe and Ng [18].

Curley and Golden [15] explored the use of belief functions within a simulated legal context. Business students acted as jurors hearing a case with four possible suspects and up to four pieces of evidence that supported one or more of these suspects. With four suspects, "a", "b", "c" and "d", many different subsets of guilty suspects could be created. For example, evidence item one might implicate subjects "a" and "d" but not the other suspects. While many subsets of subjects potentially could be assigned mass, under belief function theory only {a}, {d}, {a,d} and {a,b,c,d} should be assigned mass. Curley and Golden measured which frames were assigned mass by the participants. The researchers grouped responses into categories. For example, all responses that assigned mass only to {a,d} were included in one category. Other responses that chose only {a} and {d} were included in another category.

The researchers were able to identify patterns in the ways that participants assigned mass. With this level of complexity, the researchers wished merely to understand whether participants assigned mass in a manner consistent with Dempster-Shafer belief-function theory.

Their results show that the participants did assign mass to subsets that were logically consistent with Dempster-Shafer belief functions. This is important because it provides a basis for the current study. It is consistent with what the auditors were expected to do. However, participants in the Curley and Golden study did not combine evidence in ways that also were consistent with Dempster's Rule for the combination of belief functions. Since Dempster's Rule of combination is one of the mathematical requirements of belief functions, this difference between theory and observed behavior should be explored in future research.

Krishnamoorthy, Mock and Washington [16] compared four theoretical models for revising belief with the results produced by auditors combining pieces of evidence in an experimental setting. The four methods included Cascaded Inference Theory (Bayesian inference), Dempster-Shafer belief functions and two versions of Einhorn and Hogarth's Belief Adjustment Model. Although both belief functions and cascaded inference correctly predicted the direction of the auditor assessments, they produced results that were overly conservative when compared to the participants' responses and to the results from the Belief Adjustment Models. However, information about the values of some inputs to the models was not elicited from the participants and was assumed or interpolated. While this may be reasonable in a situation where several competing models are being tested, it appears that further work could help to determine the usefulness of belief functions for auditors. The current study addresses this issue.

Dusenbury, Reimers and Wheeler [17] asked experienced auditors from a major auditing firm to assess inherent risk, control risk and analytical procedures risk for two accounting cycles. The allowed test of details risk (TDR) was derived from these assessments. The allowed TDR was derived using three separate audit risk models. The models were (1) the model described in Statements on Auditing Standards, (2) a model developed by the auditing firm and (3) a belief-based specification of the audit risk model.

This study also asked auditors to assess two other constructs. Participants rated the sufficiency of the evidence and their own confidence ratings of their assessments of risk. The authors used these constructs to construct a belief ratio that reflected an estimate of the ignorance that auditors felt about the evidence and their evaluation of the evidence. This study differs from the current work in this key dimension. The researchers created the belief-based measures based on inputs rather than asking the participants to estimate support and ignorance directly.

The results of this study show some of the potential benefits of a belief-based model [15, p. 12]. It demonstrated that “a belief-based model can incorporate qualities of evidence into risk assessments and that belief-based assessments can be combined to derive and allowed test of details risk.” Neither of the other models captured the uncertainty that the auditors felt about the evidence that they had gathered. Similar to [16], the belief-based model in [17] produced results that were more conservative than the other models examined.

Monroe and Ng [18] compared the intuitive assessments of audit risk made by 69 practicing Australian auditors to the assessments provided by various audit risk models, including a belief function audit risk model. The estimates for the component inputs to these models were elicited from the same auditors. The results reported indicate that all of the models produced a risk assessment that was inconsistent with the risk assessment produced intuitively by the auditors. Fairly large errors were reported for the Belief-Adjustment and Belief function models. The authors identify several possible reasons for the limitations of all of the models, including the use of the assessment of risk only at the account-balance level. This study raises the possibility that many levels of risk and many characteristics of an engagement may need to be evaluated simultaneously in order to achieve an acceptable risk assessment from any audit risk model.

This paper expands the work of these studies by giving auditors the opportunity to directly assess the ignorance that remains in their minds after evaluating evidence.

5. Hypotheses

Shafer and Tversky recognize that no one model of uncertainty will fit all problems. They state [11, p.311]:

> Proponents of different theories of subjective probability have often debated which theory best describes human inductive competence. We believe that none of these theories provide an adequate account of people’s intuitive judgments of probability. On the other hand, most of these theories can be learned and used effectively. Consequently, we regard these theories as

> formal languages for expressing probability judgments rather than as psychological models, however idealized.
>
> The usefulness of these formal languages for a specific problem may depend both on the problem and on the skill of the user. There may not be a single probability language that is normative for all people and all problems. A person may find one language better for one problem and another language better for another. Furthermore, individual probability judgments made in one language may not be directly translatable into another.

This leads to a testable hypothesis for this study. If one formulates uncertain situations that the theory suggests should be expressed more effectively in one probability language as opposed to another, subjects presented with that situation should choose to model the problem using the preferred probability language. If the theory is not supported, subjects may randomly choose a particular probability language for expressing uncertainty or may consistently choose a probability representation because of its greater familiarity.

In the specific case of audit evidence, many procedures provide one-sided evidence. Often, evidence supports the audit objective being tested and provides no support for contradicting the audit objective being tested. This was true in the internal control example given earlier. This sort of one-sided support is described in the belief function framework as a simple support function. Theoretically, it would seem that participants should have a clear preference for using belief functions in auditing because simple support functions describe many auditing situations. Therefore, it becomes important to determine if the theory is consistent with practice.

This presents the basic research question of this experiment. When faced with evidence that provides support for an audit objective or account and no support against this objective or account, do auditors model the support provided by evidence through the use of belief functions? Said another way, will the auditors choose to represent ignorance as a separate, specific part of their evaluation of the audit evidence?

In the experiment, auditors were asked to determine whether a piece of evidence supported or did not support a particular audit objective (or account in the case of a combined evaluation of several pieces of evidence). They also were asked if a separate amount of ignorance remained about the audit objective or account. Auditors that chose to represent a separate amount of ignorance showed a preference for a belief function representation because it is easier to represent ignorance in belief functions than in probabilities.

Auditors that chose to identify no separate ignorance category may have been showing a preference for a probability representation. Probability is discussed in college statistics classes and included in guidance in the professional literature. Note however, that the representations that exclude a separate ignorance category can easily be modeled using belief functions and are completely consistent with a belief function approach.

If auditors were indifferent between a belief function approach and a probability approach for modeling the strength of evidence, one would expect to find that auditors choose to use them about equally. One might even expect that the probability approach might show up more frequently, since some professionals are familiar with this approach from training. However, if separate identification of ignorance occurs, this provides evidence that auditors are choosing a belief function representation over a probability representation.

Let p_{bf} represent the proportion of the auditors' evaluations that are consistent with a belief functions representation and inconsistent with a Bayesian representation. Let p_p represent the proportion of the auditors' evaluations that are consistent with a probability representation[2]. If there were a preference for a probability representation, or indifference between the two representations, one would expect that the proportion of evaluations that are consistent with a belief function representation, p_{bf}, would be less than or equal to ½. This is equivalent to $p_{bf} \leq p_p$. However, if auditors use a representation consistent with belief functions, we would expect that p_{bf} would be greater than ½ and $p_{bf} > p_p$. The hypothesis can be stated as follows:

$$H_{0:}\ p_{bf} \leq 0.5\ .$$

$$H_A\text{: } p_{bf} > 0.5\ .$$

The next sections of this paper will describe how the experiment was conducted.

---

2 Representations where the auditor feels that the evidence completely supports (or completely refutes) the audit objective or account can be represented equally well under belief functions or subjective probabilities. In this paper they have been arbitrarily assigned to the subjective Bayesian probability case because this will make it harder to reject the null hypothesis and provides a stronger test of the hypothesis.

6. Methods Used in this Experiment

6.1. Participants

In this section, the participants in the experiment are identified and the accounting firms that they represent are described. Accounting firms in Missouri and Iowa were contacted. One office from six different firms agreed to participate in the research. Of these, two firms only participated during the design phase and are not included in the table. The type of firm and the number of participants that participated in the experiment are listed in Table 1.

A reasonable level of experience in evaluating and combining audit evidence is necessary in order for an individual to complete the experiment successfully. Partners and managers from several firms and the experimenters agreed that a minimum of one year of audit experience would be needed to carry out these tasks. All 49 met that requirement.

Table 1: Summary by Firm of the Participants in the Study

Firms participating in the experiment	Number of Participants
"Big Five" firms	26
Large regional firm	15
Large local firm	8
Total number of participants in the experiment	49

Auditors assisting during the design phase evaluated whether the descriptions of the pieces of evidence made sense and asked the participants to identify the audit objective or objectives that they believed each piece of evidence supported. Most of these auditors were seniors or managers with three to six years' experience.

The pilot-testing phase used auditors from a large national accounting firm to test the accuracy and quality of the experimental materials. These auditors ranged in experience from one to five years.

Participants in the experiment came from four CPA firms located in Missouri. Two of these firms are "Big Five" accounting firms. One of the "Big Five" firms employed sixteen of the participants and the second "Big Five" firm employed ten of the participants. These participants had between one and four years of auditing experience.

Fifteen auditors from a large national firm (i.e. a firm located throughout the United States but not a “Big Five”) also participated in this study. These auditors ranged in experience from three to twenty years’ experience.

Finally, eight auditors from a very large one-office firm completed the experiment. These auditors ranged in experience from one to four years.

Auditors from one of the “Big Five” firms and from the National firm completed the experiment as part of a training meeting. Auditors from the other firms completed the experiment individually. They returned the experiment to a local manager or partner who returned the completed experiments to the researchers.

6.2. Materials Used in the Experiment

The purpose of the experiment is to determine whether auditors express their evaluations of the strength of audit evidence in terms that are more consistent with belief functions or probabilities. This requires the use of items of audit evidence. To save time, auditors were given brief descriptions of typical items of audit evidence rather than the actual audit evidence. Five descriptions of audit evidence were provided in four different audit areas. Figure 1 shows one of these descriptions.

In designing the descriptions of the evidence for the experiment, certain simplifying features were implemented. First, all of the areas of evidence pertain to one company, Midwest Industries. This company was designed to be familiar to a wide range of auditors. It is involved in a traditional, profitable industry that does not involve high-risk business or accounting activities. It was designed to have strong management and strong internal control.

This type of company provides a relatively familiar environment for the participants to work with. It allows them to focus on the particular pieces of evidence that are presented in each evidence set without having a large number of other variables influencing their decision in ways that could be hard to predict or control. Such variables could include a high risk of fraud or increased risks of litigation. This positive set of conditions would exist on some but certainly not all of an auditor’s clients. At the same time, they are not unrealistic.

Four different audit areas were used in this experiment. They include accounts receivable, accounts payable, plant, property and equipment and inventory quantities. These areas were chosen for several reasons. These areas generally are familiar to most auditors. Accounts receivable is the audit area where the amount owed to a company is tested. Accounts payable is the audit area where the amount owed by a company is tested. Plant, property and equipment is the audit area where tangible assets with long lives such as machinery and buildings are tested.

All three of these areas would be involved in almost every audit, regardless of the industry or type of business.

Inventory is the audit area where the assets that are available for sale to customers are tested. Inventory can be either purchased or manufactured. Inventory will not be involved in every audit because some companies provide services instead of goods. The inventory evidence set was limited to “quantity” issues and did not discuss the cost or value of the inventory. Eliminating the cost and value issues makes it possible for more auditors to understand the evidence because most auditors will have observed a physical inventory count.

Within each of the audit areas, the participants saw five individual pieces of evidence. They were asked to evaluate the strength of each of these pieces of evidence individually. They also were asked to evaluate the combined strength of the first four pieces of evidence before they saw the fifth piece of evidence. After evaluating the fifth piece of evidence, they evaluated the combined strength of all five pieces of evidence.

For two of the four areas, the last piece of evidence provided support for two different audit objectives within the area. This difference is not significant to the results of this experiment.

Each auditor that participated in the study received a folder of materials. These materials included:

1. Two copies of an informed consent.[3]

2. An introduction to the research project explaining what the research was being conducted to study.

3. A practice set of materials for notes payable, an area that is not included in the experiment.

4. Experimental materials for four audit areas. These materials included:

    a) Background information for the hypothetical company, Midwest Industries.

    b) Four sets of audit evidence. Each set includes:

        i) Four pieces of evidence, each on a separate page.

[3] The experimental materials and consent forms were approved through the institutional review process for human experimentation at a major midwestern university.

ii) A page for providing an evaluation of the combined strength of the first four pieces of evidence.

iii) A fifth page of evidence.

iv) A page for providing an evaluation of the combined strength of all five pieces of evidence.

5. A demographic information page that collected data related to professional characteristics of the individual involved.

A sample page for recording the level of support provided by the first piece of evidence for the inventory audit area is shown in Figure 1.

Figure 1: A Sample Page for Collecting Information about the Support that Evidence Provides

Midwest Industries Audit December 31, 1998 Working Paper I-1 Audit Objective: Existence

The audit team members observing the inventory count selected a sample of tags listed on the used tag summary sheet and found that a copy of all of these tags were attached to actual inventory. Tags not attached to inventory were all designated as unused. The staff noted that all inventory movement was halted during the count to minimize the risk of double counting inventory. Later, once the inventory had been summarized, a staff auditor reviewed the inventory summary listed by tag number and found that none of the unused tags were included. The staff auditor also verified that used tags appeared only once on the inventory summary listing.

0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10

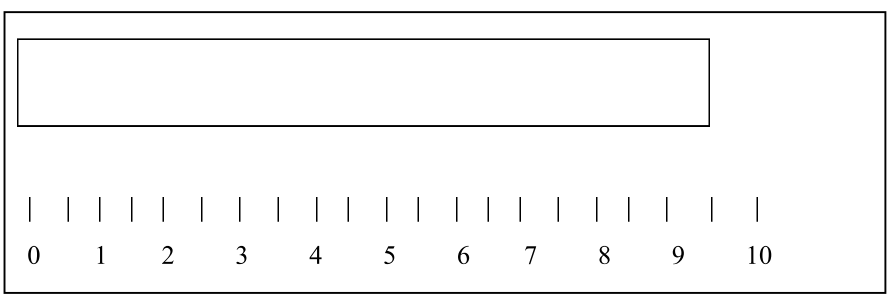
6.3. Experimental Procedures

Each auditor was asked to complete the twenty-eight individual and combined evidence evaluation tasks that were described in the previous section. Auditors at two of the firms completed the exercise during a firm training sessions and auditors at the other two firms completed the exercise individually and returned the materials to a firm coordinator. The coordinators returned the materials to the researchers. Oral instructions were provided during the firm training sessions and written instructions were provided for auditors completing the evaluations independently. The same set of instructions was provided in both cases. There was no significant statistical difference between the evaluations of auditors performing the task in training sessions and the evaluations of auditors performing the task individually.

Each of the auditors completed an informed consent form, provided some demographic data, reviewed some instructions and completed a practice exercise before beginning the experimental evidence evaluation task. No significant questions were raised during the practice exercise and none of the auditors appeared to have difficulty with the task. It appears that the materials were clear and easy for the auditors to understand.

Auditors were asked to document their evidence evaluations by dividing the judgement box at the bottom of each page. Figure 2 shows a sample judgment box. Each piece of evidence was identified with one or two audit objectives that related to the evidence. Groups of four or five pieces of evidence were associated with an account that related to all of the evidence for that area.

Figure 2: Judgment Box



Participants were asked to completely fill the large box at the bottom of each page by dividing the box into one, two or three sections and then labeling each box with a letter related to the way they thought the uncertainty in each situation would be resolved. Participants were asked to fill in the judgment box with one, two or all three of the following choices. Choice "A" represented the amount of support that the evidence set provided for the audit objective identified in the scenario. Choice "B" represented the amount of support that the evidence provided against the audit objective. Choice "C" represented the amount of ignorance that the auditor continued to have about the audit objective after seeing the piece of evidence.

The auditors divided the judgement box to show the amount they wished to assign to each of the choices. For example, the auditor drew a line at "6" if she or he wished to allocate 60% of the box to "A" and 40% of the box to "C". The auditor labeled the larger section of the box with an "A" and the smaller section of the box with a "C". The judgment box remained undivided if the auditor wished to allocate the entire box to just one of the three choices. Table 2 summarizes the choices of "A", "B", and "C" made by the auditors.

Table 2: Summary of Auditors Assignment of Support for Individual Evidence and Combinations of Evidence

	Choices that auditors made to record their support for the evidence in the various scenarios*						
Column Number	I	II	III	IV	V	VI	VII
Choice from Instructions	A	B	C	AB	BC	AC	ABC
Probability Representation	P(a)	P($\sim a$)		P(a), P($\sim a$)			
Belief function Representation (non-zero ***m***-values)	$\boldsymbol{m}(a)$	$\boldsymbol{m}(\sim a)$	$\boldsymbol{m}(\{a, \sim a\})$	$\boldsymbol{m}(a)$ and $\boldsymbol{m}(\sim a)$	$\boldsymbol{m}(\sim a)$ and $\boldsymbol{m}(\{a,\sim a\})$	$\boldsymbol{m}(a)$ and $\boldsymbol{m}(\{a,\sim a\})$	$\boldsymbol{m}(a)$, $\boldsymbol{m}(\sim a)$ and $\boldsymbol{m}(\{a,\sim a\})$
Firm 1, N = 450	50	1	0	4	3	356	36
Percentage	11.1%	0.2%	0.0%	0.9%	0.7%	79.1%	8.0%
Firm 2, N = 480	49	1	1	24[a]	1	300	104
Percentage	10.2%	0.2%	0.2%	5.0%	0.2%	62.5%	21.7%
Firm 3, N = 239	51	0	1	20[b]	0	134	33
Percentage	21.3%	0.0%	0.4%	8.4%	0.0%	56.1%	13.8%
Firm 4, N = 300	71	0	4	10	1	136	78
Percentage	23.7%	0.0%	1.3%	3.3%	0.3%	45.3%	26.0%
Total, N = 1,469	221	2	6	58	5	926	251
Percentage	15.0%	0.1%	0.4%	3.9%	0.3%	63.0%	17.1%

*The letters in these columns correspond to the types of support. Choice A shows support for the audit objective being true, choice B shows support for the audit objective not being true and choice C shows the amount of ignorance the auditor feels about the situation. Columns that have multiple letters indicate that the auditor used all of the letters shown.

Each evaluation appears in only one column. For example, no evaluations included in the column labeled "A" are included in the column labeled "AC". An auditor that used "AC" thought that the evidence provided some support for the audit objective or account but also left some ignorance about the audit objective or account.

[a] One individual recorded ten items in this category.

[b] One individual recorded ten items in this category and another individual recorded seven items in this category.

7. Results and Discussion

The evaluations of the evidence are presented in Table 2. Each auditor assigned one, two or all three of the letters (A,B,C) described above to each evidence evaluation. Any evaluation that involves letter C represents an evaluation that is consistent with a belief function representation of the evidence and inconsistent with a Bayesian representation of the evidence. Each evaluation is included in only one column. For example, if an auditor thought that evidence provided some positive support for the auditor objective but felt that some ignorance remained, that individual would choose to use both A and C and their choice would be included in column VI in the table. That evaluation would not be included in columns I or III.

In total, 1,469 evaluations were made. Each of the 49 auditors evaluated 26 items where the evidence supported one assertion or account and two items where the evidence supported two different assertions and a separate evaluation was made for each. This means that each auditor made 30 evaluations. This gives a theoretical total of 1,470 evaluations (49 x 30 = 1,470). One auditor omitted one evaluation and the total of 1,469 in Table 2 reflects this omission.

Evaluations that appear in columns I, II and IV are consistent with a probability representation. Items in columns III, V, VI, and VII are only consistent with a belief function representation of uncertainty. Table 3 summarizes the choices that are consistent with belief functions and probabilities.

Table 3: Summary of Results by Representation

	Belief Function Columns III, V, VI, VII	Probability Columns I, II, IV	Total
Number of items	1,188	281	1,469
Proportion	0.81	0.19	1.00
Probability that ($p_{bf} \leq 0.50$) = .0000			

The results from Table 3 provide strong evidence that the null hypothesis that the proportion of evaluations using a belief function representation will be less than or equal to ½ (i.e. $p_{bf} \leq 0.50$) is rejected. This probability is nearly zero and is zero to four decimal places. Over 4/5 of the evaluations use a belief function representation to express the auditor's evaluation of the evidence. This indicates that these auditors think about the evidence problem in terms of support and ignorance rather than in terms of support for and support against the audit objective being true.

Perhaps this is most clearly demonstrated in the data that suggests that just a few auditors do use a probability representation for the evidence. Three auditors accounted for 27 of the 58 uses of the "AB" category. This category shows that the evidence provides some support for and some support against but leaves no ignorance about the evidence. It appears that these three auditors were most comfortable using a probability approach to modeling evidence. Since three auditors provide almost half of the responses in the "AB" category, other auditors use it very infrequently. Most auditors used a belief function representation. For those auditors that use a Bayesian representation, a belief function representation approach can be used since all Bayesian representations can also be represented in belief functions as described earlier. Gillett [8] also found similar results with a very small sample of auditors that he contacted during his research.

This research does not address the issue of whether auditors combine evidence in a manner consistent with the axioms of belief functions. Results from Curley and Golden [15], and Krishnamoorthy, Mock and Washington [16] suggest that more work may need to be performed to understand how auditors do combine pieces of evidence. It also suggests that training of auditors might be needed if an auditing firm wished to utilize a belief function model to combine evidence and arrive at an overall evaluation for an audit area or the financial statements as a whole. This is not unusual. Most decision aids require some training before they can be properly utilized by decision-makers. Srivastava, Dutta and Johns [9] demonstrate that a realistic audit process for one portion of an audit can be constructed using a belief function based expert system. This current research suggests that auditors will be comfortable in providing inputs to such a system in a belief function format.

Further work remains to develop a full-scale belief function expert system capable of modeling an entire audit.

8. Conclusion

This paper has shown that a clear majority of the auditors in this study chose to represent the uncertainty inherent in audit evidence through belief functions rather than through probabilities, when both choices were available to them. Two reasons appear to explain this preference. First, belief functions represent ignorance as a separate explicit component of the evaluation, rather than describing it indirectly by assigning some ignorance to each of the possible outcomes. Second, belief functions can represent support for an audit objective or account without showing any support against the audit objective or account. This is not surprising since Shafer [19] designed belief functions as a mathematical theory of evidence.

This result will advance the research into developing a belief function expert system for audit decisions. It now appears that auditors can provide evaluations of evidence in ways that are consistent with belief functions and it also appears that these evaluations will be useful for these auditors. Further research is needed to determine the type of training that would be needed to improve the use of belief functions by auditors. Auditors are willing to express support for audit objectives in a belief function framework because ignorance is specifically quantified.

References

[1] G. Shafer and R. P. Srivastava, (1990) The Bayesian and belief-function formalisms: A general perspective for auditing. *Auditing: A Journal of Theory and Practice*, vol. 9 (Supplement), pp. 110-148.

[2] R. P. Srivastava and G. R. Shafer, (1992) Belief-function formulas for audit risk. *The Accounting Review*, vol. 67 (April), pp. 249-283.

[3] R. P. Srivastava, (1995a) A general scheme for aggregating evidence in auditing: Propagation of beliefs in networks. In *Artificial Intelligence in Accounting and Auditing*, vol. 3. Miklos A. Vasarhelyi, Ed. Princeton, NJ: Markus Wiener Publishers

[4] R. P. Srivastava, (1995b) The belief-function approach to aggregating audit evidence. *International Journal of Intelligent Systems*, vol. 10 (March) pp. 329-356.

[5] C. Van den Acker, (1999) A Belief-Function Model for the Representation and the Combination of Uncertain Audit Evidence. *Intelligent Systems in Accounting, Finance and Management*, vol. 8 (3), pp. 215-224.

[6] R. P. Srivastava and H. Lu, (2000) Evidential Reasoning Under Uncertainty Using Belief Functions: A Structural Analysis of Audit Evidence. *Fuzzy Sets and Systems*, Forthcoming.

[7] S. K. Dutta, K. E. Harrison and R. P. Srivastava, (1998) The audit risk model under the risk of fraud. *Applications of Fuzzy Sets and the Theory of Evidence to Accounting, II* (vol. 7 in the *Studies in Managerial and Financial Accounting* series). Philip H. Siegel, Khursheed Omer, Andre de Korvin and Awni Zebda, Eds. Stamford Connecticut: JAI Press Inc.

[8] P. R. Gillett, (1996) *A Comparative Study of Audit Evidence and Audit Planning Models Using Uncertain Reasoning.* Doctoral dissertation. University of Kansas, Lawrence, KS.

[9] R. P. Srivastava, S. K. Dutta and R. W. Johns, (1996) An Expert System Approach to Audit Planning and Evaluation in the Belief-Function Framework. *Intelligent Systems in Accounting, Finance and Management*, vol. 5, pp. 165-183.

[10] A. A. Arens and J. K. Loebbecke, (1997) *Auditing: An integrated approach*. 7th ed. Englewood Cliffs, NJ: Prentice Hall.

[11] American Institute of Certified Public Accountants, (1983) *Statement on Auditing Standards No. 47: Audit Risk and Materiality in Conducting an Audit*. New York: AICPA.

[12] T. Mock, A. Wright and R. Srivastava, (1998) Audit Program Planning Using a Belief Function Framework. *AUDITING SYMPOSIUM XIV; Proceedings of the 1998 Deloitte & Touche/University of Kansas Symposium on Auditing Problems. pp. 115-142.*

[13] G. Shafer, and A. Tversky, (1985) Languages and designs for probability judgment. *Cognitive Science*, vol. 9, pp. 309-339.

[14] G. Shafer, (1990) Perspectives on the theory and practice of belief functions. *International Journal of Approximate Reasoning*, vol. 4 pp. 323-362.

[15] S. P. Curley and J. I. Golden, (1994) Using belief functions to represent degrees of belief. *Organizational Behavior and Human Decision Processes*, vol. 58, pp. 271-303.

[16] G. Krishnamoorthy, T. J. Mock, and M. T. Washington, (1999) A Comparative Evaluation of Belief Revision Models in Auditing. *Auditing: A Journal of Theory and Practice*, vol. 18 (Fall), pp. 104-127.

[17] R. Dusenbury, J. L. Reimers, and S. Wheeler, (1996) An Empirical Study of Belief-Based and Probability-Based Specifications of Audit Risk. *Auditing: A Journal of Theory and Practice*, vol. 15 (Fall), pp. 12-28.

[18] G. Monroe and J. Ng, (2000) The Efficacy of Models of Audit Risk. *Belief-Functions in Business Decisions,* edited by R. Srivastava and T. Mock.

[19] G. Shafer, (1976) *A Mathematical Theory of Evidence*. Princeton, NJ: Princeton University Press.