Covariation Judgments and Selective Sampling in Auditing:

An Experimental Study

Ananda R. Ganguly
University of Illinois
Department of Accountancy
1206 S. Sixth St.
Champaign, IL 61820
aganguly@uiuc.edu

Jacqueline S. Hammersley
University of Georgia
Terry College of Business
J.M. Tull School of Accounting
238 Brooks Hall
Athens, GA 30602
jhammers@uga.edu

November 7, 2003

Version 2.2

Preliminary Draft. Please do not quote. We gratefully acknowledge the help of Rajib Doogar, Bob Halperin and our students in the acquisition of participants and in the running of the experiments. We also thank Kathryn Kadous, Susan Krische, and Mark Peecher for comments on earlier drafts of this paper. Research assistance of Kelly Jones is also appreciated.

Covariation Judgments and Selective Sampling in Auditing:

An Experimental Study

Abstract

This paper reports the results of an experiment wherein participants estimated the covariation between two binary variables after obtaining costly sample information. One variable was the cue that auditors might observe, and participants' task was to estimate its covariation with the other variable: material misstatement. Such tasks involve the collection and use of the frequencies in the four cells of a 2 x 2 contingency table representing the intersection of the presence or absence of the cue with the presence or absence of the condition (material misstatement). Previous research has investigated the impact of available complete and costless information on participants' covariation assessments; this study extends on this research by making the information costly to obtain, such that consequent biases in obtained information impact the ultimate covariation assessment. The findings suggest that when given a choice, participants tend to obtain more information about the presence, rather than the absence, of the cue or condition. Such selective and inappropriate sampling seems to bias participants' covariation estimates and their decisions to report material misstatements. However, performing such Selective Sampling *after* a forced non-selective random sampling mitigates the effect of the bias, indicating that forced non-random sampling may potentially be used as a training or debiasing tool prior to audit tasks potentially impacted by this bias. The findings have important implications for audit planning, the design of decision support systems, and in the training of audit personnel. They also extend our understanding of how humans estimate covariation.

Keywords: auditing; sampling; covariation; correlation.

Data Availability: Data will be available from the authors upon request.

I. INTRODUCTION

This paper reports the results of an experiment in which participants acquire and then evaluate information necessary for estimating the covariation between a binary cue and a predicted binary condition. This type of estimation task is common in auditing and normatively requires the use of four pieces of information: the frequencies in the four cells of a 2 x 2 contingency table representing the intersection of the presence or absence of the symptomatic cue with the presence or absence of the reportable condition (Cheng & Novick, 1997). Auditors often must choose what information to collect relevant to a question of whether an account is misstated or how much audit work should be done. For example, when auditing retail clients, auditors must assess the association between the presence of adequate methods of theft-deterring internal control and the risk of significant retail theft to determine, in part, how much inventory work to do. They may base this assessment on evidence from previous audit engagements about the covariation between the presence or absence of particular theft-deterring controls and the absence or presence of significant retail theft.

Previous research has yielded somewhat contradictory results. Some of the original research in the area reported that people often pay much more attention to frequencies that describe the *occurrence or presence* of cues and conditions than to frequencies that describe the *nonoccurrence or absence* of cues and conditions (see Plous 1993, Chapter 15, for a review). Other studies report that participants' judgments appear consistent with statistical models (e.g., Waller & Felix 1987) even though the covariation levels are often misestimated. Most of these previous studies provided participants with information necessary to make normatively correct covariation estimates (see Lipe 1990 for a review). Thus, the objective of these studies was to evaluate whether participants used appropriate available frequency information in making covariation judgments. These studies did not investigate what information participants would *obtain* in order to evaluate the covariation between two variables, especially in circumstances in which obtaining the information can be difficult or costly. In an auditing setting, the task of information evaluation can be subdivided into (1) the obtaining of requisite information based on some search strategy and (2) subsequent cognitive processing of the information obtained. In such settings, the information that auditors obtain can thus affect the availability of information, and this in turn can affect the processing of the information (in this case, the covariation estimation). This paper conducts a test of covariation estimations when costly information must first be obtained and then processed to obtain covariation estimations. The joint impact on covariation estimation of (1) the information search strategy and (2) the information processing is not currently well understood. This paper provides evidence on such joint impact in an auditing setting.

Investigating the joint impact of information search and information processing on covariation estimations is especially important in auditing because obtaining information is difficult and costly. When information is difficult or costly to obtain, auditors are likely to emphasize collection of types of information that they deem most relevant to their task. Consequently, if they evaluate their information needs inappropriately, their covariation estimates are likely to be affected and this, in turn, will likely affect audit efficiency and effectiveness. This is especially important when auditors are faced with new cue-condition pairs (e.g., while auditing an unfamiliar business process) about which they have little prior empirical frequency information.

This study uses laboratory experimentation to investigate this issue. The results suggest that when participants choose the type of frequency information they wish to collect (i.e., a selective or "judgmental" sample), they are more likely to obtain information on those frequencies that are informative about the presence, rather than the absence (i.e., the occurrence, rather than the non-occurrence), of either the cue or the condition, or both.[1] Further, such obtained samples tend to result in exaggerated estimates of covariation between cue and condition in comparison with normatively appropriate estimates.

However, the results also suggest that obtaining a non-selective sample across all cells of the frequency table does not result in these exaggerated (biased) covariation estimates. This is consistent with previous research demonstrating that people use available information appropriately in making covariation estimates when all relevant information is made available (Lipe 1990; Waller & Felix 1987). Moreover, we find that when participants perform selective sampling *after* they first perform non-selective sampling in a similar task, they tend to choose a wider variety of frequency information in the selective sampling, and tend to make less biased covariation estimates. Thus, performing non-selective sampling prior to performing selective sampling seems to have a de-biasing effect on the latter.

The findings have important implications for audit planning, the design of decision aids in auditing, and the training of audit personnel. In addition, this paper extends covariation estimation research by investigating the joint impact of both aspects

[1] Statistical and non-statistical sampling are both acceptable under current professional standards.

of covariation estimation making: the acquiring and the subsequent processing of information. Finally, by making information costly, this study makes the methodological contribution of forcing participants to reveal, through their decisions to purchase information, the information that they really consider valuable in evaluating covariation between variables. (See Lipe 1990 for an explanation of the difficulties of testing cell usage.)

The rest of this paper is organized as follows. Section II presents the theory and develops the hypotheses. Section III describes the methods. The results are analyzed in Section IV, followed by a summary and discussion of the contributions and limitations in Section V.

II. HYPOTHESES DEVELOPMENT

"Covariation assessment" is a term used in psychology to describe the assessment of whether two binary variables are related (Cheng and Novick 1997; Nisbett and Ross 1980). Consider, for example, the information presented in Table 1 relating to the association between the presence and absence of a long-term contract (the cue) with the presence or absence of material misstatement (the condition). The normatively correct assessment 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 there is thus no association between the two variables according to most normatively accepted statistical measures.[2]

****

Insert Table 1 Here

****

Some previous research has suggested that there are various non-normative ways in which people process this type of information (see Plous 1993, Chapter 15, for a review). Some people appear to attend very closely to the present-present cell (Cell A in the table) and assess a high covariation between the variables if the frequency in that cell is high (Smedslund 1963, p.165). Others appear to assess covariation on the basis of whether the sum of the present-present and absent-absent cells (Cells A and D) is larger than the sum of the present-absent and absent-present cells (Cells B and C).

Recent research questions these earlier findings of non-normative judgment, reinterprets prior findings, and reports evidence consistent with accurate covariation assessment (Vallée-Tourangeau et al. 1998; Shanks 1995; Lipe 1990; Waller & Felix 1987). These later studies report that people appear to be sensitive to differences between positive, zero, and negative correlations (Vallée-Tourangeau et al. 1998). Lipe (1990) performed a lens model analysis using data from several prior studies, concluding that participants' aggregate covariation judgments appear highly correlated with

[2] 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 $\chi^2=0$.

Pearson's *phi*, a result that is consistent with the use of all four cells by participants in making their covariation assessments.

However, one factor that may contribute to inaccurate covariation assessments in auditing has been largely overlooked in previous research: namely, auditors often must decide which types of information are relevant to a covariation estimation and then collect that information for use in the estimation process. The acquisition of that information is costly, making it likely that auditors will collect the information that they deem most relevant to the task. Previous research has investigated only whether people appropriately use information already contained in a completed frequency table when making covariation estimates. However, since the information collection strategy may lead to incomplete information in the frequency table, such collection of information may introduce covariation estimation inaccuracies not detected in prior research.

Auditors are generally required to choose both the sample size and the sample cells (i.e., which cells of the frequency table to complete). They are also expected to balance the cost of collecting data with the need for obtaining a sample suitable for their purpose. They are therefore expected to choose information on those cells that seem most relevant to the question of whether a symptom is related to a suspected condition (e.g., whether presence or absence of a theft deterring device is related to absence or presence of significant theft) when they choose sample cells. Previous research related to the 'confirmation bias' has found that peoples' search strategies are often erroneously directed toward tests in which people look for instances they believe should fit a rule, with relatively less attention paid to instances that should not (Klayman & Ha 1989). For example, people seek information that supports their social stereotypes (Johnston 1996) or expectations in negotiations (Pinkley et al. 1995). Extant auditing research has shown that even professional auditors are often confirmation-bias prone (Bamber et al. 1997), and especially so when faced with incentives that reward efficiency, and that they tend to become disconfirmation-bias prone when faced with incentives that reward effectiveness (Brown et al. 1999).

Therefore, if auditors are faced with the task of searching for information in support of covariation between a symptom and an outcome, they are likely to look for information that fits the hypothesis, i.e., for information indicating the frequency with which the symptom and condition occur together. In other words, information about the presence of the symptom and/or the presence of the outcome is expected to be deemed most relevant to the judgment about the level of covariation between the two factors, and participants are likely to search for this type of information in preference to other frequencies. This yields hypothesis H1:

> **H1**: When choosing information on specific combinations of the cue and condition (i.e., in the Selective Sampling condition), participants will more often obtain information on the *occurrence* of a cue or condition than on the non-occurrence of a cue or condition.

The information thus obtained by participants is expected to ultimately impact their covariation assessments. Specifically, participants may have individual prior expectations regarding the covariation between the cue and the condition, and the information obtained by sampling will be used to update these priors. As will be explained in detail in Section III, the information made available to participants through sampling of all four cells of the contingency table in this study was the same as given in Table 1. That is, the statistically correct covariation level determined from an unbiased sample in this study is zero. Therefore, irrespective of participants' priors, the post-sampling covariation assessment can normatively only be less than or equal to the prior.[3] However, the expectation of our theory is that when participants have an option to select information pertaining to particular cells in the frequency table (i.e., Selective Sampling), they will collect more information on the occurrence rather than the non-occurrence of the cue or condition; and their estimates will correspondingly be biased towards a finding of a higher-than-normative level of covariation between the two variables. Therefore, it is expected that the Selective Sampling condition will lead to a non-normative increase in the estimated covariation. On the other hand, when participants do *not* have an option to select information pertaining to particular cells in the frequency table and can only conduct non-directed sampling (i.e., non-Selective Sampling), they will *not* systematically collect more information of one type rather than another, and their estimates will correspondingly *not* be biased in any systematic way. Hence Hypothesis H2:

> **H2:** Given sample covariation lower than participants' priors, participants' pre-sampling to post-sampling increase in perceived covariation will be greater after a Selective Sampling for frequency information in specific cells than after a Non-Selective Sampling.

Hypotheses H1 and H2 are based on the premise of selective search strategies that excessively emphasize collection of information in cells showing presence of cue or condition. This premise is formalized in hypothesis H1, and hypothesis H2 extends on this premise by arguing that such inappropriate collection of information would in turn lead to a covariation estimation bias. However, previous studies have shown that when people have access to all the information in the frequency table, their covariation

[3] Participants either rely on the sample information or they do not. If they rely on the sample information, their revised covariation estimates for the population should move toward zero. Therefore, covariation estimates should normatively decrease (or not change, if the prior was already zero). Alternatively, if participants do not rely on the sample information, they should not revise their priors.

assessments are consistent with the use of decision rules incorporating all relevant bits of information (Lipe 1990; Waller and Felix 1987; Arkes and Harkness 1983; Beyth-Marom 1982; Shaklee and Mims 1982; Shacklee and Tucker 1980). Therefore, when participants perform *Non-Selective* sampling, they are likely to apply one of these decision rules. Consequently, the availability in their memory of these decision rules is likely to be high (Tversky and Kahneman 1973, 1974), and in immediately succeeding similar tasks they are likely to seek information utilized by these decision rules. Therefore, when participants perform Non-Selective sampling immediately prior to Selective Sampling, they will request more cells of information in such subsequent Selective Sampling (as compared to situations in which they had not conducted Non-Selective Sampling in the immediate past). Consequently, the effect hypothesized in hypothesis H2 is expected to be weaker in Selective Sampling performed immediately *after* Non-selective Sampling than otherwise. This yields hypothesis H3:

> **H3:** Given sample covariation lower than participants' priors, when participants perform Selective Sampling *before* performing any Non-Selective sampling, the increase in their covariation estimate in the Selective Sampling condition is greater than the corresponding increase when they perform Selective Sampling *after* performing Non-selective Sampling.

III. EXPERIMENTAL DESIGN AND METHOD

Our experimental method is designed to examine the effect of sample selection type (i.e., Selective or Non-Selective Sampling) and the order of sample selection (i.e., whether Selective Sampling was performed before or after Non-Selective sampling) on decision makers' choice of information and covariation assessments. We use a full factorial 2 x 2 design with sample selection (Selective Sample versus Non-Selective Sample) manipulated within participants and order of sample selection manipulated between participants as independent variables. The primary dependent variables are participants' sample selections and revisions in covariation estimates.

Participants

Participants are 117 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. Participants received nominal course-participation credit and cash payments based on profits earned during the experiment. The participants are largely accounting (87%) or finance (6%) majors. They had taken an average of 5.9 accounting courses and 1.9 statistics courses.

Since the theory we test in this paper involves estimation of covariation between cues and conditions, using professional auditors in our experiments presents special difficulties. Professional auditors are likely to have very strong confounding priors about specific cues and conditions relating to auditing cause and effect relationships that they have developed with their experience in the field. These priors are likely to prove difficult to overcome through experimental manipulation. Therefore, we chose accounting and auditing students for our experiment, taking care to ensure that we selected participants who had the requisite accounting and statistics background necessary to understand and perform the task. Our experiments are thus aimed at testing the theory we postulate without making any assertions about the impact of professional expertise.

Experimental Task and Procedures

Participants completed two cases separated by a distracter task and completed a post-experimental questionnaire containing demographic questions. Each of the two cases operationalized either the Selective Sample or the Non-selective Sample manipulations. The cases described an audit client in either the construction (Selective Sample condition) or leasing (Non-selective sample condition) industries. The order of the cases was counter-balanced. Participants were asked to estimate the association between the existence of long-term contracts and material misstatement (i.e., between the cue and the condition) in order to determine what type of audit opinion to issue when the cue is present. (Extracts from the Instructions for participants are provided in Appendix A.)

For each case, participants read the case, reported prior assessments of the covariation between the cue and the condition, purchased sample information, reported revised assessments of the covariation, and then decided to issue an unqualified or a qualified audit opinion given that the cue was present. Participants were endowed in each case with 2000 "points" and used this experimental currency to purchase one of four levels of cell-frequency information. In the Selective Sample condition, participants chose the cells for which they wanted to purchase frequency information (see Table 1). Participants were required to buy information for at least one cell, but could buy information for up to all four. 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 Sample condition, participants purchased a sample size of 25, 50, 75, or 100 observations. In this condition, participants received the frequencies in all four cells of the frequency table that summed to the sample size they chose.

The frequency information provided in the samples were the same across the two sampling conditions, and the same as the relative frequencies reported in Table 1.[4] Each incremental level of cell-frequency information cost 200 points; so, for example, purchasing one cell (two cells) in the Selective Sample condition cost 200 points (400 points) and purchasing a sample of 25 (50) observations in the Non-selective Sample condition also cost 200 points (400 points). The cost of information purchased was subtracted from participants' endowments.

Immediately after participants made the decision on what information to purchase, their frequency tables were filled in with appropriate frequency information. Participants then provided updated estimates of the covariation between the cue and condition. We elicited participants' prior and updated covariation assessments on a 101-point scale anchored by "not at all correlated" and "perfectly correlated." Next, participants reported their estimated probability of the material misstatement condition occurring, contingent upon observing the cue, and whether they would issue a qualified opinion on observing the cue.

Lastly, participants were told in each case whether a misstatement actually occurred in the company. Whether participants were told that a misstatement had or had not occurred depended actually on a random draw based on the normatively correct probability of a misstatement occurring as calculated from available information. As

[4] Please note from Appendix A that the positioning and naming of cells in the table of information that participants saw was organized a little differently than the organization of cells 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 (shown in Appendix A), we changed the order of the cells to bias any such habit-based preferences against our hypotheses. However, for expositional ease, and to be consistent with the way these covariation tables have been presented in previous research, our discussion assumes the cell labels of Table 1.

indicated earlier in Table 1, the covariation between cue and condition in our experiments was zero. Therefore, the random draw was made from a population which contained misstatements and 'no misstatements' in the base rate ratio of 20:80, i.e., the column totals in Table 1.

Participants earned profits or losses based on the outcome of the draw and their audit opinion choice according to their payoff tables. The profits were added to (losses were subtracted from) their endowments. Participants were paid in cash at the end of the experiment at a rate of 1000 points = $1.00. In addition, the person with the highest profit in each session received a $25 cash prize.

IV. RESULTS

Hypothesis H1

Recall, hypothesis H1 states that when participants choose frequency information using Selective Sampling, they will more often obtain information on the occurrence of a cue or condition than on the non-occurrence of a cue or condition. For this to be true, Cell A (from Table 1) must be chosen more often than any other cell, and Cell B or Cell C must be chosen more often than Cell D.

Table 2, Panel A presents the percentages of participants who chose each cell in the Selective Sampling condition. Panel B contains pairwise comparisons among all the possible pairs. It is clear from Panel B of Table 2 that a significantly larger proportion of participants purchased frequency information on the presence of both cue and condition than on the absence of both cue and condition (Cell A chosen more frequently than Cell D, one-sided $p < 0.01$). Frequency information on the presence of both cue and condition was not chosen significantly more often than frequency information on the presence of one and the absence of the other (Cell A was not chosen significantly more often than either Cell B or Cell C, $p>0.10$ in each case). In addition, frequency information on the presence of either the cue or the condition (but not both) was purchased more frequently than information on the absence of both cue and condition (Cell B and Cell C each chosen more frequently than Cell D, $p<0.01$ in each case). Thus, H1 is supported on the whole by the data because participants chose information about presence of a cue or condition over absence. The only surprise was that Cell A was not chosen significantly more frequently than Cell B or Cell C, but the choices are still directionally consistent with our hypothesis.

****

Insert Table 2 Here

****

One potential confounding factor that may partially explain the results reported above is that participants make two decisions at the same time: the decision on what amount of points (money) to spend on acquiring information, and the decision on which cells to choose in the Selective Sampling condition for a given expenditure in points. Since the desire to minimize expenses is expected to force participants to choose fewer than four cells in the Selective Sampling condition, it can be argued that the results reported above are driven partially by this constraint.[5] To investigate whether the above results are driven *entirely* by the desire to minimize costs, we examine cell choices *for*

[5] Note that even in that case, the fact that participants chose cells conveying information on 'occurrence' of cue or condition over cells conveying information on 'non-occurrence' of cue or condition when trying to minimize costs is evidence in support of hypothesis H1.

each possible level of spending. Panel A of Table 3 presents the percentage of time participants chose a particular combination of cells in the Selective Sampling condition.[6]

****

Insert Table 3 Here

****

<u>Selective Sampling condition choices while spending 200 points</u>: Panel B of Table 3 presents comparison tests between pairs of choices at a spending level of 200 points. Since only one cell can be chosen at this expenditure level, the hypothesis unambiguously predicts that Cell A will be chosen more frequently than any other cell, and Cell B and Cell C will each be chosen more frequently than Cell D. The results show that Cell A is chosen more frequently than Cell C or Cell D (p=0.01 and p=0.03, respectively). The difference between Cell A and Cell B is not significant (p=0.36), although it is directionally consistent with the hypothesis. Additionally, Cell B is chosen significantly more frequently than Cell D as predicted (p=0.05). However, Cell C is not chosen more frequently than Cell D (p=0.73), and the direction of the difference is inconsistent with the hypothesis. Also, Cell B is chosen significantly more frequently than Cell C (p=0.02), a finding not addressed by, nor necessarily inconsistent with, our theory. Except for these last two comparisons, the results all support the hypothesis.

[6] Since we are concerned with the choice of information across different levels of spending, a one-way ANOVA seems at first glance to be the most obvious test. However, a careful evaluation of the design shows that the cost level is linked by design to the extent of information that can be obtained. For example, participants electing to spend 600 points and obtain 3 cells of information will *always* obtain more information about different occurrence frequencies than participants electing to spend only 200 points and obtain information contained in only one cell. Therefore, the cell choices are compared separately for each level of spending.

Selective Sampling condition choices while spending 400 points: Panel C of Table 3 presents *significant* comparisons between possible cell choices at an expenditure level of 400 points. There are 15 possible comparisons between pairs of cells that participants could have purchased with 400 points. In the interest of brevity, only combinations significantly different at conventional levels are presented in Panel C of Table 3. (None of the unreported comparisons were directionally inconsistent with our theory.)

As hypothesized, the results suggest a general preference for combinations involving Cell A (cue and condition both present) and a general avoidance of combinations involving Cell D (cue and condition both absent). Thus AB is preferred to BD and CD; AC is preferred to AD, BD and CD; finally, BC is preferred to AD, BD and CD. In addition, the diagonal AD, containing the key presence-presence and absence-absence cells, is preferred to the other possible combinations involving Cell D, namely BD and CD. All these results are significant and support the hypothesis.

Selective Sampling Condition Choices while spending 600 points: Table 3, Panel D presents the six possible comparisons between the four possible cell choices with a sampling expenditure of 600 points. Hypothesis H1 would predict that ABC would be the most frequently chosen of these and that BCD (the only choice not containing Cell A) would be the least frequently chosen. The results show that ABC is indeed chosen more frequently than ABD or ACD ($p=0.06$ and $p=0.01$ respectively). However, the comparison between ABC and BCD is in the predicted direction but not significant ($p=0.11$, one-tailed). This is surprising given that the only difference between ABC and BCD is the inclusion or exclusion of Cell A and Cell D. Thus, the hypothesis would have predicted the most significant result in this comparison. Also surprising is the finding that BCD is chosen significantly more frequently than ACD (p=0.06) although ACD included Cell A.

Thus, the results from analyses of cell choices at all possible levels of Selective Sampling expenditures also supports H1, except that Cell C is not chosen more frequently than Cell D, and BCD is chosen more frequently than ABC. One possible explanation for these two unexpected results is that some participants may be more aware of the need for information on a variety of cells. This may be especially true for participants who purchase information for a larger number of cells than other participants. Therefore, it is possible that these participants are less prone to the cell-selection biases that result from Selective Sampling, and may indeed be targeting the Cell D information as a crucial piece of evidence which they believe they need.

Hypothesis H2

Recall, hypothesis H2 states that given sample covariation lower than participants' priors, participants' pre-sampling to post-sampling increase in covariation estimation will be greater after a Selective Sampling than after a Non-selective Sampling. The normative level of covariation in the sample was zero. Therefore, whatever participants' covariation estimates prior to sampling, their pre-to-post sampling change in the covariation estimate should be a movement towards zero (unless the participant considered the sample non-representative of the population, in which case there should be no change in estimate).

Panel A of Table 4 presents the mean estimates of covariation made by the participants before and after obtaining the sample information, in both the Selective Sampling and Non-selective Sampling conditions. The expectation was that in the Selective Sampling condition participants would obtain mostly occurrence-related frequencies that would enhance/reinforce their estimate of covariation between the cue and condition. Conversely, the expectation was that Non-Selective sampling would highlight the lack of covariation between the two variables in the sample, and participants would thus revise their covariation estimates downward towards zero. Therefore, H2 predicts that the increase in covariation estimates would be greater in the Selective Sampling condition than in the Non-Selective Sampling condition. The results show that in the Selective Sampling condition, participants' mean estimates of covariation increase from 49.66 to 49.89, although not significantly ($p=0.94$). However, in the Non-Selective Sampling condition, participants mean estimates of covariation *decrease* significantly from 51.79 to 42.06 ($p<0.01$). As hypothesized, the increase in the Selective Sampling condition is thus greater than in the Non-Selective Sampling condition and highly significant (difference of 9.95, $p<0.01$). [A manipulation check confirmed that participants' priors had not been significantly different between the Selective Sampling and Non-selective Sampling conditions ($p=0.36$).]

****

Insert Table 4 Here

****

Since hypothesis H2 is based on underlying cell choices in Selective Sampling of information, it implies that the results are driven in large part by inappropriate cell selection. Consequently, when all four cells are chosen in Selective Sampling, the impact on the covariation assessment should diminish. As expected, the number of cells chosen by participants affected the mean downward revision of their estimates in the Selective Sampling (ANOVA: $F=3.97$, $p<0.01$, see Panel B of Table 4). Specifically, participants who chose all four cells made a downward adjustment in the normative direction(mean change = -24.83, significantly different from zero, $p<0.01$).

In sum, the data show that when participants were offered the opportunity to perform Selective Sampling, they did not update their covariation estimates downwards as they *normatively* should have done. On the other hand, when they were forced to use a Non-Selective sampling, their revision of estimates was significantly in the normative direction. Both these results were consistent with our hypothesis. Lastly, as hypothesized, the downward revision was significantly greater in the Non-Selective Sampling condition than in the Selective Sampling condition. Together, these results strongly support hypothesis H2.

<u>Tests for Hypothesis H3</u>

Recall that H3 predicts that given sample covariation lower than participants' priors, when participants perform Selective Sampling *before* performing Non-Selective Sampling the increase in their covariation estimate in the Selective Sampling condition is greater than when they perform Selective Sampling *after* performing Non-selective Sampling. Recall also that there were two differences between the two cases seen by each participant: (1) in one case they were allowed to choose the cells for which they wanted information (the Selective Sampling condition) and in the other case they could only choose sample sizes (the Non-selective Sampling condition); and (2) the cue in the Selective Sampling case was the presence or absence of a long-term *construction* contract, while the cue in the Non-selective Sampling case was a long-term *lease* contract. The first difference was due to the manipulation of the sampling type, and the second difference was nominal. In all other respects, the two cases were identical. In fact, unknown to the participants ex-ante, the information received by sampling 100 observations in each case (i.e., all four cells in the Selective Sampling condition or a sample size of 100 in the Non-selective Sampling condition) was exactly the same. Therefore, a within-participant comparison of participants' revisions of the covariation estimate was a comparison between isomorphic tasks.

Thus, hypothesis H3 is tested by analyzing the order effect on participants' pre-sampling to post-sampling change in covariation estimates in the Selective Sampling condition. As expected, participants chose significantly more cells in the Selective Sampling condition when Non-selective Sampling came first than when Selective Sampling came first (mean number of cells chosen 2.56 vs. 1.96 respectively, $p<0.01$). Consequently, participants who were in the Selective Sampling condition *after* performing Non-selective Sampling were expected to request information on a wider variety of cells in order to use the decision rules fresh in their memory. Their covariation estimate results are reported in Table 5. Panel A of Table 5 shows that when Selective Sampling was performed before Non-selective Sampling, participants' covariation estimates in the Selective Sampling condition increased from a mean of 54.20 to a mean of 57.59. That is, the Selective Sampling biased their results away from the normative direction (of movement towards zero). Conversely, the revision in the Selective Sampling condition when presented after the Non-selective Sampling condition was negative, moving from a mean of 46.29 to 42.90, consistent with the normative movement towards zero. That is, as hypothesized, performance of Non-selective Sampling before Selective Sampling reduced participants' tendency to make incorrect covariation assessments as compared to the condition in which they had no prior experience of the Non-selective Sampling condition. However, a one-way ANOVA with change in pre- to post-sample covariation estimate as the dependent variable (Table 5, Panel B) showed that this between-participants effect was only marginally significant ($p=0.096$, one-tailed) due to high variance. (Although the pre-sampling covariation estimates in the Selective Sampling condition were marginally different between the two order conditions at the outset ($t=1.64$, $p=0.104$), presumably due to the effect of prior Non-selective Sampling in one order condition, the tests reported relied on the *change* from pre-sampling to post-sampling estimates in each case, and should not have been affected.) Together, these results strongly support hypothesis H3, and its implications are discussed later.

****

Insert Table 5 Here

****

<u>Additional analyses</u>

<u>Degree of difference between Selective and Non-Selective Sampling</u>. Although for ease of exposition no formal hypotheses were presented regarding the comparability

of the Selective Sampling versus Non-selective Sampling differences across different levels of expenditure, the general expectation was that the differences would be larger for smaller sample sizes. This was due to two factors: (1) the expected tendency of relatively lower-cost Selective Sampling conditions to favor selection of Cell A; and (2) the fact that at progressively higher levels of cost the Selective Sampling and Non-selective Sampling conditions led to progressively similar samples. (Note that Selective Sampling and Non-Selective Sampling provided the same sample information at an expenditure of 800 points, the maximum amount of information that could be collected in either condition.) To examine this general qualitative expectation, Table 6 presents details of changes in the dependent variables between the Selective Sampling and Non-selective Sampling across the four possible expenditure levels.[7]

****

Insert Table 6 Here

****

Table 6 shows that at a cost level of 800 points the Selective Sampling condition does not produce more non-normative behavior than the Non-selective Sampling. In each condition, there is a decrease in both the estimated covariation and the reported likelihood of issuing a qualified audit opinion when the cue is present. The decrease in covariation estimates is not significantly different between the groups (–24.83 vs. –23.46, $|t|=0.14$, 2 –tailed $p= 0.88$), and the decrease in reported likelihood of issuing a qualified opinion is significantly greater (and therefore more in the normative direction) in the

[7] Since none of the order effects is statistically significant within these small subsets of the data, the results are reported collapsed across the order condition.

Selective Sampling condition than in the Non-selective Sampling condition (–28.67 vs. –8.9, |t|=1.98, 2-tailed p=0.06). Relatedly, more qualified opinions are issued in the Non-selective Sampling condition at this level than in the Selective Sampling condition (18.2% vs. 0%). Thus, participants' reporting decisions in the Selective Sampling condition are more normative than participants in the Non-selective Sampling condition, in the sense that the former make decisions consistent with zero covariation between cue and condition (while the latter do not). Since these Selective-Sampling-condition participants purchased all the information available, it is likely they were appropriately aware of the importance of all cells in the covariation assessment task, and acted upon the information received.

At a cost level of 600 points, the decrease in covariation estimates is less for the Selective Sampling condition (-2.09) than for the Non-selective Sampling condition (-12.50) as predicted, but the result is not significant (|t|=1.08, 2-tailed p=0.29). Also, the decrease in the reported likelihood of issuing a qualified opinion is now significantly less in the Selective Sampling condition than in the Non-selective Sampling condition (–12.07 vs. –36.64, |t|=2.5, 2-tailed p=0.02). Lastly, the number of qualified audit opinions issued at this cost level is now significantly greater for the Selective Sampling condition than for the Non-selective Sampling condition (35% vs. 9%, |z|=2.07, 2-tailed p= 0.04). All these results are consistent with the hypotheses. Overall, at this level the Non-selective Sampling condition leads to more normative judgments and decisions than the Selective Sampling condition, and most effects are significant, consistent with expectations.

At a cost level of 400 points, the Selective Sampling condition leads to *increases* in covariation estimates away from the normative while the Non-selective Sampling condition still leads to decreases in the normatively correct direction (+6.68 vs. –6.60, |t|= 2.20, 2-tailed p=0.03). Also, the decrease in reported likelihood of issuing qualified opinions is significantly less in the Selective Sampling condition than in the Non-selective Sampling condition (–5.71 vs. -25.51, |t|=3.89, 2-tailed p<0.01). Finally, there is again a significantly greater issuance of qualified opinions in the Selective Sampling condition than in the Non-selective Sampling condition at this cost level (46.8% vs. 12.3%, |z|=4.09, 2-tailed p<0.01). Thus, overall at this level, the Selective Sampling condition leads to qualitatively greater non-normative judgments and decisions than the Non-selective Sampling condition, and all reported effects are now statistically significant.

Lastly, the differences between the Selective Sampling and Non-selective Sampling conditions were expected to be greatest when participants chose very small sample sizes. This expectation was supported qualitatively by the results at a cost level of 200 points (when a single cell is chosen in the Selective Sampling condition or a sample size of 25 is requested in the Non-selective Sampling condition) but the differences between the two conditions were smaller than at the 400-point level. The covariation estimates at this level decrease less in the Selective Sampling condition than in the Non-selective Sampling condition as hypothesized, but not significantly so (–5.47 vs. –10.88, |t|=0.56, 2-tailed p=0.58). The likelihood of issuing a qualified opinion decreases significantly less in the Selective Sampling condition than in the Non-selective Sampling condition as hypothesized (–5.67 vs. –25.60, |t|=2.48, 2-tailed p=0.02). Finally, there is a greater issuance of qualified opinions in the Selective Sampling condition than in the Non-selective Sampling condition at this cost level, but the effect is not significant

(43.7% vs. 32%, $|z|=0.76$, 2-tailed $p=0.45$). One explanation for these somewhat weaker results at a cost level of 200 points may be that participants who acquire such small sample sizes in either condition likely place less reliance on the sample information than at higher levels of expenditure.

Links between covariation estimates and likelihood of issuing a qualified opinion. The covariation estimates and inaccuracies therein are meaningful in an auditing context only if they lead to decision errors in issuing qualified opinions. Therefore, we conducted reasonability tests on the link between participants' covariation estimates (i.e., their judgments), their reported likelihood of issuing qualified opinions contingent upon presence of the cue (i.e., their self-reported reliance on their judgments), and their hypothetical issuances of qualified opinions given presence of the cue (i.e., their decisions).

Throughout the Selective Sampling condition, participants' reported likelihood of issuing a qualified opinion contingent upon the presence of the cue was significantly positively correlated ($r=0.51$, $p<0.01$) with their revised covariation estimates on the relationship between the cue and the condition. In the Non-selective Sampling condition, the corresponding correlation was directionally consistent but not significant ($r=0.094$, $p=0.32$). Further, the *change* in covariation estimate (i.e., the revision in the estimate) was significantly positively correlated with the *change* in the reporting likelihood, both in the Selective Sampling condition ($r=0.45$, $p<0.01$) and in the Non-selective Sampling condition ($r=0.21$, $p=0.03$). It can be concluded, therefore, that participants' revisions of likelihood of issuing qualified opinions were indeed linked to their revisions in covariation estimates.

Finally, we investigated the link between the reported likelihood of issuing qualified opinions and the actual decisions to issue qualified opinions contingent upon presence of the cue. We ran logistic regression models in both Selective Sampling and Non-selective Sampling conditions with the audit report issued as the dependent categorical variable (0 if Unqualified Opinion and 1 if Qualified Opinion) and the reported inclination to issue a qualified opinion when the cue is present as the explanatory variable. The reported inclination measure was significant in both the Selective Sampling condition (odds ratio=1.10, $t=6.06$, $p<0.01$) and the Non-selective Sampling condition (odds ratio=1.08, $t=4.87$, $p<0.01$). Thus, it can be concluded that participants' reported likelihood of issuing qualified opinions significantly predicted their actual decisions to issue hypothetical qualified opinions.[8]

V. Discussion

The experiment reported in this paper presented participants with a covariation estimation task in which they had to pay to get the information they needed to estimate the covariation between two binary variables. These binary variables were of the nature that might be frequently experienced in an auditing context, a context with which the participants were familiar through their academic training. One variable was the cue that auditors might observe, and the task required evaluating how that observed cue was associated with a possible material misstatement. Participants were rewarded for accurately predicting the outcome of a random draw of material misstatement. This reward was netted with participants' costs incurred to obtain the information, resulting in cash profits.

[8] A separate model including order as an explanatory variable indicated that the order effect was non significant in both the Selective Sampling and the Non-selective Sampling conditions and did not change the reported conclusions.

The results showed that when participants tried to minimize costs by selecting the types of frequencies about which they would need to be informed, they mostly chose frequencies informative about the presence of cues or conditions, consequently overestimating the covariation between cue and condition. This led to associated errors in the likelihood of reporting the misstatement (i.e., reported probability of issuing a qualified opinion) contingent upon presence of the cue, and the ultimate frequency with which the condition was reported (i.e., the frequency of issuing qualified opinions).

The results have considerable impact on the field of auditing. When auditors are exposed to a new cue-condition relationship that they need to estimate (e.g. in a new business model or an unfamiliar statutory regime), their estimates may well be affected by whether they conduct a directed non-random sampling of specifically requested information. Relatedly, the finding that performing a non-selective sampling prior to a selective sampling reduces the impact of the biases is noteworthy for de-biasing and decision-support applications. Thus audit programs that include a mandatory random sampling component for assessing covariation between variables are likely to reduce biases possible in immediately subsequent judgmental samplings.

Although student participants were purposefully used in this study to conduct a 'pure' test of the theory, the extent of this bias at different levels of expertise would be an interesting issue to investigate. It can be hypothesized that the similarity of professional auditors' biases to those exhibited by these student participants is likely to increase with unfamiliarity with the cue-condition pair in the past. It may be argued that professional auditors might have less strong biases due to experience with available historical information. However, given that audit failures generally become evident only after business failures, it is debatable whether accumulation of frequency information from past erroneous Selective Sampling strategies likely to bias or de-bias.

Incentives to issue or not issue qualified opinions also have an impact on the extent to which the biases reported in this paper may impact actual issuances of audit opinions. The impact of the general incentive-driven low probability of issuing qualified opinions may moderate the ultimate impact of these covariation assessment errors. However, before studying the moderating impact of incentives, the underlying nature of the biasing effect must be understood. This paper makes such a contribution in the cognitive domain. In addition to the contributions to auditing, this paper also extends our understanding of how people estimate covariation. While previous research had reported contradictory evidence on whether people generally *use* all necessary information in covariation estimation, this paper separates the covariation estimation task into obtaining and processing of the information, thus showing that people do not *obtain* all necessary information, and this often impacts the covariation estimation in predictable ways. The inadequate information availability then impacts usage, resulting in biased covariation estimates.

Lastly, this paper makes the methodological contribution of utilizing a novel design wherein participants are forced to reveal the types of information they consider valuable in making covariation assessments. While previous research had generally inferred value attached to information by estimating statistical models of information usage (cf. Waller and Felix 1987; Lipe 1990), this design makes information expensive to obtain, thereby directly obtaining data about types of information that participants consider valuable for the task.

REFERENCES

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

Arkes, H. R. and A. R. Harkness. 1983. Estimates of contingency between two dichotomous variables. *Journal of Experimental Psychology: General*, 112: 117-135.

Bamber, E. M., Ramsay, R. J. 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, pp. 249-268.

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

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

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.

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

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

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

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

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

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

Shaklee, H. and M. Mims. 1982. Sources of error in judging event covariations: Effects of memory demands. *Journal of Experimental Psychology: Learning, Memory and Cognition,* 8: 208-224

Shaklee, H. and D. Tucker. 1980. A rule analysis of judgments of covariation between events. *Memory and Cognition,* 8: 459-467.

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

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

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

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

Table 1
The association between Cue and Condition when all available information is obtained.

	Condition: Material Misstatement	
	Present	Absent
Cue:		
	(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 a participant in the Selective Sampling condition who has purchased information on all four cells, or to a participant in the Non-selective Sampling condition who has purchased a sample size of 100 observations. Information available to participants who have purchased other samples are presented in Appendix A.

The most commonly accepted measure of covariation between "Cue" 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

The percentage of participants purchasing information on a particular cell in the Selective Sampling condition

Panel A:

Percentage of participants purchasing information in each cell

	Condition: Material Misstatement	
	Present	Absent
Cue:		
	(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 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
Cell B vs. Cell C	0.41	0.34

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 a binary variable, these tests were based on the normal approximation to the binomial distribution.

Table 3
The percentage of time participants chose a particular combination of cells in the Selective Sampling condition

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	10.5%

Panel B: Pairwise comparisons at 200 points spent:

Difference tested	|z-stat|	1-tail p
A only vs. B only	0.36	0.36
A only vs. C only	2.45	0.01
A only vs. D only	1.97	0.03
B only vs. C only	2.14	0.02
B only vs. D only	1.63	0.05
C only vs. D only	0.61	0.73

Panel C: *Significant* pairwise comparisons at 400 points spent:

Difference tested	|z-stat|	1-tail p
AB vs. BD	2.68	0.00
AB vs. CD	2.35	0.01
AC vs. AD	1.76	0.04
AC vs. BD	3.41	0.00
AC vs. CD	3.11	0.00
AD vs. BD	1.82	0.03
AD vs. CD	1.46	0.07
BC vs. AD	1.76	0.04
BC vs. BD	3.41	0.00
BC vs. CD	3.11	0.00

Panel D: Pairwise comparisons at 600 points spent:

Difference tested	|z-stat|	1-tail p
ABC vs. ABD	1.57	0.06
ABC vs. ACD	2.69	0.01
ABC vs. BCD	1.24	0.11
ABD vs. ACD	1.23	0.11
ABD vs. BCD	0.35	0.36
ACD vs. BCD	1.56	0.06

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

Table 4
Changes in pre- to post-sampling perceptions of
covariation between Cue and Condition

Panel A

	Estimated Covariation Pre-sampling	Post-sampling	Mean Change	|t-stat|	2-tail p
Selective Sampling (SS)	49.66	49.89	0.23	0.04	0.94
Non-Selective Sampling (NS)	51.79	42.06	-9.72	3.15	0.00
Matched pairs difference			9.95	2.89	0.00
			N=112		

Panel B
ONE-WAY ANOVA of Change in Estimate in the Selective Sampling condition:

	F	p	R^2
Effect of no. of cells chosen	3.97	0.01	0.10

No. of cells chosen:	1	2	3	4
Mean change in estimate:	-5.14	6.68	-2.09	-24.83***

***Significantly different from zero at a 1% level using Bonferroni adjusted t-tests.
Notes:
(1) Pre-sampling and post-sampling estimates are on a 101-point scale anchored at the two ends by 0 (not at all correlated) and 100 (perfectly correlated).
(2) All comparisons in Panel A use matched-pairs t-tests. For this, 2 observations were dropped due to incomplete data. Unpaired tests using all observations did not change any conclusions, although the difference of 0.23 in the Selective Search condition became a difference of -0.13, p=0.97. The difference of -0.13, *if significant,* would have been inconsistent with the hypothesis.
(3) The covariation being estimated is between the Cue and Condition described in Table 1.

Table 5

Changes in Covariation Estimates in the Selective Sampling Condition
based on whether a Non-Selective Sampling was performed beforehand

Panel A

	Estimated Covariation		
	Pre-sampling	Post-sampling	Mean Change
Selective Sampling (SS) *before* Non-Selective Sampling (NS)	54.20	57.59	3.39 N=54
Selective Sampling (SS) *after* Non-Selective Sampling (NS)	46.29	42.90	-3.40 N=58

Panel B

One-way ANOVA:

Dependent variable: Change in estimated covariation	F	p (one-tail)
Between-Subjects order effect	1.72	0.096

Notes: (1) The two mean changes were individually not significantly different from zero at conventional levels of significance.
(2) Pre-sampling and post-sampling estimates are on 101-point scale scale anchored by 0 (not at all correlated) to 100 (perfectly correlated).
(3) The covariation being estimated is between the Cue and Condition described in Table 1.

Table 6

	N	Estimated Covariation between Cue and Condition: Pre-sampling	Post-sampling	Mean Change	|t-stat|	2-tail p	Reported likelihood of issuing a Qualified Opinion contingent upon presence of cue: Pre-sampling	Post-sampling	Mean Change	|t-stat|	2-tail p	Qualified Opinion issued when cue present: Mean
Participants' Expenditure: 800 points												
Selective Sampling	12	50.75	25.92	-24.83	4.33	0.00	50.00	21.33	-28.67	5.42	0.00	0.0%
Non-Selective Sampling	11	58.27	34.82	-23.46	2.99	0.01	50.00	41.09	-8.91	1.03	0.33	18.2%
Participants' Expenditure: 600 points												
Selective Sampling	23	50.96	48.87	-2.09	0.31	0.76	54.70	42.63	-12.07	1.66	0.11	34.8%
Non-Selective Sampling	22	55.86	43.36	-12.50	1.83	0.08	63.82	27.18	-36.64	5.57	0.00	0.9%
Participants' Expenditure: 400 points												
Selective Sampling	62	48.13	54.81	6.68	1.67	0.10	57.29	51.58	-5.71	1.46	0.15	46.8%
Non-Selective Sampling	57	47.30	40.70	-6.60	1.45	0.15	57.05	31.54	-25.51	8.11	0.00	12.3%
Participants' Expenditure: 200 points												
Selective Sampling	16	56.47	51.00	-5.47	0.92	0.38	54.40	48.73	-5.67	1.07	0.30	43.7%
Non-Selective Sampling	25	56.88	46.00	-10.88	1.64	0.11	60.36	34.76	-25.60	4.79	0.00	32.0%

Note:
(1) All reported pre-sampling and post-sampling estimates are on a 101-point scale. For estimated covariation, the anchors of the scale are 0 (not at all correlated) and 100 (perfectly correlated). For reported likelihood of issuing a qualified opinion, the anchors of the scale are 0 (Certain to issue Unqualified Opinion when cue is present) and 100 (Certain to issue Qualified opinion when cue is absent)
(2) The cue and condition are shown in Table 1.

APPENDIX A
EXTRACTS FROM EXPERIMENTAL INSTRUMENTS

SELECTIVE SAMPLING CONDITION:

{…*Vignette about company, omitted…*}

There are four possibilities:
1. The client has long-term construction contracts and a material misstatement exists.
2. The client does not have long-term construction contracts and a material misstatement exists.
3. The client has long-term construction contracts and a material misstatement does not exist.
4. The client does not have long-term construction contracts and a material misstatement does not exist.

Before you obtain additional information, please provide us with your estimate of… {*Several scales eliciting pre-sampling beliefs, omitted*}

You will now have an opportunity to buy information that will help you improve the three responses you gave on the previous page.

The client database of your audit firm has a large sample of construction contractor clients. Based on these sample data, we can tell you how many times the above four conditions were found. The table below represents four cells corresponding to the "four conditions" explained above. You should decide for which of the four cells of the table below you wish to see the data; you may request data for any one cell, any two cells, any three cells, or for all four cells. (Remember, the information you buy should help you in making judgments of the type you made on the previous page.)

Obtaining information is costly. You will be charged for the information as follows:

A cost of 200 points for data in any 1 cell below
A cost of 400 points for data in any 2 cells below
A cost of 600 points for data in any 3 cells below
A cost of 800 points for data in any 4 cells below

The data can be used to help you decide which audit opinion to issue and the probability that a material misstatement is present.

When you have decided which cell data you would like, please raise your hand and one of the experimenters will provide you with that information by filling in the relevant cell(s) below.

	Material Misstatement Present	No Material Misstatement Present
Client **does not have** long-term construction contracts	A	B
Client **has** long-term construction contracts	C	D

{THESE CELLS ARE NOT IN THE SAME FORMAT AS TABLE 1. See Footnote 4}

Based on the information you received, please provide us with your estimate of… {*Several scales eliciting post-sampling beliefs, omitted*}

Based on the information you received, which type of audit report would you issue if Company I had a long-term construction contract? (circle one) {*Reponse area omitted}*

Outcome: __________________________(to be determined at the end of this part)

NON-SELECTIVE SAMPLING CONDITION:

{*Vignette about company, omitted*}

There are four possibilities:

5. The client has long-term lease contracts and a material misstatement exists.
6. The client does not have long-term lease contracts and a material misstatement exists.
7. The client has long-term lease contracts and a material misstatement does not exist.
8. The client does not have long-term lease contracts and a material misstatement does not exist.

Before you obtain additional information, please provide us with your estimates of … {*Several scales eliciting pre-sampling beliefs, omitted*}

The client database of your audit firm has a large sample of industrial machinery vendor clients. Based on these sample data, we can tell you how many times the above four conditions were found. The table below represents four cells corresponding to the "four conditions" explained above. You should decide on the size of the sample for which you would like to see the data. You may request a sample of 25, 50, 75 or 100 items. (Remember, the information you buy should help you in making judgments of the type you made on the previous page.)

Obtaining information is costly. You will be charged for the sample as follows:

A cost of 200 points for a sample of 25 observations
A cost of 400 points for a sample of 50 observations
A cost of 600 points for a sample of 75 observations
A cost of 800 points for a sample of 100 observations

The data can be used to help you decide which audit opinion to issue and the probability that a misstatement is present.

When you have decided how large of a sample you would like, please raise your hand and one of the experimenters will provide you with that information by filling in the relevant cell(s) below.

	Material Misstatement Present	No Material Misstatement Present
Client **does not have** long-term lease contracts		
Client **has** long-term lease contracts		

{THESE CELLS ARE NOT IN THE SAME FORMAT AS TABLE 1. See Footnote 4}

Based on the information you received, please provide us with your estimate of …{*Several scales eliciting post-sampling beliefs, omitted*}

Based on the information you received, which type of audit report would you issue if Company S had a long-term lease contract? (circle one) {*Response area omitted}*

Outcome: ________________________ (to be determined at the end of this part)