Covariation Assessments with Costly Information Collection in Audit Planning: An Experimental Study

Ananda R. Ganguly*
Visiting Assistant Professor
Purdue University
Krannert School of Management
403 W. State Street
West Lafayette, IN 47907-2056
United States of America

Jacqueline S. Hammersley
Assistant Professor
University of Georgia
JM Tull School of Accounting
238 Brooks Hall
Athens, GA 30602
United States of America

Version 5.0.8

* Corresponding author. Fax: +1 765 494-9658.
E-mail address: argangul@purdue.edu

We thank Mark Bagnoli, Efrim Boritz, Carla Carnaghan, Audrey Gramling, Tom Hall, Kathryn Kadous, Derek Koehler, Natalia Kotchetova, Susan Krische, Marlys Lipe, Mary-Beth Mohrman, Mark Peecher, Alan McNaughton, Steve Salterio, Teri Shearer, Scott Vandervelde, Alan Webb and participants at the 2004 AAA Auditing Section Midyear Meeting and at workshops at Purdue University, Queen's University, and the Universities of Missouri-St. Louis, Texas-Arlington and Waterloo for helpful comments on earlier drafts of this paper. We also gratefully acknowledge the research assistance of Kelly Jones and the help of Rajib Doogar, Bob Halperin and our students in the acquisition of participants and in the running of the experiments.

Covariation Assessments with Costly Information Collection in Audit Planning:

An Experimental Study

Abstract

In this paper we report the results of an experiment that investigates the impact of two types of costly information acquisition on subsequent covariation-estimate revisions in audit planning. Covariation estimates assess the degree of association between a 'Clue' that auditors might observe, and its potentially associated 'Condition.' We find that when participants can choose which cells of the covariation table to purchase frequency information about (i.e., Selective Collection), they purchase less than all the required information in specific biased ways. These biased information choices lead to predictable biases in covariation-estimate revisions. However, when participants buy a sample of representative information containing all four cells of the covariation table (i.e., Non-Selective Collection), their covariation-estimate revisions are in the normative direction. This suggests that covariation-estimation errors are likely to be biased in a predictable direction when information acquisition is judgmental and costly. However, we find that when participants are sensitized to the importance of all four cells of information, they purchase more cells of the covariation table in Selective Collection, which increases the likelihood the information they receive is representative. Participants subsequently make covariation revisions in the normative direction. This suggests that sensitizing participants to the importance of all cells of the covariation table improves covariation revisions mostly due to improvements in information selection rather than due to changes in information processing.

Keywords: Auditing; Audit Planning; Covariation; Correlation; Costly Information

Introduction

During audit planning, auditors often must estimate the possible association between clues observed when gaining an understanding of the client and the presence of material misstatement in the financial statements. For example, while obtaining an understanding of the internal control system during the planning stages of a new audit, auditors may discover a material weakness in the design of the internal control system (i.e., a clue). The auditor must estimate the likelihood that the presence of this material weakness is related to the presence of a material misstatement (i.e., an associated condition) in order to efficiently plan the extent of substantive testing necessary.

In such cases, auditors are likely to start with initial beliefs about the degree of association (or covariation) between the presence or absence of the clue and the presence or absence of the condition based on their theoretical knowledge of accounting practice and transaction cycles as well as by comparison with analogues (Bedard and Biggs 1991; Marchant 1989). They will update their beliefs by collecting empirical information regarding the relationship between this type of material weakness and a material misstatement from previous audit engagements, searches of databases directed at the question, and industry statistics, etc.[1] The updated estimate will then be used to help determine how much substantive testing is necessary. If this estimate is too high (low), auditors will likely overemphasize (underemphasize) the related substantive testing.

A covariation estimation task normatively requires the use of four pieces of information: the frequencies in the four cells of a 2 x 2 contingency table representing the

[1] Covariation estimations can occur in substantive testing as well, but are much more complex than what we study in this paper. Since substantive testing is almost always iterative, complete historical frequency data as presented in a contingency table likely will not be available on a client through one query without numerous iterative searches.

intersection of the presence or absence of the observable clue with the presence or absence of the associated condition (Cheng and Novick 1997). Some covariation research reports that people place too much emphasis on the presence of clues and conditions relative to the absence of clues and conditions (see Baron 1994 or Plous 1993, Chapter 15, for reviews). Most frequently, the result of such emphasis is a tendency to overestimate these associations even when no association exists (Garnham and Oakhill 1994). However, Lipe (1990) performs a lens model analysis combining data from five previous studies. She concludes that participants' aggregate covariation judgments match normative benchmarks quite well, and that information from all cells of the 2 x 2 table is utilized appropriately. This contradicts the findings of individual studies that show that some cells are over- or under-weighted. Additionally, other studies report that participants' covariation-estimate revisions generally are consistent with statistical models of cell usage (e.g., Vallée-Tourangeau et al. 1998; Waller and Felix 1987) even though the covariation levels often are misestimated. Thus, previous research is inconclusive about people's ability to make covariation estimations.

Moreover, previous research has not examined covariation assessments when obtaining requisite information is costly; most importantly for auditing settings, previous research has not separated the impact of costly information collection from cognitive processing in covariation estimation. Any insufficiency in obtaining information affects the information available for cognitive processing and the resulting covariation estimate. Some research in other domains demonstrates that biased information search due to a motivational cause biases evaluation of authoritative support in tax research (Cloyd and Spilker 1999) and that time pressure may have similar detrimental effects on audit performance (McDaniel 1990). However, other research demonstrates that allowing tax professionals control over the order of evidence collection mitigates the recency bias

(Cuccia and McGill 2000). So the impact of choosing information on the subsequent covariation assessment is not clear when the information is costly to obtain. Further, the specific impacts of information search patterns on covariation estimation have not been examined previously in either auditing or psychology research.

This paper reports the results of an experiment in which participants must choose the type of costly information necessary for estimating the covariation (i.e., degree of contingency or correlation) between an observable clue and its associated condition in an audit planning setting. We manipulate how participants receive information. Participants in the Selective Collection condition choose about which cells of the contingency table to obtain data; participants in the Non-Selective Collection condition receive information about all four cells of the contingency table. We hypothesize that participants in the Selective Collection condition will use certain predictable strategies to collect data and that use of this information will lead to predictably biased upward covariation-estimate revisions for some of these strategies.[2] Alternatively, we hypothesize that participants in the Non-Selective Collection condition, who receive all of the necessary data, will make covariation-estimate revisions in the normative direction.

Consistent with expectations, we find that participants in the Selective Collection condition obtain frequency information on a limited number of cells using predicted collection strategies. Covariation estimates then are updated based on the information collected, favoring the initial hypothesis of association when a more adequate collection of data would not support the hypothesis. Overall, these results suggest that when decision makers with limited resources collect costly information, they collect less than

[2] When previous research has reported a bias in covariation assessments, generally the bias has been a tendency towards over-estimation rather than under-estimation [Garnham and Oakhill 1994, p. 175]. Accordingly, we do not focus on erroneous under-estimation.

all the necessary information in predictable patterns, despite monetary incentives. The resulting covariation assessments tend to erroneously reinforce initial beliefs, but this appears due more to inadequacies in information collection rather than to inappropriate cognitive processing of information obtained.

Also as predicted, participants in the Non-Selective Collection condition update their covariation estimates downward (i.e., in a normative direction). This is consistent with previous research showing that people use information appropriately in making covariation estimates when all relevant information is provided (Lipe 1990; Waller and Felix 1987; Ward and Jenkins 1965). Overall our results suggest that people can make predictable covariation estimation errors despite using all *available* information correctly because the errors are likely due more to inadequacies in information collection than to information processing.[3]

Lastly, we find evidence that when participants are sensitized to the importance of all four cells of information, they choose to purchase more cells of the contingency table in Selective Collection. They continue to update their covariation estimates appropriately based on the information selected; however, because the information they select is now more complete, they generally make covariation revisions in the normative direction. It appears that sensitizing participants to the importance of all four cells of the covariation table improves covariation revisions mostly due to improvements in information selection, rather than due to changes in information processing.

Our findings have important implications for audit planning and our understanding of covariation estimation. When auditors try to estimate the association between a clue and a condition while planning the audit for a new client, a new business

[3] The terms "available information" and "availability of information" are used throughout this paper in the sense of availability for cognitive processing (Plous 1993; Libby 1985; Tversky and Kahneman 1973). They are not used in the sense of being available for collection or available for purchase.

model, or in an unfamiliar statutory regime, their estimates are likely affected by whether they use their limited resources to ask directed questions to elicit specific frequencies, or whether they methodically obtain frequency information to complete all four cells of the contingency table. In addition, our results indicate that sensitization to the importance of all the information in the covariation table should have important applications in the design of decision aids in auditing, and the training of audit personnel. Relatedly, our results suggest that decision aids may need to focus more on information collection than on information processing. Finally, this study makes the methodological contribution of using a design in which participants reveal, through their information-purchase decisions, the real demand for information. Thus, instead of using a regression framework to identify cells of information that seem to "explain" participants' covariation assessments, or a questionnaire methodology wherein the preference reported may differ from the actual information obtained, our design allows us to directly identify the pieces of information participants consider most valuable from a set of costly pieces of information.

These results also have applicability outside the audit planning domain. Similar covariation estimation tasks are common at other stages of the auditing process and in managerial, financial, and tax accounting, and in business decision-making in general. Examples include associations between (the presence or absence of) specific misstatements discovered in substantive testing and (the presence or absence of) misstatements in other accounts; between particular business processes and 'red flags' in key performance indicators; between specific debt covenants and certain accrual decisions; between particular tax rules and tax-avoiding or tax-evading actions; etc.

The rest of this paper is organized as follows. Section II presents the theory and develops the hypotheses. Section III describes the method. The results are analyzed in

Section IV, followed by a summary and discussion of the contributions and limitations in Section V.

Hypotheses Development

"Covariation assessment" and "contingency assessment" are terms used in psychology to describe the assessment of the degree to which two binary variables are related (Cheng and Novick 1997; Nisbett and Ross 1980). Consider, for example, the information presented in Table 1 about the association between the presence and absence of a long-term contract (the Clue) with the presence or absence of material misstatement (the Condition). The normatively correct assessment of covariation requires the use of all four cells of the table. One must make the following comparisons: material misstatement present versus material misstatement absent when long-term contracts exist (16:64) and the same ratio when long-term contracts do not exist (4:16). In the case of Table 1, the ratio is 1:4 in both cases, and this sample thus shows no association between the two variables according to most normatively accepted statistical measures.[4]

****

Insert Table 1 here

****

Some previous research suggests that there are various ways people process this type of information when all of the data is available to them and is costless. Some people are reported to attend very closely to the present-present cell (Cell A in the table) and assess a high covariation if the frequency in that cell is high (Smedslund 1963, p.165).

[4] There is some disagreement in the psychological judgment literature on the appropriate measure of the relationship between the two variables of interest (see Allan 1980 or Lipe 1990 for reviews). However, using statistical theory as the norm, a summary measure of the covariation between two binary variables is the χ^2 statistic. As shown in a footnote to Table 1, the χ^2 measure for the data in Table 1 is zero. Another normative measure of covariation used in statistical literature is Pearson's phi (ϕ). Since $\phi=\sqrt{(\chi^2/N)}$, it is also equal to zero when $\chi^2=0$.

But one of the most robust and oft-replicated findings has been that people are overly influenced by the co-occurrence of two events and insufficiently influenced by instances in which one event occurs without the other (see Baron 1994; Plous 1993, Chapter 15, for reviews). Garnham and Oakhill (1994) conclude that "people have difficulty in detecting correlations or, more particularly, detecting their absence" (p. 175).

However, other research questions these findings of non-normative judgment, reinterprets prior findings, and ultimately reports evidence consistent with accurate covariation assessment (Vallée-Tourangeau et al. 1998; Baker et al. 1996; Shanks 1995; Waller and Felix 1987). Lipe (1990) performs a lens model analysis using data from several prior studies. She concludes that participants' aggregate covariation judgments appear highly correlated with Pearson's *phi*, a result that is consistent with the use of all four cells by participants in making their covariation assessments.

The datasets that Lipe (1990) and others analyze differ in ways that make overall conclusions difficult. First, unlike in audit planning settings, experimental instructions in these studies often emphasize presence-related and absence-related information differently. For example, of 34 data points used in Lipe (1990), 16 are from studies stressing all cells, 17 are from studies stressing at least some absence-related information in addition to the presence-related information, and one stresses only presence-related information. Some other studies (e.g., Beyth-Marom 1982) as well as sensitivity analyses reported by Lipe (1990) indicate that these instructions likely affect participants' cell-choice strategies. Second, unlike in auditing settings where information has to be collected, some studies provide the data to participants in summarized form, which leads to greater utilization of data from all cells than when data is provided on an item-by-item basis (Ward and Jenkins 1965). Lastly, none of these studies make data costly to obtain.

To summarize, three conditions typical of audit planning settings have not been investigated in previous covariation estimation research. First, auditors must often choose which information to obtain in order to confirm or disconfirm an initial hypothesis or estimate before they begin the cognitive estimation task. Second, collecting the information is difficult, time consuming, or otherwise costly. Third, the auditors almost always have a limited budget. For example, auditors must start in the planning phase of an audit with expectations of relationships between different types of misstatement conditions and clues that would be associated with them (Brown et al. 1999; Heiman-Hoffman et al. 1995; McMillan and White 1993; Church 1991; Libby 1985; Einhorn 1976). These initial hypotheses will likely be based on (1) knowledge of the business model, accounting procedures, and transaction cycle, (2) theoretical predictions of the impact of different control characteristics (or lack thereof) on different financial statement conditions, and (3) casual or anecdotal observations of analogues (Bedard and Biggs 1991; Marchant 1989). The initial hypotheses or estimates of covariation likely will be tested for empirical support by collecting frequency information from past audit engagements, obtainable peer data, and industry historical and base-rate statistics, but such evidence will generally be obtained on a limited budget.[5]

These conditions are particularly important when auditors are faced with new clue-condition pairs with which they have little prior experience (e.g., auditors planning internal control audits in the United States in 2005 or audits of dot.com companies in the late 1990s). Initial hypotheses are then largely "theoretical," and collection of base-rate information is particularly costly. Searches of databases such as Lexis-Nexus or queries submitted to an accounting firm's own research department for information are common

[5] There may or may not be a formal budget for this stage of background planning. However, auditors have limited resources, so they will have to be selective.

at this stage of planning. In any of these cases, an insufficiency in the collection of information will affect the information available for cognitive processing, and thus the resulting covariation estimate. In turn, excessively high covariation estimates will adversely affect the efficiency of planned substantive audit procedures.

Previous psychology research demonstrates that people often test hypotheses using a *positive-test strategy*. That is, they tend to test cases that are expected to have the property of interest rather than those expected to lack that property (Johnston 1996; Pinkley et al. 1995; Church 1990; Klayman and Ha 1989, 1987). Previous research also reports that people tend to use a positive-test strategy more frequently under adverse time pressure and other constraints (Gorman and Gorman 1984; Klayman and Ha 1985). Auditing research reports that professional auditors are often prone to decision making biases (Bamber et al. 1997), especially when incentives reward efficiency (Brown et al. 1999). Therefore, when people search for information to update their covariation estimates, the information is costly, it can be selectively obtained, and budgets are limited, we expect that they will use a positive-test strategy.

Klayman and Ha (1987) discuss a variety of ways in which people can use a positive-test strategy. One strategy, using only two cells of information, is to examine instances in which the Clue is present (Cells A and B) and estimate how often, among those cases, the Condition is also present (Cell A). Klayman and Ha (1987) call this "+H testing," or testing for instances that fit the Hypothesized Clue. The other two-cell positive-test strategy is to examine instances in which the targeted Condition occurs (Cells A and C) and estimate how often, among those cases, the Clue is also present (Cell A). Klayman and Ha (1987) call this "+T testing," or testing for instances that fit the Targeted Condition. When using a positive-test strategy, we do not expect that participants will prefer cell pair AB over cell pair AC or vice versa. However, we do expect that they will choose these cell pairs more frequently than any others.

In addition to employing two-cell strategies, some participants may choose to do both a +H test and a +T test, employing a three-cell strategy collecting information from cell triple ABC. There is no other three-cell strategy that is consistent with positive testing. Therefore, we expect participants to choose this triple more frequently than other three-cell strategies.

We expect that other strategies will be much less prevalent than these two- and three-cell strategies. In particular, we do not expect participants to use a four-cell strategy (although it is normatively the best strategy for an accurate estimate) because of the cost involved. *If single-cell strategies are used*, we expect participants will choose Cell A most often as it represents an extreme application of a positive-test strategy. Although Cell A-only is reported as a common choice in previous covariation research which provided all data to participants (Baron 1994; Plous 1993; Nisbett and Ross 1980; Jenkins and Ward 1965; Smedslund 1963), our theory leads us to predict that this single-cell strategy will <u>not</u> be employed more frequently than the two-cell strategies predicted above. We expect that our subjects, due to economic incentives, will be more concerned about the adequacy of collected information and generally will not use simplistic single-cell strategies.

However, note that all positive-test strategies require the use of Cell A. Two-, three- and four-cell positive-test strategies additionally require the use of Cells B and C. Only a four-cell strategy requires the use of Cell D. This leads to the following hypothesis:

H1: When choosing which cells of costly information to obtain to inform a covariation estimate, Selective Collection condition participants will use a positive-test strategy which implies they will choose Cell A most often, followed by Cells B and C, followed by Cell D.

We expect that the information-collection effects predicted above will affect the resulting covariation revisions. That is, people who choose fewer than four cells of information will believe that they have all the relevant information and will update their covariation assessments based on the sample information. Based on evidence from prior research, we expect that once the information is collected, it will be processed appropriately (e.g., Vallée-Tourangeau et al. 1998; Shanks 1995; Waller and Felix 1987; Beyth-Marom 1982). We expect that any bias in the covariation-estimate revisions will mainly be due to the information available, not due to information processing. Specifically, we expect that participants will start with priors (i.e., initial hypotheses) of a positive association between the Clue and the Condition. The frequency information that our participants could then obtain is shown in Table 1. The statistically correct covariation level determined from this 'sample' is zero. Since the information provided represents a sample and not the entire population, participants need not believe that the information is perfectly representative of the population. Participants can either rely on the information obtained or not (Hirst 1994; Hackenbrack 1992). However, irrespective of the degree to which participants rely on the information obtained, *normatively* the updated covariation assessment can be only less than or equal to the prior, provided that participants have obtained the information in all four cells.[6]

Participants who use a pure +H test will choose Cell AB information. In processing this information they will note that the Condition is absent four times as often

[6] If participants rely on the collected information (i.e., the sample) to any degree at all, then their covariation revisions should be negative, given that they have data from all cells. Alternatively, if participants do not rely on the collected information, they should not revise their priors, and covariation estimates should not change.

as it is present when the Clue is present (64:16, Table 1), indicating a highly negative relationship between the variables. Consequently, we expect participants will either not change their priors or will revise their estimates downwards. Conversely, participants who use a pure +T test will select Cells AC. These participants will see that the Clue is present four times as often as it is absent when the Condition is present (16:4, Table 1). This will indicate a highly positive relationship between the variables and we expect that participants either will not change their priors or will revise their estimates upwards.[7]

As indicated earlier, we do not expect that other two-cell strategies will be as common as the choice of Cells AB and AC. However, when participants choose other two-cell combinations, the above logic can be extended to predict the revision direction. Thus, participants who choose Cells AD will find that presence of the condition occurs with presence of the clue, and absence of the condition occurs with absence of the clue, with equal frequency. This seems to indicate (erroneously) a perfect correlation between the two variables, so we expect positive covariation revisions. Similarly, we expect that participants who choose Cells BD will revise their estimates downward as the Clue is *present* more often than not (64:16) when the Condition is *absent*; participants who choose Cells CD will revise their estimates upward as the Condition is *absent* more often than not (16:4) when the Clue is *absent.* However, choice of Cells BC leads to a weaker prediction because both cells show cases when the Condition is *not* associated with the

[7] There is no statistically acceptable unambiguous rule for updating priors based on such incomplete information. The predictions made in this section therefore are descriptive, not normative. Further, we expect these predictions to hold only when participants perceive the information as indicative of a more extreme association (negative or positive) between the two variables than the participants' own priors. Otherwise, if participants perceive the 16:4 ratio in the Cell AC example above as indicative of a smaller correlation than their priors, they might revise their estimates downwards even after collecting Cell AC information. However, this was not generally expected to happen as the priors were expected to be moderate, and we designed the available information (Table 1) to be 'extreme' in order to get unambiguous results.

presence of the Clue, indicating a downward revision; but the row or column totals cannot be determined, leading to some ambiguity.

We expect participants who choose three-cell strategies to obtain Cells ABC most often as we expect they will use both +H and +T testing. These participants will find their initial hypothesis disconfirmed in 68/84 cases, and we expect them to revise their covariation estimates downward. Using similar logic on other three-cell choices, we expect choosing Cells ABD will lead to downward covariation revisions (as the initial hypothesis is disconfirmed in 64/96 cases); we expect choosing Cells ACD will lead to upward revisions (as the initial hypothesis is confirmed in 32/36 cases); finally we expect choosing Cells BCD will lead to downward revisions (as the initial hypothesis is disconfirmed in 68/84 cases).

Recall that we expect that biases in covariation estimates are due mainly to information availability, not information processing. Consequently, we expect that participants who receive information from all four cells (i.e., in Non-Selective Collection or by choosing Cells ABCD in Selective Collection) will revise their covariation estimates downward using this information. As before, participants' initial covariation estimates, their confidence in those estimates, and the extent of their reliance on the obtained information will all likely influence whether the estimate will remain unchanged or be revised.

Combining our predictions about cell-combination choices in Selective Collection with the above revision predictions, we expect that Selective Collection will result in some upward and some downward revisions depending on what cell combinations participants choose. Conversely, we expect Non-Selective Collection will always result in downward revisions. Thus, overall, we expect that Non-Selective Collection will result in more normatively directed covariation revisions than will Selective Collection.

> **H2:** Given sample covariation lower than participants' priors, Non-Selective Collection condition participants' revisions of covariation estimates will be more in the normative direction than will revisions by Selective Collection condition participants.

We expect that participants in the Non-Selective Collection condition will use some linear or non-linear combination of the four cells of information to form their covariation estimates (Lipe 1990). Consequently, performing Non-Selective Collection before Selective Collection will sensitize participants to the importance of the four cells of information. The combination rules used in Non-Selective Collection will be available in participants' memories in immediately succeeding tasks and will cause participants to purchase more cells in later Selective Collection so that they can apply these decision rules (Moser 1989; Libby 1985; Tversky and Kahneman 1974, 1973). Therefore, participants who perform Selective Collection soon after Non-Selective Collection, especially if they are additionally sensitized to the importance of all four cells of information before performing Selective Collection, are expected to purchase more information in Selective Collection than participants who have not been sensitized to the importance of all four cells of information relatively recently.

Purchasing more cells of information generally implies that participants are more likely to choose cell combinations that we expect to lead to negative covariation revisions. As discussed earlier, we expect that three out of six two-cell combinations will lead to negative covariation revisions. We also expect that three out of four three-cell combinations will lead to negative revisions. Finally, we expect that choosing the four-cell combination will lead to negative covariation revisions. Consequently, participants will be more likely to produce negative covariation revisions as they choose a larger number of cells. For example, consider participants who follow positive test strategies and move from purchasing two cells to three cells because of the insight gained from sensitization. These participants would choose either cells AB or AC when they purchase two cells; we predict a negative revision if they choose AB and a positive revision if they choose AC. If instead these participants purchase three cells they would choose cells ABC; we predict a negative revision for this cell combination. Consequently, moving from purchasing two cells to three cells increases the likelihood of a negative revision, not because participants process the information any differently, but because the information they choose is more likely to lead to a negative revision. Consequently, we expect the effects hypothesized in H1 (and consequently the bias hypothesized in H2) for Selective Collection to be weaker for participants who are sensitized to the importance of all four cells of information than for participants who are not.

> **H3A:** Selective Collection condition participants will obtain information about more cells when Selective Collection follows Non-Selective Collection and Sensitizing Tasks than when Selective Collection is performed first.
>
> **H3B:** Given sample covariation lower than participants' priors, Selective Collection condition participants' covariation revisions will be more negative when Selective Collection follows Non-Selective Collection and Sensitizing Tasks than when Selective Collection is performed first.

Experimental Design and Method

Since our theory deals with updating priors on an unfamiliar Clue-Condition pair, we use participants who have no definite knowledge or experience with the Clue-Condition pairs used in our experiment, who have enough accounting, auditing, and general business background to understand the task and context. Participants are 114 undergraduate business students from a large public university who volunteered to participate in one of five experimental sessions after being recruited from junior-level accounting classes. The participants are largely accounting (87%) or finance (6%) majors and had taken an average of 5.9 accounting and auditing courses and 1.9 statistics courses. Each participant received nominal course-participation credit and cash payments based on profits earned during the experiment.

General Experimental Task and Procedures

We use a full factorial 2 x 2 design with information-collection strategy (Selective Collection vs. Non-Selective Collection) manipulated within participants and order of information-collection strategy manipulated between participants as independent variables. The dependent variables are the cells of information collected by participants and their covariation-estimate revisions (pre- to post-information collection).

Each participant completed two cases, separated by a sensitization task, and a post-experimental questionnaire containing demographic questions. The cases operationalized either Selective Collection or Non-Selective Collection. For each case, participants were endowed at the start with 2000 “points” as experimental currency (convertible to cash at the end of the experiment at 1000 points = $1.00) for use in purchasing frequency information from the contingency table. The sensitization task involved answering two open-ended questions regarding which cells of a 2 x 2 table (similar to Table 1) managers should emphasize in disclosures to shareholders, and how these disclosures should be regulated by the Securities and Exchange Commission.

For each case, participants performed the sequence of activities outlined in Figure 1. The case activities were designed as analogues to what auditors would do at the planning stage of an audit before collecting client-specific information. First, participants read the case and reported prior estimates of the covariation between the Clue and the Condition. Second, they purchased frequency information (details later in this section) and reported updated estimates of the covariation between the Clue and Condition. Both the prior and updated covariation estimates were elicited on a 101-point scale (0 to 100) anchored by “not at all correlated” and “perfectly correlated.” Third, they issued a hypothetical unqualified or a qualified audit opinion given that the Clue was present, without any company-specific or substantive testing.[8] Lastly, participants were told in each case whether a material misstatement had actually occurred in the company.

****

Insert Figure 1 here

****

For the purposes of the experiment, whether a material misstatement had occurred was determined by a random draw based on the correct probability of a misstatement occurring as calculated from available information. Specifically, the random draw was made from a population that contained 'misstatements' and 'no misstatements' in the ratio 1:4, which was the sample ratio for total misstatements (Table 1 column totals) since this ratio did not depend on the presence or absence of a long-term contract. Participants were informed that a random draw was being made based on "statistically correct principles."

Profits or losses were based on the outcome of the draw, whether participants had issued a qualified or non-qualified opinion, the costs participants had incurred in purchasing information, and their payoff tables.[9] The profits were added to (losses were

[8] The purpose of having participants issue an opinion without client-specific testing in our experiment is to design incentives for them to identify misstatements accurately (because they would be paid or penalized for correct/incorrect identifications). This is not intended to emulate actual practice.

[9] Three different payoff tables were used, randomly assigned between participants separately for Selective Collection and Non-Selective Collection. When participants issued a qualified opinion, some of them earned 50 points and others earned 1000 points when a misstatement was subsequently found present; conversely, they were penalized either 50 points or 1000 points if a qualified opinion had been issued and no misstatement was present (i.e., false positive). When participants issued an unqualified opinion, they earned 1000 points when no misstatement was present, and were penalized 1000 points if a misstatement was present (i.e., false negative).The three different payoff tables were designed with appropriate combinations of these numbers such that they either (a) encouraged the issuance of qualified opinions with high rewards for identifying misstatements and low penalties for false positives; (b) encouraged the issuance of unqualified opinions with low rewards for identifying misstatements and high penalties for false positives; or (c) were neutral between issuing qualified or unqualified opinions. The results did not differ significantly across these incentive structures, so we report results collapsed across incentive conditions.

subtracted from) participants' endowments. In addition to participants being paid their 'points' in cash at the end of the experiment at a rate of 1000 points = $1.00, the person with the highest profit in each session received a $25 cash prize.

Selective Collection and Non-Selective Collection

As mentioned earlier, each participant performed the above procedures twice, once each for the Selective Collection and the Non-Selective Collection cases, separated by a sensitization task. Case order was manipulated. The cases described an audit client in either the construction industry (for Selective Collection) or in the leasing industry (for Non-Selective Collection). The only "real" difference between the Selective Collection and the Non-Selective Collection cases was in the nature of additional information that could be purchased by participants, and how it could be purchased. These differences are detailed in the next few paragraphs. However, care was taken to ensure that there were no other differences, and that the vignettes introducing each case did not reveal any differences due to industry membership. Thus, industry membership of the hypothetical clients served merely as a label to disguise the underlying similarities of the two cases. (A manipulation check confirmed that there was no difference in the priors. See footnote 12.)

In each case, the Clue was the presence or absence of a long-term contract, and the Condition was the presence or absence of material misstatement. In the Selective Collection condition, participants chose the cells of the contingency table for which they wanted to purchase frequency information. Participants were required to buy information for at least one cell, but could buy information for up to all four. Participants were required to buy at least one cell to prevent a "house money" effect wherein they might gamble away their endowment by making random judgments and decisions without obtaining any information (Ackert et al. 2003; Clark 2002; Thaler and Johnston 1990).

Forcing participants to buy at least one cell in such cases reveals the information such participants consider most important.

When a participant requested information, the experimenters supplied the requested information privately. The information provided is given in Table 1. (For example, a participant requesting information on Cells A and B was told that the numbers were 16 and 64, respectively.) Only one purchase request was allowed per participant. That is, they were not allowed to obtain information iteratively. This is consistent with usual auditing practice at the planning stages, when auditors request a 'research' or 'survey' of information from previous audit engagements, obtainable peer data and industry base-rate statistics, etc. Participants who purchased fewer than four cells of information were not told the total number of observations, so they were not able to compute the frequencies in the missing cells.

In the Non-Selective Collection condition, most of the above procedures remained the same, except that participants purchased a sample size of 25, 50, 75, or 100 observations instead of information from particular cells. In this case, participants received the frequencies in all four cells of the frequency table that summed to the sample size they chose, in the same relative frequencies as reported in Table 1.[10] For example, a participant requesting a sample size of 50 got the Cell A, B, C, and D information as 8, 32, 2, and 8, respectively, and could compute all row and column totals. Thus, a participant requesting a sample size of 100 in Non-Selective Collection got exactly the

[10] The positioning and naming of cells in the table of information that participants saw was different than that shown in Table 1. Since our theory broadly predicts that participants would predominantly choose Cell A in Table 1, and since this also happens to be the first cell the reader encounters when reading from left to right, we were concerned that some participants might choose Cell A simply because of their reading habits. Therefore, in the table that the participants saw, we reversed the order of the rows to bias any such habit-based preferences against our hypotheses. However, for expositional ease, and to be consistent with the way these contingency tables have been presented in the literature, our discussion assumes the cell labels of Table 1.

same information as a participant requesting all four cells in Selective Collection. Of course, participants were not informed that the data in the two cases were so linked.

For both Selective and Non-Selective Collection, each incremental level of cell-frequency information cost 200 points; so, for example, purchasing one cell (two cells) in Selective Collection cost 200 points (400 points) and purchasing a sample of 25 (50) observations in Non-Selective Collection also cost 200 points (400 points). The cost of information purchased was subtracted from participants' endowments as an expense.

Results

Hypothesis 1

Recall that H1 predicts that Selective Collection participants will use one of a number of possible positive-test strategies and will obtain Cell A information more often than all other cells, followed by Cells B and C, and then Cell D. Table 2, Panel A presents the proportion of Selective Collection participants who include each cell in their purchase (as participants could choose more than one cell, percentages add to more than 100 percent); Panel B presents one-tailed tests of the predicted comparisons. As predicted, participants choose Cells A, B, or C more often than Cell D as predicted ($p < 0.01$ in each case). However, Cell A is not chosen significantly more frequently than either Cell B or Cell C. This is consistent with our theory that participants' strategies would focus beyond Cell A in the presence of economic incentives.

****

Insert Table 2 here

****

To investigate H1 in greater detail, we look at one-, two- and three-cell strategies separately. Table 3, Panel A lists all of the possible cell combinations that Selective Collection participants could choose and the frequency with which each is chosen. As predicted, few participants (10.4%) choose all four cells of information. Additionally, as expected, single-cell strategies are uncommon. However, Table 3, Panel B shows that when participants do employ a single-cell strategy they choose Cell A more frequently than Cell C ($p = 0.02$) or Cell D ($p = 0.04$), but not more frequently than Cell B ($p = 0.39$). This is consistent with our predictions and contrary to the Cell A-only dominance reported in some research with non-incentivized participants.

****

Insert Table 3 here

****

Among two-cell strategies, we predict that Cells AB and AC will dominate other pairs. Table 3, Panel C lists these comparisons. As expected, participants choose Cells AB (i.e., consistent with a +H test) more frequently than Cells BD ($p < 0.01$) or CD ($p = 0.01$) but not significantly more frequently than the diagonal Cells AD or BC.[11] Additionally, as predicted, participants choose Cells AC (i.e., consistent with a +T test) more frequently than Cells AD ($p = 0.05$), BD ($p < 0.01$), and CD ($p < 0.01$). However, unexpectedly, participants choose Cells AC as frequently as the diagonal BC ($p = 0.50$). The surprising number of participants choosing Cells BC may be consistent with using a reduced form of both +H (Cells AB) and +T (Cells AC) strategies.

Finally, for three-cell strategies, we predict that Cells ABC will dominate other triples. Table 3, Panel D tabulates these comparisons and shows that Cells ABC are chosen more frequently than ACD ($p = 0.01$) and marginally significantly more frequently than ABD ($p = 0.09$) but not more frequently than BCD ($p = 0.15$). The

[11] It can be argued that our incentive scheme promotes the choice of Cells AB (a +H test) as participants are rewarded for identifying the presence/absence of misstatement *given that the Clue was present.* However, the data do not suggest that participants use +H testing more often than +T testing; participants choose Cells AC (14.9%) as often as AB (11.4%), as expected ($p = 0.78$, 2-tailed).

choice of Cells BCD may be evidence of some use of a *negative* test strategy, or the "counterpart" of our predicted strategy of Cells ABC; this is not predicted by our theory.

In sum, the results support the use of positive test strategies which lead to biased population of participants' information sets as predicted. This allows us to investigate the expected covariation-estimation biases predicted in H2.

Hypothesis 2

Recall that H2 predicts that Non-Selective Collection participants' covariation-estimate revisions will be more in the normative direction than Selective Collection participants' revisions. In order to test this hypothesis, we first compute the covariation-estimate revision (i.e., [posterior – prior]) for each participant in each collection condition.[12] Next, for each participant, we compute the difference between the Selective Collection revision and the Non-Selective Collection revision. Panel A of Table 4 presents the above differences and their tests of significance.

****

Insert Table 4 here

****

As hypothesized, the covariation revision in Non-Selective Collection is significantly more negative (i.e., more in the normative direction) than the revision in Selective Collection (difference of 10.13, $p < 0.01$). Parsing this result into its component parts, the results show that participants do not significantly revise their covariation estimates after Selective Collection (49.93 vs. 49.89, $p = 0.99$). With Non-Selective Collection, participants' mean covariation revisions are significantly negative (51.98 to 41.81, $p < 0.01$) as predicted. Further, there is no significant difference in the

[12] Before conducting tests for H2, we first confirm that participants' priors are not significantly different across collection conditions (49.66 vs. 51.79, $p = 0.36$), and that therefore any "label differences" between the cases do not significantly affect participant's priors.

covariation revision in Non-Selective Collection done before versus after the Sensitization task (p=0.41, not tabulated).

The H2 predictions for Selective Collection are based on the expectation that cell selection will affect covariation revision. As expected, the number of cells participants choose does affect the mean covariation revision within Selective Collection (ANOVA: $F = 4.03$, $p < 0.01$, see Panel B of Table 4). Specifically, participants who choose all four cells make significant covariation revisions in the normative direction as expected (mean change = -24.83, $p < 0.01$).

In order to further analyze whether cell choices lead to predicted covariation revisions, we tabulate participants' mean covariation revisions in Selective Collection by cell combination choice. We present the results in Table 5. The "All Participants" data in Table 5 show that the covariation revisions across order are always consistent with our predictions when significant, and are generally directionally consistent when not significant. This provides direct support for our prediction that the Selective Collection covariation estimate biases are driven most by information selection, and not by processing of the information selected.

****

Insert Table 5 about here

****

Hypotheses 3A and 3B

Recall that H3A predicts that participants who perform Selective Collection after Non-Selective Collection and the Sensitization task (hereafter, these two tasks are referred to as simply "Sensitization") will choose a larger number of cells *in Selective Collection* than will participants who perform Selective Collection before Sensitization. Table 6, Panel A shows the number of cells participants choose in Selective Collection by

Sensitization order. As expected, participants choose significantly more cells during Selective Collection when it follows rather than precedes Sensitization (an average of 2.59 vs. 1.96 per participant respectively, $p < 0.01$). Thus H3A is supported.

****

Insert Table 6 about here

****

In order to further investigate the impact of Sensitization on Selective Collection, we tabulate how frequently participants choose each Selective Collection cell combination by Sensitization order. We present these results in Table 7. Recall that we expect that Sensitization prior to Selective Collection will alert participants to the need for information from all four cells in Selective Collection. Examining Table 7 reveals that while the percentage of participants choosing one-cell strategies does not decrease significantly ($p = 0.49$), a smaller percentage of participants choose two-cell strategies (74.5 percent versus 37.3 percent, $p < 0.001$) when they perform Selective Collection after Sensitization. Additionally, a larger percentage of participants choose three-cell strategies (33.9 percent versus 5.4 percent, $p < 0.001$) and the four-cell strategy (16.9 percent versus 3.6 percent, $p = 0.02$) when Selective Collection follows Sensitization.

We examine the overall effect of Sensitization order on cell choice by estimating an ANOVA with cell-choice frequency as the dependent variable and two independent variables: (1) the number of cells chosen (a measured independent variable); and (2) Sensitization order (a manipulated independent variable). Results (not tabulated) show a significant interaction between the number of cells chosen and order ($p=0.014$). Participants purchase more cells when Selective Collection follows Sensitization, supporting our prediction.

****

Insert Table 7 about here

****

H3B predicts that participants' Selective Collection covariation-estimate revisions will be larger and more in the normative direction when Selective Collection follows Sensitization than when Selective Collection precedes Sensitization. To test this, we divide the Selective Collection data based on Sensitization order and compute the revision in Selective Collection for each order. Table 6 presents the revisions for each order and the related tests of significance.

Table 6, Panel B shows that when Selective Collection precedes Sensitization, Selective Collection participants' mean covariation revisions are positive as expected (54.20 to 57.59). Conversely, when Selective Collection follows Sensitization, participants' mean covariation revisions are negative and in the normative direction as expected (46.29 to 42.90). To compare the revisions, we use a contrast based on an ANOVA model with the pre-to-post collection revision in covariation estimate as the dependent variable and Sensitization order as the independent variable. Contrast weights of (0, 1, 0, -1) are used to compare the following cells: prior covariation assessment when Selective Collection precedes Sensitization, posterior covariation assessment when Selective Collection precedes Sensitization, prior covariation assessment when Selective Collection follows Sensitization, and posterior covariation assessment when Selective Collection follows Sensitization. The results indicate that the pattern of means is consistent with our hypothesis and significant ($F = 9.40$, $p = 0.002$). Thus, H3B is also supported.

Lastly, recall from the discussion of H2 that the "All Participants" data in Table 5 show that the covariation revisions across sensitization order are always consistent with our predictions when significant, and are generally directionally consistent when not significant. Table 5 also shows that the same revision patterns hold regardless of sensitization order, indicating that the processing of information does not change depending on whether participants complete Selective Collection before or after sensitization. Therefore, it appears that the improvement in revisions seen when Selective Collection follows sensitization is driven more by which data participants collect than by how they process the data once it is collected. This provides further support for our theory.

Additional Analysis

One possible alternative explanation for our results is that participants who first performed Non-Selective Collection had more money left for the second case (i.e., Selective Collection) for some reason than did participants who performed Selective Collection first. We designed the experiment so that participants would not have a resource constraint in either case, as each participant received an endowment more than sufficient to purchase all available information for each case. Table 8, Panel A, shows that participants who performed Non-Selective Collection first (413.56) did not spend less on information than did participants who performed Selective Collection first (392.73). Moreover, Table 8, Panel B, shows that participants who performed Non-Selective Collection first did not have significantly higher net earnings (1990.68, i.e., endowment less cost of information purchase plus or minus earnings/losses) in the first round than did participants who performed Selective Collection first (1710.00). As Panel B shows, on average participants lost money in all conditions. Overall, this evidence is not consistent with a wealth constraint driving information choices.

****

Insert Table 8 about here

****

Discussion

In this paper, we report the results of an experiment that investigates the impact of two types of information purchases on subsequent covariation revisions in audit planning. We find that when participants buy information selectively, they are likely to buy less than all the necessary information using positive test strategies. This biased information selection leads to biases in covariation revision for certain predictable cell-combinations. When participants buy information non-selectively, their covariation revisions are in the normative direction. Finally, we find that performing Selective Collection after being sensitized to the importance of all four cells of information leads to purchasing more cells of information and more covariation-estimation revisions in the normative direction in Selective Collection than when Selective Collection precedes sensitization.

Our results suggest that covariation-estimation errors in incentivized settings are more likely due to inadequacies in information collection than to information processing errors. Additionally, our results suggest that when auditors choose to do professional research about the covariation of a clue and a condition under budgetary constraints, they are likely to request information about more cells than sometimes documented in previous research, but these strategies may still be inadequate for appropriate covariation assessment. Consequently, auditors are likely to overestimate the covariation between such a weakness and misstatement, which may lead to inefficiencies in designing substantive tests. Our results further suggest that experience with the importance of investigating all four cells of covariation data may be an effective way to reduce the bias in information selection and, consequently, the bias in covariation estimation. This last finding may be especially helpful in designing audit training material.

This paper also extends our understanding of how people estimate covariation in an incentivized setting. While previous research has reported inconclusive evidence on whether people generally *use* all available information in covariation estimation, this paper parses the covariation estimation task into collecting and processing of the information, thus showing that people do not *obtain* all necessary information when collection of information is costly, and this impacts the covariation estimation in predictable ways despite appropriate use of the information available for decision making. The inadequate information obtained then results in biased covariation estimates.

While the specific task we examine is analogous to one that auditors might face in planning audit engagements, similar covariation estimation tasks are common in other accounting-related disciplines whenever estimates need to be made on the potential association between two binary variables that are predicted to covary. For example, investors and other stakeholders may want to estimate the association between the existence of third-world investments and the potential of losses from political turmoil; regulators might be interested in the association between a particular type of debt covenant and the existence of off-balance-sheet financing; tax regulators would like estimates of the association between the existence of a particular tax rule and an increase in tax-compliance; boards of directors would like to assess the association between particular executive compensation plans and the presence of particular schemes of overhead misallocation, etc.

The evaluation of all these situations often include (i) an initial "theoretical" estimate of the association between a clue and a condition for a similar representative firm; (ii) obtaining of costly base-rate empirical frequency information; and, (iii) revision of the initial estimate for a similar representative firm. As such, this paper provides insights into steps (ii) and (iii) of this multi-step process across a wide variety of applicable settings.

Lastly, this paper makes the methodological contribution of utilizing a novel design wherein participants are forced to reveal the information they consider valuable in making covariation assessments. While previous research has generally inferred value attached to information by estimating statistical models of information usage, our design makes obtaining information costly, thereby forcing participants to reveal what information they consider most valuable for the task. Thus, this paper also extends Lipe's (1990) findings and adds to our understanding of the relative frequency of different cell choice strategies in covariation estimation.

As with any experimental research, our study has important limitations in the extent to which it can be generalized to an auditing setting. First, the experimental task is abstract and its generalizability to actual auditing settings is not tested. In particular, the generalizability of our results to iterative information-collection procedures, as in substantive testing, is left to future research.

Second, our Sensitization task was composed of two components: (1) a sensitization treatment which required participants to answer open-ended questions on suggestions for disclosures by CEOs of a hypothetical firm with incentives to boost stock prices, and suggestions about regulation of these disclosures from the SEC's point of view; and (2) performance of Non-Selective Collection before the Selective Collection task of interest. While this two-pronged approach increased the power of Sensitization, we cannot distinguish the impact of performing Non-Selective Collection from the impact of the open-ended questionnaire task. Therefore, while we can safely conclude that sensitization to the importance of all four cells of information in a covariation table increases the number of cells participants purchase, we cannot determine what kinds of sensitization are likely to be most efficient. Separating the effects of the individual debiasers should be the subject of future research.

Lastly, although student participants were purposefully used in this study to conduct a ‘pure’ test of the theory free of professional auditors’ “home-grown” priors (Butler 1986), the extent of this bias at different levels of expertise would be an interesting issue to investigate. It can be argued that professional experience would mitigate these errors even under budget pressure; on the other hand, the biases might increase with unfamiliarity with the Clue-Condition pair. These issues are not addressed in this paper but would be useful extensions of this research.

References

Ackert, L. F., N. Charupat, B. K. Church, and R. Deaves. (2003). An Experimental Examination of the House Money Effect in a Multi-Period Setting. Working Paper 2003-13, Federal Reserve Bank of Atlanta, Atlanta, GA.

Allan, L. G. (1980). A Note on Measurement of Contingency Between Two Binary Variables in Judgment Tasks. *Bulletin of the Psychonomic Society 15*(3), 147-149.

Baker, A. G., R. A. Murphy, and F. Vallée-Tourangeau. (1996). Associative and Normative Models of Causal Induction: Reacting to versus Understanding Cause. In Shanks, D. R., K. J. Holyoak, and D. L. Medin, eds. *The Psychology of Learning and Motivation 34*, 1-45. Academic Press: San Diego.

Bamber, E. M., R. J. Ramsay, and R. M. Tubbs. (1997). An Examination of the Descriptive Validity of the Belief-Adjustment Model and Alternative Attitudes to Evidence in Auditing. *Accounting, Organizations and Society 22* (April/May), 249-268.

Bedard, J. C. and S. F. Biggs. (1991). Pattern recognition, Hypothesis Generation, and Auditor Performance in Analytical Review. *The Accounting Review 66*, 622-642.

Baron, J. (1994). *Thinking and Deciding.* 2nd edition. Cambridge University Press.

Beyth-Marom, M. (1982). Perception of correlation reexamined. *Memory and Cognition 10*, 511-519.

Brown, C. E., M. E. Peecher and I. Solomon. (1999). Auditors' Hypothesis Testing in Diagnostic Inference Tasks. *Journal of Accounting Research 37* (Spring), 1-26.

Butler, S. (1986). Anchoring in the Judgmental Evaluation of Audit Samples. *The Accounting Review 61*, 101-111.

Cheng, P. W. and L. R. Novick. (1997). Covariation in Natural Causal Induction. In *Research on Judgment and Decision Making*, W. M. Goldstein and R. M. Hogarth (eds.), Cambridge, United Kingdom: Cambridge University Press.

Church, B. K. (1991). An Examination of the Effect that Commitment to a Hypothesis has on Auditors' Evaluations of Confirming and Disconfirming Evidence. *Contemporary Accounting Research 7*, 513-524.

Church, B. K. (1990). Auditors' Use of Confirmatory Processes. *Journal of Accounting Literature 9*, 81-112.

Clark, J. (2002). House Money Effects in Public Goods Experiments. *Experimental Economics 5*, 223-231.

Cloyd C. B. and B. C. Spilker (1999). The Influence of Client Preferences on Tax Professionals' Search for Judicial Precedents, Subsequent Judgments and Recommendations. *The Accounting Review 74* (3), 299-322.

Cuccia A. D. and G. A. McGill. (2000). The Role of Decision Strategies in Understanding Professionals' Susceptibility to Judgment Biases. *Journal of Accounting Research 38*(2) Autumn, 419-435.

Einhorn, H. J. (1976). A Synthesis: Accounting and Behavioral Science. *Journal of Accounting Research* (Supplement), 196-206.

Garnham, A and J. Oakhill. (1994). *Thinking and Reasoning.* Oxford: Blackwell.

Gorman, M. E. and M. E. Gorman. (1984). A comparison of disconfirmatory, confirmatory, and a control strategy on Wason's 2-4-6 task. *Quarterly Journal of Experimental Psychology* 36A, 629-648.

Hackenbrack, K. (1992). Implications of Seemingly Irrelevant Evidence in Audit Judgment. *Journal of Accounting Research 30*(1) Spring, 126-136.

Heiman-Hoffman V. B., D. V. Moser, and J. A. Joseph. (1995). The Impact of an Auditor's Initial Hypothesis on Subsequent Performance at Identifying Actual Errors. *Contemporary Accounting Research 11*(2), 763-779.

Hirst, D. E. (1994). Auditors' Sensitivity to Source Reliability. *Journal of Accounting Research 32*(1), 113-126.

Jenkins, H. M. and W. C. Ward. (1965). Judgment of Contingency between Responses and Outcomes. *Psychological Monographs: General and Applied 79* (Supplement 1, Whole No. 594), 1-17.

Johnston, L. (1996). Resisting Change: Information-seeking and Stereotype Change. *European Journal of Social Psychology 26*, 799-825.

Klayman, J. and Y. Ha. (1989). Hypothesis Testing in Rule Discovery: Strategy, Structure and Content. *Journal of Experimental Psychology 15*(4), 596-604.

Klayman, J. and Y. Ha. (1987). Confirmation, Disconfirmation, and Information in Hypothesis Testing. *Psychological Review 94*(2), 211-228.

Klayman, J. and Y. Ha. (1985). Strategy and Structure in Rule Discovery. Paper presented at the *Tenth Research Conference on Subjective Probability, Utility and Decision Making*, Helsinki, Finland, quoted in Klayman and Ha 1987.

Libby, R. (1985). Availability and Generation of Hypotheses in Analytical Review. *Journal of Accounting Research* 23(2) Autumn, 648-667.

Lipe, M. G. (1990). A lens model analysis of covariation research. *Journal of Behavioral Decision Making 3*, 47-59.

Marchant G. (1989). Analogical Reasoning and Hypothesis Generation in Auditing. *The Accounting Review 64*(3), 500-512.

McDaniel, L. S. (1990). The Effects of Time Pressure and Audit Program Structure on Audit Performance. *Journal of Accounting Research, 28*(2), 267-285.

McMillan J. J. and R. A. White. (1993). Auditors' Belief Revisions and Evidence Search: The Effect of Hypothesis Frame, Confirmation Bias, and Professional Skepticism. *The Accounting Review 68*(3), 443-465.

Moser, D. V. (1989). The Effects of Output Interference, Availability, and Accounting Information on Investors' Predictive Judgments. *The Accounting Review 64*(3), 433-448.

Nisbett, R. E., and L. Ross. (1980). *Human inference: Strategies and shortcomings of social judgment.* Englewood Cliffs, NJ: Prentice-Hall.

Pinkley, R. L., T. L. Griffith, and G. B. Northcraft. (1995). "Fixed Pie" a la Mode: Information Availability, Information Processing, and The Negotiation of Suboptimal Agreements. *Organizational Behavior and Human Decision Processes 62*, 101-112.

Plous, S. (1993). *The Psychology of Judgment and Decision Making*. New York: McGraw-Hill Inc.

Shanks, D. R. (1995). *The Psychology of Associative Learning.* Cambridge: Cambridge University Press.

Smedslund, J. (1963). The concept of correlation in adults. *Scandinavian Journal of Psychology 4*, 165-173.

Thaler, R., and E. J. Johnston. (1990). Gambling with the House Money and Trying to Break Even: The Effects of Prior Outcomes on Risky Choice. *Management Science 36*, 643-660.

Tversky, A. and D. Kahneman. (1973). Availability: A Heuristic for Judging Frequency and Probability. *Cognitive Psychology* (September), 207-232.

Tversky A. and D. Kahneman. (1974). Judgment Under Uncertainty: Heuristics and Biases. *Science 185*, 1124-1130.

Vallée-Tourangeau, F., L. Hollingsworth, and R. A. Murphy. (1998). ‘Attentional Bias’ in Correlation Judgments? Smedslund (1963) Revisited. *Scandinavian Journal of Psychology 39*, 221-233.

Waller, W. S. and W. L. Felix, Jr. (1987). Auditors’ covariation judgments. *The Accounting Review 62*, 275-292.

Ward, W. C. and H. M. Jenkins. (1965). The Display of Information and the Judgment of Contingency. *Canadian Journal of Psychology 19*, 231-241.

Figure 1
Summary Time Line of Experimental Procedures

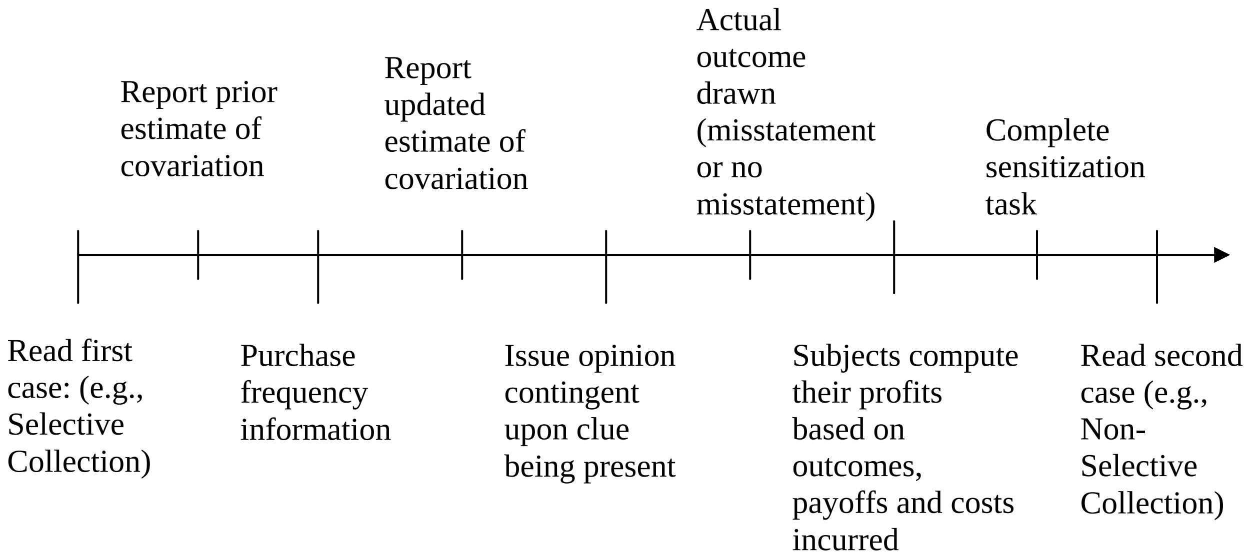

Table 1
The Association Between Clue and Condition When All Available Information is Obtained

	Condition: Material Misstatement	
	Present	Absent
Clue:		
	(Cell A)	(Cell B)
Long Term Contract Present	16	64
	(Cell C)	(Cell D)
Long Term Contract Absent	4	16

The above table presents the frequency information available to participants in the Selective Collection condition who purchase information on all four cells, or to participants in the Non-Selective Collection condition who purchase a sample size of 100 observations.

The most commonly accepted measure of covariation between "Clue" and "Condition" based on the above information is zero:

$\chi^2 = [N(AD-BC)^2]/[(A+B)(C+D)(A+C)(B+D)]=0$

Table 2
Selective Collection Information-Purchase Data and Comparisons

Panel A:
Percentage of participants purchasing information in each cell

	Condition: Material Misstatement	
	Present	Absent
Clue:		
	Cell A	Cell B
Long Term Contract Present	65.80%	63.20%
	Cell C	Cell D
Long Term Contract Absent	60.50%	37.70%

Panel B:
Significance tests on predicted paired differences:

Difference tested	|z-stat|	1-tail p
Cell A vs. Cell B	0.41	0.34
Cell A vs. Cell C	0.82	0.21
Cell A vs. Cell D	4.24	0.00
Cell B vs. Cell D	3.84	0.00
Cell C vs. Cell D	3.44	0.00

Notes:
(1) The percentages do not add to 100% as participants could choose to purchase data for one to four cells. The 15 possible choices are presented in Table 3.
(2) Since the choice of a cell is binary, these tests were based on the normal approximation to the binomial distribution.

Table 3
Selective-Collection Condition Cell-Combination Choices and Comparisons

Panel A: Combinations Chosen

A only	6.1%
B only	5.3%
C only	0.9%
D only	1.8%
AB	11.4%
AC	14.9%
AD	7.9%
BC	14.9%
BD	2.6%
CD	3.5%
ABC	8.8%
ABD	4.4%
ACD	1.8%
BCD	5.3%
ABCD	10.4%
Total	**100.0%**

Panel B: Predicted comparisons of one-cell strategies:

Difference tested	|z-stat|	1-tail p
A only vs. B only	0.29	0.39
A only vs. C only	2.16	0.02
A only vs. D only	1.70	0.04

Panel C: Predicted comparisons of two-cell strategies:

Difference tested	|z-stat|	1-tail p
AB vs. AD	0.89	0.18
AB vs. BC	0.78	0.78
AB vs. BD	2.59	0.00
AB vs. CD	2.27	0.01
AC vs. AD	1.67	0.05
AC vs. BC	0.00	0.50
AC vs. BD	3.28	0.00
AC vs. CD	2.98	0.00

Panel D: Predicted comparisons of three-cell strategies:

Difference tested	|z-stat|	1-tail p
ABC vs. ABD	1.34	0.09
ABC vs. ACD	2.37	0.01
ABC vs. BCD	1.04	0.15

Note: Since the choice of a cell is binary, these tests were based on the normal approximation to the binomial distribution.

Table 4
H2: Covariation-Estimation Revisions

Panel A

	Estimated Covariation Prior	Posterior	Mean Revision	|t-stat|	2-tail p
Selective Collection	49.93	49.89	-0.04	0.01	0.99
Non-Selective Collection	51.98	41.81	-10.17	3.30	0.00
Matched pairs difference of matched-pair differences			10.13 N=111	2.92	0.00

Panel B

One-way ANOVA of Change in Estimate in the Selective Collection Condition:

	F	p	R^2
Effect of number of cells chosen	4.03	0.01	0.10

Number of cells chosen:	1	2	3	4
Mean change in estimate:	-5.47	6.68	-2.09	-24.83***

***Significantly different from zero at a 1% level.

Notes:

(1) The covariation being estimated is between the Clue and Condition with 'sample' information collected as shown in Table 1.

(2) Prior and updated covariation estimates are on a 101-point scale anchored by 0 (not at all correlated) and 100 (perfectly correlated).

(3) All comparisons in Panel A use matched-pairs t-tests. For this, three observations were dropped due to incomplete matched data. Unpaired tests using all observations did not change any conclusions.

Table 5
Analysis of Mean Revisions in Selective Collection by Sensitization Order

		All Participants			Order of Selective Collection: Selective Collection before Sensitization*		Selective Collection after Sensitization*	
Combinations Chosen:	Expected Revision	n	Mean Revision	p†	n	Mean Revision	n	Mean Revision
A only	?	6	-12.0	0.41	4	-10.5	2	-15.0
B only	?	6	-7.2	0.27	2	-6.5	4	-7.5
C only	?	1	10.0	n/a	1	10.0	0	n/a
D only	?	2	11.5	0.08	2	11.5	0	n/a
One cell		15			9		6	
AB	-	13	-19.2	0.02	8	-23.5	5	-12.8
AC	+	17	22.5	0.00	12	14.3	5	42.0
AD	+	9	17.9	0.11	7	14.4	2	30.0
BC	-/?	17	3.9	0.61	9	13.4	8	-6.9
BD	-	3	-6.7	0.73	1	14.0	2	-17.0
CD	+	3	25.0	0.19	3	25.0	0	n/a
Two cell		62			40		22	
ABC	-	10	-10.6	0.12	2	-10.0	8	-10.8
ABD	-	5	-11.0	0.53	0	n/a	5	-11.0
ACD	+	2	46.5	0.40	1	13.0	1	80.0
BCD	-	6	3.3	0.83	0	n/a	6	3.3
Three cell		23			3		20	
ABCD	-	12	-24.8	0.00	2	-42.5	10	-21.3
Total		112			54		58	

Notes:
†2-tailed p-value for tests that the mean revision in the Selective Collection condition across order is different than zero.
*Sensitization includes performance of both the Non-Selective Collection case and the Sensitization Task.

Table 6
Covariation-Estimate Revisions in the *Selective-Collection Condition*
Based on Sensitization

Panel A

Number of Cells Chosen in Selective Collection:	**1**	**2**	**3**	**4**	**Mean**
Selective Collection *before* Sensitization	9	40	3	2	1.96
Selective Collection *after* Sensitization	6	22	20	10	2.59

Panel B

	Estimated Covariation		
	Prior	Posterior	Mean Revision
Selective Collection *precedes* Sensitization	54.20	57.59	3.39 N=54
Selective Collection *follows* Sensitization	46.29	42.90	-3.39 N=58

Notes: (1) Prior and Posterior Covariation estimates are on 101-point scale anchored By 0 (not at all correlated) to 100 (perfectly correlated).
(2) The covariation being estimated is between the Clue and Condition described in Table 1.
(3) Sensitization was accomplished by having participants perform both Non-Selective Collection and the Sensitization Task.

Table 7
Selective-Collection Condition Cell Choices

	Effect of Sensitization:							
	Selective Collection precedes Sensitization*			Selective Collection follows Sensitization*				
Strategy	**n**	**Percent**	**1-tail p†**	**n**	**Percent**	**1-tail p†**	**Difference**	**2-tail p‡**
A only	4	7.3%	0.12	3	5.1%	0.17	(2.2)%	0.63
B only	2	3.6%	0.57	4	6.8%	0.04	3.1%	0.45
C only	1	1.8%	0.80	0	0%	0.91	(1.8)%	0.30
D only	2	3.6%	0.57	0	0%	0.91	(3.6)%	0.14
One-cell	9	16.4%		7	11.9%		(4.5)%	0.49
AB	8	14.5%	0.32	5	8.5%	0.24	(6.0)%	0.31
AC	12	21.8%	0.02	5	8.5%	0.24	(13.3)%	0.05
AD	7	12.7%	0.47	2	3.4%	0.82	(9.3)%	0.06
BC	9	16.3%	0.19	8	13.6%	0.01	(2.8)%	0.67
BD	1	1.8%	0.99	2	3.4%	0.82	1.6%	0.60
CD	4	7.3%	0.88	0	0%	0.98	(7.3)%	0.04
Two-cell	41	74.5%		22	37.3%		(37.2)%	<0.001
ABC	2	3.6%	0.07	8	13.6%	0.08	10.0%	0.06
ABD	0	0%	0.81	5	8.5%	0.50	8.5%	0.03
ACD	1	1.8%	0.39	1	1.7%	0.97	(0.1)%	0.96
BCD	0	0%	0.81	6	10.2%	0.32	10.2%	0.02
Three-cell	3	5.4%		20	33.9%		28.5%	<0.001
ABCD	2	3.6%		10	16.9%		13.3%	0.02
Total	55			59				

†Testing that the percentage is greater than the mean percentage choosing one-cell, two-cell, or three-cell strategies, as appropriate.
‡Testing that the "selective-collection last" percentage is different from the "selective-collection first" percentage.
*Sensitization was accomplished by having participants perform both Non-Selective Collection and the Sensitization Task.

Table 8
Spending and Payments

Panel A: Spending on information collection.

	Case order	
	Selective First	**Non-Selective First**
Selective Collection	392.73[a, b]	511.86[a, c]
Non-Selective Collection	450.91[b]	413.56[c]

Note: Entries with matching superscripts are different from one another at $p < 0.05$.

Panel B: Earnings (endowment minus cost of information purchased plus or minus earnings/losses).

	Case order	
	Selective First	**Non-Selective First**
Selective Collection	1710.00	1809.48
Non-Selective Collection	1710.91	1990.68

Note: None of the above entries differ from each other at $p < 0.05$.