Debiasing Auditor Judgments from the Influence of Information Foraging Behavior

Benjamin P. Commerford
Culverhouse School of Accountancy
The University of Alabama

Richard C. Hatfield
Culverhouse School of Accountancy
The University of Alabama
Tuscaloosa, AL 35487-0220
(205) 348-2901
rhatfiel@cba.ua.edu

Richard Houston
Culverhouse School of Accountancy
The University of Alabama

Curtis Mullis
Culverhouse School of Accountancy
The University of Alabama

May 2014
Debiasing Auditor Judgments from the Influence of Information Foraging Behavior

ABSTRACT

This study examines a potential bias in the collection of audit evidence that originates in information foraging behavior. In an electronic information environment, foraging behavior suggests that auditors selecting items for a sample will select those items that will “cost” the least in terms of time and effort. Specifically, in our setting, where an auditor must haphazardly select sample items across multiple divisions with multiple managers, foraging theory predicts that auditors will avoid those divisions whose managers are slower in responding to information requests. Results demonstrate an interactive effect where auditors collect fewer items from a slow manager, but this reduction in sampling is mitigated when a deviation previously occurred that is associated with the slow manager. By definition, haphazard sampling requires a representative sample, such that the avoidance of a single division introduces bias to the sample. In addition to identifying this bias, we provide evidence of how additional contextual cues in an audit context can change auditors’ interpretations. Since information foraging is self-serving, the bias is eliminated by a simple intervention of having one auditor select sample items to be tested by another auditor. Given the importance of sampling to audit quality, we contribute to the literature by identifying a potentially harmful bias. Perhaps most importantly, we propose a simple and efficient intervention that can eliminate such a bias.
I. INTRODUCTION

We consider how information foraging for audit evidence may bias auditors’ collection of electronic information. Non-statistical sampling, and haphazard sampling in particular, is a very common method for collecting audit evidence (Hall et al. 2002; Christensen et al. 2013) during both controls and substantive testing. These sampling methods play an important role as auditors conduct audits and opine on financial statements (Messier et al. 2014). Haphazard selection requires that the auditor identify a sample that approximates a random collection of information, is representative of the population, and is free from bias or predictability (Christensen et al. 2013). We identify a bias predicted by information foraging theory whereby auditors’ judgments, in an electronic audit environment, will potentially deviate from the requirements of haphazard sampling by not collecting a representative sample. Additionally, we develop a bias-mitigating intervention based on the antecedents to the behavior. Beyond informing auditors and researchers regarding haphazard sampling in a modern audit, our study contributes more broadly to the audit literature as it examines how an auditor’s personal incentives, as well as client behavior, can influence the quality of audit judgments in an electronic audit environment.

Time is a key constraint when completing most audit procedures (e.g., DeZoort and Lord 1997), and auditors are under considerable pressure to increase the efficiency of their work (Willett and Page 1996). Adding to the pressure, auditors must commonly deal with information delays. For example, according to a survey by Bobek et al. (2012), auditors indicated that the late delivery of information from the client was the principal cause of audit difficulties. Prior research also indicates that it is not uncommon for auditors to engage in irregular auditing methods in order to speed up testing. These irregular methods include not testing all items in a reported sample, accepting doubtful audit evidence, and prematurely signing-off on audit
workpapers (Raghunathan 1991; Willett and Page 1996). In addition, extant literature suggests that auditors are prone to decision biases when they evaluate samples non-statistically, which may allow auditors to accept populations that are considered unacceptable based on specified levels of tolerable deviation rate or tolerable misstatement (Elder et al. 2013). These irregular auditing methods and biases are undesirable as nonrepresentative samples and testing of such evidence might compromise overall audit quality (Messier et al. 2014).

We consider a scenario in which, for a given sampling procedure, auditors must obtain information from several client employees in order to complete testing. Information foraging theory (Resnikof 1989), applied to an audit setting, would propose that auditors will try to gather evidence to support financial statement assertions while minimizing effort and time. This suggests that if possible, auditors may avoid collecting evidence from members of client management who appear relatively slow when providing evidence (i.e., auditors are biased towards collecting audit evidence from timely information providers). While selecting a haphazard audit sample, such directed information search is contrary to the fundamental requirement that sample selection should ensure that auditors obtain a sample that is as close as possible to a random, representative sample (i.e., objective, representative of the population, and devoid of bias).

While a delayed response might be attributable to benign reasons (e.g., perhaps the individual is simply slow or busy with other work), psychology research indicates that additional contextual cues help to inform a person’s perception of another’s actions (e.g., Marchand and Vonk 2005). For example, delayed responses can be a cue to deception (DePaulo et al. 2003). In an audit sampling context, the results from prior audit testing may provide additional cues that can influence how the auditor interprets a client’s response. Specifically, audit standards state
that deviations from prescribed controls may cast doubt on the operating effectiveness of such controls (PCAOB AU 350). Hence, if a slow manager is associated with prior deviations, the auditor may interpret the manager’s delayed response as an indication of deception and no longer exhibit the bias predicted by foraging theory. Thus, we consider if the reaction to delayed conveyance of audit evidence is moderated by the occurrence of a deviation in a previously tested sample.

We employ a unique experimental design in which participants select and evaluate samples obtained electronically from a hypothetical client; in other words, we simulate an actual audit setting in which some client employees provide information in a timely manner, while one does not. Results suggest that auditors’ haphazard sample selection is biased in the presence of a client divisional manager who is slow in responding to information requests. Specifically, participants decreased sampling from a slow manager when no deviation occurred in a prior sample. However, if that manager was associated with a deviation in a prior sample, participants did not decrease sampling for the slow manager.

Our expectation for this bias is based on the individual auditor’s incentives. That is, foraging theory predicts that the individual will minimize time and effort when selecting a sample. Based on this incentive, we consider the effectiveness of an intervention that separates the sample selection from the testing outcomes (i.e., potential delays) in an effort to debias the sample selection process. In our setting, when an auditor haphazardly selects a sample for testing to be performed by another auditor (rather than him/herself), the bias described above disappears. Prior research has been unsuccessful in developing a mechanism, such as increasing sample size (e.g., Guy et al. 1998; Hall et al. 2001), to eliminate bias in haphazard sampling. Our
The proposed intervention is a unique, efficient, and easily implementable solution to mitigate sampling bias in comparison to prior research.

Finally, we conduct a survey of practicing staff auditors to gain a better understanding of the sampling context we consider. We find that haphazard sampling is very common, the auditor selecting the sample typically performs the test of that sample, and that client delays in the delivery of information are quite common. We also find that practicing auditors perceive the bias we observe in our experiment to be relatively likely to occur in auditing.

This study informs audit research as well as audit practice. First, we contribute to the sampling literature by identifying a bias that may be quite common in the current audit environment, in which most audit evidence is collected electronically. In so doing, we answer the call for additional research examining how auditors select samples and whether haphazard selection leads to bias (Elder et al. 2013). We also provide evidence concerning how auditors use additional contextual cues to interpret client behavior, such as a delayed response. Finally, this study shows how understanding the cause for biases in audit evidence collection can directly yield an efficient intervention to mitigate the bias. Specifically, separating the judgment (sample selection) from the evaluation of evidence gathered eliminates the bias. Such an intervention provides audit firms with an easy to implement improvement to current audit practice that results in more representative haphazard samples, a commonly cited audit deficiency in PCAOB inspection reports (Church and Schefchik 2012).

The remainder of the paper includes hypothesis development in Section II; the experimental design in Section III; and results following in Section IV. Section V presents a survey of practicing auditors and Section VI discusses our findings and provides implications for audit research.
II. BACKGROUND AND HYPOTHESES DEVELOPMENT

Sampling is an integral part of auditing, and the importance of properly determining and evaluating audit samples is highlighted by recent updates to standards and guides (e.g., AICPA [2011] AU-C 530; AICPA [2012]; PCAOB AU 350). Sampling is necessary to efficiently evaluate characteristics of an account balance or class of transactions without viewing all of the items within that balance or class. Samples are widely used by auditors to test both financial statement balances and the effectiveness of internal controls (Christensen et al. 2013). Auditing standards allow sampling to be performed either statistically or non-statistically (PCAOB AU 350). While statistical sampling allows for the precise measurement of sampling risk and the extrapolation of discovered errors, non-statistical (e.g., haphazard) sampling is the most common method used in practice (Hall et al. 2002). Haphazard sampling improves efficiency and in many cases is preferable to a statistical sample when the expected deviation rate is low.

Haphazard Sampling

Non-statistical sampling is the most commonly employed form of sampling. For example, Hall et al. (2002) find that non-statistical sampling make up 85 percent of sampling applications. Further, Hall et al. (2002) find that haphazard sampling accounts for 87-90 percent of non-statistical sampling. Additionally, this method of sampling is especially common in the context of internal controls testing (Christensen et al. 2013). Haphazard sampling is a non-statistical sampling technique in which the auditor purposefully does not employ a structured approach and attempts to avoid conscious bias and predictability (PCAOB AU 350; Elder et al. 2013). The objective is to obtain a sample that does not selectively avoid or include any particular items in the population, with the ultimate goal of obtaining a sample that is representative of the population (PCAOB AU 350; Christensen et al. 2013).
However, haphazard sampling has been shown to be susceptible to selection bias (Hall et al. 2000). Hall et al.’s results indicate that, when haphazard sampling is used to test inventory, participants are influenced by physical attributes of the inventory items, including the item’s size, color, and location. While Hall et al. examine the effect of physical attributes of inventory samples, the current audit environment involves a high level of reliance on technology for both communications and completing audit testing; in other words, much of the audit is completed electronically (Bierstaker et al. 2001). As noted by Christensen et al. (2013), research has not examined potential biases in sample selection and evaluation in an electronic environment.

**Information Foraging Theory**

Foraging theory is based on evolutionary psychology. Generally, foraging theory asserts that the goal of behaviors, such as foraging for food, is to maximize net energy gain while minimizing costs; in this basic context, collecting the most food while minimizing the amount of time taken and effort expended (e.g., Abarca and Fantino 1982). Foraging theory need not apply only to searching for food, it also applies to other areas regarding the acquisition of required materials where risk and uncertainty are involved (e.g., DiClemente and Hantula 2003). The key currency in foraging is maximizing the acquisition of required resources relative to time spent. When applied to information foraging, this translates to maximizing the value of desired information relative to the cost of obtaining it (Resnikof 1989).

Information foraging has been considered in the context of obtaining information in electronic environments. The link to foraging theory is that information seekers will modify their behavior, or modify the context in which they seek information, if possible, to maximize the value of the information (e.g., Pirolli and Card 1999). For example, on-line shoppers develop a preference for stores (i.e., higher satisfaction ratings) with shorter delays, and are more likely to
make purchases from stores with shorter delays (Rajala and Hantula 2000). Additionally, when time sensitivity is increased, these effects are amplified (DiClemente and Hantula 2003).

Auditors work in an environment in which time constraints are a fundamental contextual characteristic (DeZoort and Lord 1997). Time constraints affect the audit in many ways, including the overall profitability and timeliness of the audit (e.g., meeting client filing deadlines) and individual auditors’ performance evaluations (Wright 1982). On an individual auditor basis, auditors obtain better personal outcomes if they report being at or under budget, as compared to being over the budgeted amount of time (e.g., Shapeero et al. 2003; Agoglia et al. 2014). Research also has shown that these constraints can lead auditors to resort to undesirable actions in an effort to reduce the amount of time spent on audit testing (Raghunathan 1991; Willett and Page 1996). In the context of haphazard sampling, auditors must determine a sample of items to be tested that is representative of the population. Given that auditors choosing the sample also will perform the testing, information foraging theory suggests that they will tend to choose items that will result in the least cost per sampled item (which has the potential to yield a biased, nonrepresentative sample).

Hall et al.’s (2000) findings regarding haphazard sampling are consistent with what foraging theory would predict. They find that auditors demonstrate a bias when choosing samples that involves being more likely to include in their sample physical inventory items that are large, nearby, or brightly colored. Prior audit research has focused on resolving such biases simply by increasing sample sizes (e.g., Guy et al. 1998). However, research suggests that, even when auditors double their sample size, the biases may be reduced, but not fully mitigated (e.g., Hall et al. 2001). As a result, such a solution likely hurts efficiency while not resolving the effectiveness issue. As audits are increasingly based on electronic information and evidence, it is
important to examine how predictions based on information foraging theory would affect audit evidence collection.

Clients are the primary source of audit information and auditors typically must ask clients to provide information (Hirst and Koonce 1996). However, it is very common for auditors to experience information delays. For instance, auditors indicated that the late delivery of information from the client was the leading cause of the audit difficulties they had experienced (Bobek et al. 2012). Auditors encounter varying levels of client responsiveness when they ask for evidence. For example, particular members of client management, when asked by auditors to provide data, may be very busy, may prioritize auditor requests as less important than other responsibilities, or may simply take longer to meet data requests (e.g., because of personal characteristics), resulting in delayed responses to the auditor. In this context, foraging theory suggests that auditors will avoid slow client managers in an effort to minimize their cost (i.e., time spent on the task).

**H1:** When haphazardly collecting a sample, auditors will sample less evidence from a client manager that is associated with delayed responses than a timely manager.

**Prior Deviations as an Additional Contextual Cue**

While information foraging theory suggests that delays in client responses to auditor requests affect auditors’ sampling behavior, the presence of additional contextual cues may moderate this effect. There are many different ways to interpret a late response to an auditor’s request for information from a specific client employee. Possible benign explanations of a slow response include a busy person, a person with more pressing issues and/or unexpected

---

1 We conducted interviews with four audit partners from a regional accounting firm who verified that there is a large amount of variability in the timeliness of the delivery of evidence requested from clients. They suggested that such causes could be varying client workloads, the rank of the auditor making the request (i.e., clients may respond quicker to a partner relative to a staff auditor), and idiosyncrasies. These interviews also are consistent with our survey results in section V.
interruptions, or simply a slow or disorganized person. None of these causal explanations indicate inappropriate behavior on the part of the client employee. Alternatively, psychology research indicates that a delay also can be viewed as a cue indicating potential deception (DePaulo et al. 2003). This cue may be particularly salient in an audit context as auditors are required to exhibit professional skepticism (i.e., Hirst and Koonce 1996; PCAOB 2006).

When there are multiple explanations for observed actions (i.e., delayed responses), individuals must rely on contextual cues to interpret the behavior (e.g., Fein et al. 1990; Hilton et al. 1993; Marchand and Vonk 2005). In an internal control testing environment, one important contextual cue is the presence or absences of deviations. According to audit standards, auditors should evaluate the operating effectiveness of internal controls in terms of deviations from prescribed controls (PCAOB AU 350). The presence of a deviation(s) may not necessitate expanding the sample or increasing future sample sizes if the number of deviations identified in the sample does not exceed the tolerable deviation rate (PCAOB AU 350; Elder et al. 2013). However, audit standards indicate that auditors should consider the nature of deviations and that deviations require more broad consideration when they are attributable to irregularities rather than errors (PCAOB AU 350).

This discussion suggests that a deviation will serve as a contextual cue that may influence the auditor’s interpretation of the delay. If no deviations are found, the auditor has little reason to suspect irregularities from a delayed response (i.e., perhaps the individual is simply slow). However, if an additional contextual cue exists (e.g., a deviation was discovered in this manager’s division) the auditor may change their interpretation of the delayed response. That is, a deviation may increase the likelihood that an auditor interprets a delay as a signal of some type of problem or deception, likely moderating foraging behavior. Thus, we expect an interactive
effect where information foraging will cause a decrease in sampling from the slow manager in
the absence, but not presence, of prior deviations, leading to the following interactive hypothesis:

**H2:** Auditors will sample less evidence from a client manager who is associated with
delayed responses in the absence, but not presence, of a prior deviation.

**An Intervention to Mitigate the Bias**

Recall that prior research focuses primarily on increasing sample size as a method of
debiasing haphazard sample selection (Guy et al. 1998; Hall et al. 2001) and Christensen et al.
(2013) indicate that firms also take this approach. However, it has been shown that this solution
is not entirely effective at eliminating the bias (Hall et al. 2001) and is also likely to reduce audit
efficiency by creating unnecessarily large samples.

Information foraging is based on self-preserving behavior associated with evolutionary
psychology (Rajala and Hantula 2000). Strategies employing foraging behavior are based on the
personal cost of obtaining the resource relative to the value of the resource (Pirolli and Card
1999). Thus, a logical solution to removing the bias brought on by information foraging behavior
is to remove the personal cost (time and convenience) from consideration when an auditor selects
a sample. We consider the simple and efficient debiasing solution of having one member of the
audit team select the sample for another member of the audit team (i.e., independent selection
and testing of samples). That is, when selecting a sample for another member of the audit team,
we expect the auditor’s incentives that yield foraging behavior to be eliminated; specifically, we
expect auditors to select a more representative sample when they are not the individuals that will
collect the evidence and perform the tests, leading to the following hypothesis:

**H3:** The predicted undersampling resulting from delayed response will be eliminated
when selecting a sample for another auditor.
III. EXPERIMENT

Participants

Participants for which we report results include 77 Masters of Accountancy students enrolled in a graduate auditing class at a large state university in the Southeast. Given the logistical complexities of this experiment, practicing auditors were not a feasible participant pool. Participants averaged approximately three months of internship experience over one busy season. Since prior literature has established that audit sampling is commonly performed by entry-level staff whose training in bias avoidance is concentrated in their undergraduate accounting studies (Hall et al. 2001), we consider graduate students (most with internship experience) to be an appropriate proxy for staff-level auditors. Forty-eight participants (55 percent) were male and 39 were female.

Experimental Audit Case

After receiving general instructions from a research assistant, participants were taken to separate offices where they received case materials via computer. Participants were asked to assume the role of a staff-level auditor responsible for testing internal controls on a hypothetical audit engagement. Specifically, they were to test the operating effectiveness of a manual control over cash disbursements that required dual signatures on all check disbursements. Participants completed three tasks during the experiment.

2 A total of 99 students participated in the experiment as a part of their course credit. Twelve participants were removed due to either incomplete responses (3), tasks performed in the wrong order (4), or participant error resulting in a significant time delay (5). Additionally, 10 participants failed at least one manipulation check, resulting in the final sample size of 77. Removed participants were not concentrated in any one cell.
3 All participants had completed an undergraduate course in auditing and were currently enrolled in a graduate-level course in auditing. Participants also demonstrated an understanding of the concept of haphazard sampling, with 83% correctly identifying the appropriate definition.
4 Additionally, we interviewed four partners from a regional accounting firm who all indicated that staff-level auditors and interns were an appropriate auditor rank to perform such audit tasks (i.e., selecting internal control testing samples).
5 Because of logistical demands, the experiment was spread over three days in a manner that allowed up to four individuals to complete the experiment at any given time.
In Task 1, participants received a pre-selected sample of checks from four operating divisions of the audit client for the second quarter of 2012. Participants then sent an email to the managers in charge of each division requesting the necessary check copies relating to the pre-selected sample. After receiving copies of the requested checks from division managers (where timeliness was manipulated), participants completed the internal control testing and emailed the completed audit workpaper to their hypothetical audit supervisor, played by one of the researchers (located in a remote room). Participants were told that the time budget for Task 1 was 15 minutes, and they were provided with a timer on their computer screen in order to monitor their budget.

For Task 2, the audit supervisor sent a new Excel spreadsheet containing the client’s third quarter check disbursements and instructed the participant to haphazardly select a sample of 20 checks in order to complete the third quarter internal control testing. After participants sent the third quarter check request to the client division managers, the audit supervisor sent a third and final task to the participants with the following instructions. “I know that you haven’t completed the third quarter testing yet, but I need you to go ahead and select a sample of fourth quarter disbursements as well. We would like to give another staff on the job an opportunity to perform some internal control testing. So, the fourth quarter sample you select will be tested by the other staff, not you. You should select the sample haphazardly.” After selecting the fourth quarter

---

6 Participants were told that the sample was selected using a haphazard method. The materials included a short definition of haphazard sampling consistent with Elder et al. (2013). Additionally, instructions indicated that, if more than one deviation (i.e., two deviations) was identified during testing, the sample would need to be expanded.

7 Pilot testing of the instrument on over 50 senior level undergraduate auditing students, multiple accounting Ph.D. students, and accounting faculty members indicated that 15 minutes was the appropriate amount of time needed to complete the first task.

8 The check detail was created with a purpose of keeping the risk and materiality constant among all four divisions. We randomly populated 75 checks with amounts between $10,000 and $20,000. We then used the same 75 check amounts for each operating division. For each division we altered the check numbers and check payees, and randomized the order between divisions. The same detail was provided for second, third, and fourth quarters with the order randomized and the check numbers and dates altered.
sample, participants were told that they had finished the case portion of the experiment and were directed to open an envelope that contained a post-experimental questionnaire regarding their experience with the tasks and audit client, along with manipulation check and demographic measures.

**Independent Variables**

To test the hypotheses, we use a 2 x 2 x 2 ANOVA design where the first two factors, *timeliness* and *deviation* are manipulated between-subjects at two levels. The third factor, *testing responsibility*, is manipulated within-subjects at two levels. We manipulated timeliness by randomly assigning participants into one of two groups (fast versus slow). In these experimental conditions, participants interacted with four client managers through email in order to complete second quarter internal control testing. In the fast condition, all four client managers responded to participants’ sample selection within two minutes of the email request. In the slow condition, three of the managers responded in exactly the same manner as in the fast condition. The fourth manager responded in three separate emails, each partially fulfilling the sample request. The three emails together fulfilled the request and were sent approximately 2, 10, and 14 minutes after the participant sent the initial email.

To manipulate the second independent variable, deviation, participants in both timeliness groups received either zero or one seeded deviations in the second quarter testing (no deviation versus deviation). Recall that participants were informed that one deviation was tolerable, but that if two or more deviations were identified the sample would need to be expanded. In all cases in which a deviation was present, the deviation came from the same manager who was manipulated as either slow or fast. In order to protect against ordering effects as suggested by
prior audit sampling studies (Hall et al. 2010), we varied the order of the four divisions in the disbursements spreadsheet.⁹

We manipulated testing responsibility by telling participants that either they would be testing the sample selection or someone else on the hypothetical audit team would be (i.e., self versus other). Recall that participants selected a sample for two sequential accounting quarters (quarters 3 and 4). After selecting the sample for the third quarter, but before receiving any client documentation pertaining to testing, participants were instructed to select a sample of fourth quarter transactions that would be tested by another member of the audit team.

**Dependent Variables**

The appropriateness of haphazard selection as a nonstatistical audit sampling technique depends on whether this method can reliably provide samples that are representative of some population characteristic (Elder et al. 2013). Thus, the dependent variables measured in this study pertain to whether or not the third and fourth quarter samples selected by participants were representative of the population of check disbursements. We calculate the number of sample selections for each division as a percentage of the total sample size (20) as a measure of representativeness. Specifically, we isolate the division in which the manipulations occurred when measuring our dependent variables. Because the check amounts and payees for each operating division and each quarter were exactly the same (order was randomized), the normative selection percentage (on average) from each division equals 25 percent (five checks).

**IV. RESULTS**

H1 predicts that auditors will sample less from client managers that are previously associated with delayed responses to auditor requests for sample items. We test this hypothesis

---

⁹ We tested for order effects by including a control variable in the model. There were no significant differences in any of the results when including this variable.
by examining the main effect of timeliness, which is not significant (p-value = 0.49; see Table 1, Panel B). Specifically, participants selected 5.08 items, on average, from the division of interest when there was no delay in prior sample requests. Participants selected 4.98 items on average when the division of interest was associated with a delayed response in the prior quarter testing. However, examination of Figure 1 indicates that the interaction predicted in H2 requires an examination of the simple effects to better understand the main effect of timeliness.

Recall that the effect of timeliness is predicted to be strongest when no deviations are discovered in prior testing. H2 predicts an interaction effect where auditors will sample less from a slow manager when no prior deviations exist and that this undersampling will be mitigated when a prior deviation occurred. This interaction is significant (p-value = 0.01; see Table 1, Panel B) and examination of Figure 1 indicates that the form of the interaction is consistent with our prediction. The simple effect of delay when no deviation occurred is consistent with the prediction in H2. Specifically, participants selected 5.10 items when the client manager was timely in response in the prior quarter and 4.39 items when the client manager was slow in response in the prior quarter (see Panel A of Table 1). This difference is significant (directional p-value = 0.03; see Table 1, Panel C). Interestingly, when a deviation occurred, participants appeared to increase their sampling of the slow manager (means of 5.77 and 5.06 respectively; p-value = 0.11; Table 1, Panel C), although the difference is not significant. These results support the interactive effect predicted in H2, and are consistent with information foraging behavior occurring when no additional contextual cues were present.

H3 examines a potential intervention to reduce the bias in haphazard sampling. We test this hypothesis by considering whether the reduced sampling predicted in H1 is mitigated when the sample is chosen for another auditor to test, thereby removing the biases that may be based
on personal incentives. Figure 1 illustrates that the undersampling brought on by a slow manager without prior deviations is eliminated when the sample is selected for another auditor. Note that the cell means in Table 2 indicate that the simple effect of delayed response is eliminated when the sample is chosen for another auditor.¹⁰ We also find that the marginal increase in sampling brought on by the occurrence of a delay and deviation is eliminated when the sample is drawn for another auditor. Consistent with these findings, the 3-way interaction is significant (p-value = 0.01; Table 3). To interpret this 3-way interaction we look closer at the cell means. The pattern of means demonstrates that the 2-way interaction predicted in H2 does not occur when selecting a sample for another auditor (p-value = 0.73, see Table 2, Panel B). In total, our results suggest that auditors choose a more representative sample when selecting the sample for another auditor.

V. AUDITOR SURVEY

Participants

Given our use of students having internship experience as participants, we conducted an electronic survey of practicing auditors to provide additional context to our experimental results. The participant pool was 55.5 percent female and had an average of 3.4 years of auditing experience. Of the participants, seven were staff-level, eight were senior-level, and three were manager-level. With the exception of one participant working for a large regional firm, all participants worked for various Big 4 firms.

Survey Results

We asked participants several questions to validate our assumptions and provide additional information about sampling practices. It is important to note that these responses represent the perceptions of auditors. Results reported in Table 4, Panel A indicate that

¹⁰ Additionally, we examined whether cell means were different than the normative expectation of 5 selections from each manager/division, noting that no significant differences existed, suggesting a representative sample.
participants identify haphazard sampling as the most employed sampling technique during internal control testing (44.7 percent haphazard vs 40.3 percent random) and the second most used sampling technique for substantive testing (28.6 percent haphazard vs. 58.3 percent monetary-unit). Over half of the participants (55.5 percent) indicate that the person choosing the sample is typically the one who tests the sample, 33.3 percent indicate that this is “sometimes” the case, while only 11.1 percent say this rarely happens (0 percent saying it never happens). Participants also suggest that it is unlikely that reviewers will comment on whether their selected sample is representative (66.7 percent indicating it is somewhat unlikely to very unlikely), with only two respondents stating such a review comment would be likely.

Results reported in Table 4, Panel B indicate that 94.4 percent of respondents state that client delays when responding to information requests are common or very common. We also provided two scenarios for these respondents to consider how another auditor may act when making sample selections. The first scenario was one similar to the experiment, in which we asked how likely an auditor would be to avoid selecting items from a slow manager. The mean response on our 6 point scale (from very unlikely to very likely) was 3.89 with 66.7 percent of respondents beyond the midpoint of the scale indicating significant agreement with this possibility. The second scenario considers an auditor selecting a random sample with a random number generator. The auditor’s random sample has selected some items that may be difficult to test (e.g., slow manager or complex item). We ask the likelihood of whether the auditor may select another random sample. This scenario was provided in multiple anecdotal examples during the development of this study by auditors and former auditors. The mean for this question is 2.56. While this mean is below the midpoint, 22.2 percent of our respondents believed this action to be at least somewhat likely and only 27.7 percent found it to be very unlikely, suggesting that
a large portion of auditors believe that rerunning random sample selections occurs in practice. This latter scenario provides additional evidence to the idea that simply having the auditor select a “random” sample may not solve the issue of information foraging, but that the underlying incentives should be considered.

VI. DISCUSSION

The majority of audit procedures on which audit evidence is obtained and conclusions are developed are founded on samples developed by the audit team. Of these samples, most are non-statistical and haphazardly selected. This study demonstrates, based on information foraging theory, that auditors will select haphazard samples with a bias that will reduce delays in their own work by undersampling slow managers. However, this effect is mitigated when additional contextual cues (i.e., a deviation identified during prior period testing that is associated with the slow manager) are present that may trigger suspicion. Finally, a simple intervention in the development of sample selection and evaluation procedures eliminates this bias. Since this bias is based on a personal incentive (i.e., delays in completing testing), it can be eliminated when samples are selected for another auditor to perform the test work.

Sampling bias has been a concern for several years (e.g., Shanteau 1989). Audit regulations have long allowed non-statistical sampling (e.g., AICPA [2012] Sampling Guide; PCAOB AU 350), and evidence suggests that auditors rely primarily on such sampling, principally employing haphazard sampling (e.g., Hall et al., 2002). This study contributes to the understanding of the potential personal incentives that can yield sampling bias in an electronic audit evidence environment. This sampling bias results in nonrepresentative samples that may compromise the overall quality of the audit. Further, the avoidance of requesting audit evidence from managers who exhibit delayed responses is consistent with other undesirable irregular
auditing methods used by auditors to speed up testing (Willett and Page 1996; Donnelly et al. 2003). Given the nature of the antecedents to these biases and its importance to audit quality, we contribute to the literature by exploring a simple, effective, and efficient method for eliminating these biases. Prior focus by researchers and audit firms has been on increasing sample size (e.g., Guy et al. 1998; Hall et al. 2001; Christensen et al. 2013), which is not be completely effective in eliminating bias and therefore is likely to reduce audit efficiency. Our study contributes to this line of research by introducing a simple intervention in which auditors select samples for other auditors to test, thereby removing the precursor to the bias.

This study is subject to certain limitations. First, given the time and location requirements of this study, participants were graduate-level students rather than staff auditors. Additional experience could influence our findings. However, many of the student participants had completed “busy season” internships, in which they had experience similar to the staff for whom they proxy.\(^{11}\) Further, while incentives in practice cannot be completely recreated, we attempted to develop an incentive structure parallel to that experienced in practice. For example, participants were given a time budget and were held accountable for their work through a debriefing with their audit class instructor that simulated a review. It is important to note that all participants selected their own sample prior to selecting a sample for another auditor. Therefore, an order effect may influence our results relating to the intervention. However, this design choice was made due to logistical and participant limitations. Additionally, we are unaware of any theoretical reasons why order would affect our hypotheses and cause participants in only one of the four experimental conditions to undersample in initial sample selections, but not subsequent sample selections. Finally, future research could consider how information foraging behavior affects the audit engagement in areas other than sampling. While our study focuses on how

\(^{11}\) Results did not differ between participants who did or did not have internship experience.
information foraging behavior can negatively impact sampling, it is possible that, in other areas of the audit, this type of behavior can be influential (e.g., fraud interviews, walk-throughs, follow-up questions, etc.). Such research would further our understanding of how information foraging impacts auditors and audit quality, such as the biases described in this study.
REFERENCES


### TABLE 1

Panel A: Mean (Std. Dev.) Sample Selections from Client Manager of Interest for Self to Perform Testing

<table>
<thead>
<tr>
<th>Prior Deviation</th>
<th>Low Timeliness</th>
<th>High Timeliness</th>
<th>Collapsed Across Timeliness</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean (Std. Dev.)</td>
<td>Mean (Std. Dev.)</td>
<td>Mean (Std. Dev.)</td>
</tr>
<tr>
<td>Prior Deviation</td>
<td>5.77 (1.48)</td>
<td>5.06 (0.99)</td>
<td>5.40 (1.29)</td>
</tr>
<tr>
<td></td>
<td>n = 17</td>
<td>n = 18</td>
<td>n = 35</td>
</tr>
<tr>
<td>No Prior Deviations</td>
<td>4.39 (1.43)</td>
<td>5.10 (0.81)</td>
<td>4.71 (1.24)</td>
</tr>
<tr>
<td></td>
<td>n = 23</td>
<td>n = 19</td>
<td>n = 42</td>
</tr>
<tr>
<td>Collapsed Across Deviation</td>
<td>4.98 (1.59)</td>
<td>5.08 (0.89)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>n = 40</td>
<td>n = 37</td>
<td></td>
</tr>
</tbody>
</table>

Panel B: ANOVA Results

<table>
<thead>
<tr>
<th></th>
<th>F</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Timeliness</td>
<td>0.00</td>
<td>0.49</td>
</tr>
<tr>
<td>Deviation</td>
<td>5.56</td>
<td>0.01</td>
</tr>
<tr>
<td>Timeliness*Deviation</td>
<td>6.43</td>
<td>0.01</td>
</tr>
</tbody>
</table>

Panel C: Simple Effects

<table>
<thead>
<tr>
<th></th>
<th>Low Timeliness Mean (Std. Dev.)</th>
<th>High Timeliness Mean (Std. Dev.)</th>
<th>F (p-value)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Timeliness Across No Prior Deviation</td>
<td>4.39 (1.43)</td>
<td>5.10 (0.81)</td>
<td>3.70 (0.03)</td>
</tr>
<tr>
<td>Timeliness Across Prior Deviation</td>
<td>5.77 (1.48)</td>
<td>5.06 (0.99)</td>
<td>2.80 (0.11)</td>
</tr>
</tbody>
</table>

Table 1 provides descriptive statistics, ANOVA results, and simple effects of participants' sample selections from manager 3 (the manager of interest) when instructed to select a sample that they would test. Consistent with H1, all reported p-values are one-tailed.
<table>
<thead>
<tr>
<th></th>
<th>Low Timeliness</th>
<th>High Timeliness</th>
<th>Collapsed Across Timeliness</th>
</tr>
</thead>
<tbody>
<tr>
<td>Prior Deviation</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>4.88 (0.86)</td>
<td>5.06 (0.97)</td>
<td>4.91 (0.90)</td>
</tr>
<tr>
<td></td>
<td>n = 17</td>
<td>n = 18</td>
<td>n = 35</td>
</tr>
<tr>
<td>No Prior Deviations</td>
<td>5.13 (0.92)</td>
<td>5.17 (0.62)</td>
<td>5.14 (0.79)</td>
</tr>
<tr>
<td></td>
<td>n = 23</td>
<td>n = 19</td>
<td>n = 42</td>
</tr>
<tr>
<td>Collapsed Across Deviation</td>
<td>5.03 (0.89)</td>
<td>5.11 (0.80)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>n = 40</td>
<td>n = 35</td>
<td></td>
</tr>
</tbody>
</table>

**Panel B: ANOVA Results**

<table>
<thead>
<tr>
<th></th>
<th>F</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Timeliness</td>
<td>0.29</td>
<td>0.59</td>
</tr>
<tr>
<td>Deviation</td>
<td>0.80</td>
<td>0.37</td>
</tr>
<tr>
<td>Timeliness*Deviation</td>
<td>0.12</td>
<td>0.73</td>
</tr>
</tbody>
</table>

**Panel C: Simple Effects**

<table>
<thead>
<tr>
<th></th>
<th>Low Timeliness</th>
<th>High Timeliness</th>
<th>F (p-value)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean (Std. Dev.)</td>
<td>Mean (Std. Dev.)</td>
<td>(p-value)</td>
</tr>
<tr>
<td>Timeliness Across Prior Deviation</td>
<td>4.88 (0.86)</td>
<td>5.06 (0.97)</td>
<td>0.32 (0.58)</td>
</tr>
<tr>
<td>Timeliness Across No Prior Deviation</td>
<td>5.13 (0.92)</td>
<td>5.17 (0.62)</td>
<td>0.02 (0.89)</td>
</tr>
</tbody>
</table>

Table 2 provides descriptive statistics, and ANOVA results of participants' sample selections from manager 3 (the manager of interest) when instructed to select a sample that someone else would test. Consistent with H3, all reported p-values are two-tailed.
### TABLE 3

ANOVA Results for Full Model

<table>
<thead>
<tr>
<th></th>
<th>F</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Timeliness</td>
<td>0.05</td>
<td>0.40</td>
</tr>
<tr>
<td>Deviation</td>
<td>1.91</td>
<td>0.09</td>
</tr>
<tr>
<td>Selection</td>
<td>0.00</td>
<td>0.48</td>
</tr>
<tr>
<td>Timeliness*Deviation</td>
<td>2.78</td>
<td>0.05</td>
</tr>
<tr>
<td>Timeliness*Selection</td>
<td>0.15</td>
<td>0.35</td>
</tr>
<tr>
<td>Deviation*Selection</td>
<td>6.93</td>
<td>&lt;0.01</td>
</tr>
<tr>
<td>Timeliness<em>Deviation</em>Selection</td>
<td>5.27</td>
<td>0.01</td>
</tr>
</tbody>
</table>

Table 3 provides results for a mixed ANOVA model with Timeliness and Deviation as Between-Subjects factors and Selection as a Repeated Measure Within-Subjects factor. All reported p-values are one-tailed.
TABLE 4
Results of Survey

Panel A: Audit Sampling Experiences (n=18)

<table>
<thead>
<tr>
<th>(1)</th>
<th>Who commonly performs sample selections?</th>
<th>(2)</th>
<th>Participants' specific audit experience:</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>n (% of 18)</td>
<td>Mean (Std. Dev)</td>
<td>Scale Endpoints:</td>
</tr>
<tr>
<td>Staff</td>
<td>10 (56%)</td>
<td>6.6 (2.38)</td>
<td>1 = None</td>
</tr>
<tr>
<td>Senior</td>
<td>13 (72%)</td>
<td>5.9 (2.46)</td>
<td>9 = Extensive</td>
</tr>
<tr>
<td>Manager</td>
<td>1 (6%)</td>
<td>5.8 (2.43)</td>
<td></td>
</tr>
</tbody>
</table>

(3) In your experience, how often does the person selecting the sample also complete the testing of that sample?

Typically 10 (56%)
Sometimes 6 (33%)
Rarely 2 (11%)

(4) For each type of audit testing, how often do you use the following sampling methods? a

<table>
<thead>
<tr>
<th></th>
<th>Haphazard sampling</th>
<th>Random sampling</th>
<th>Statistical sampling (i.e., monetary-unit)</th>
<th>Other</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Internal Controls (relative %)</td>
<td>44.7</td>
<td></td>
<td></td>
<td></td>
<td>100</td>
</tr>
<tr>
<td>Substantive (relative %)</td>
<td>28.6</td>
<td>6.7</td>
<td>58.3</td>
<td>6.4</td>
<td>100</td>
</tr>
</tbody>
</table>

(5) When using haphazard sampling to test internal controls, how likely is it that a preparer would receive workpaper review comments related to the following topics?

<table>
<thead>
<tr>
<th>Comment Type</th>
<th>Mean (Std. Dev)</th>
<th>% Very Unlikely (1)</th>
<th>% Unlikely to Somewhat Unlikely (2-3)</th>
<th>% Likely to Somewhat Likely (4-5)</th>
<th>% Very Likely (6)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Conclusions Drawn</td>
<td>3.61 (1.20)</td>
<td>0</td>
<td>33.4</td>
<td>61.1</td>
<td>5.6</td>
</tr>
<tr>
<td>Documentation</td>
<td>4.17 (0.92)</td>
<td>0</td>
<td>16.7</td>
<td>77.8</td>
<td>5.6</td>
</tr>
<tr>
<td>Sample Representativeness</td>
<td>2.78 (1.31)</td>
<td>16.7</td>
<td>50</td>
<td>33.3</td>
<td>0</td>
</tr>
</tbody>
</table>

Panel B: Client Timeliness and Scenarios (n=18)

(1) How common is it that the timeliness of a client response delays the auditor while performing an audit procedure?

<table>
<thead>
<tr>
<th>Scale Endpoints</th>
<th>Mean (Std. Dev)</th>
<th>% Very Uncommon (1)</th>
<th>% Responses below midpoint (1-3)</th>
<th>% Responses above midpoint (4-6)</th>
<th>% Very Common (6)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 = Very uncommon</td>
<td>5.17 (0.86)</td>
<td>0</td>
<td>5.6</td>
<td>94.4</td>
<td>38.9</td>
</tr>
<tr>
<td>6 = Very common</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

(2) Joe Auditor at XYZ Firm is haphazardly selecting a sample across four divisions of the client. Joe knows that the manager of division 2 has been slow in responding to past audit requests. Joe is working against a deadline. What is the likelihood that Joe may avoid selecting items from the manager of division 2, if possible?

<table>
<thead>
<tr>
<th>Scale Endpoints</th>
<th>Mean (Std. Dev)</th>
<th>% Not Likely at All (1)</th>
<th>% Responses below midpoint (1-3)</th>
<th>% Responses above midpoint (4-6)</th>
<th>% Very Likely (6)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 = Not likely at all</td>
<td>3.89 (1.08)</td>
<td>0</td>
<td>33.3</td>
<td>66.7</td>
<td>5.6</td>
</tr>
<tr>
<td>6 = Very likely</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

(3) Joe Auditor at XYZ Audit Firm is selecting a sample using a random number generator. After using the random number generator, Joe notices that a few items in the sample will be difficult to test (i.e. source of information is known to be slow, items appear to be complex, etc.). What is the likelihood that Joe may choose to run the random number generator again in order to obtain a sample that is more desirable?

<table>
<thead>
<tr>
<th>Scale Endpoints</th>
<th>Mean (Std. Dev)</th>
<th>% Not Likely at All (1)</th>
<th>% Responses below midpoint (1-3)</th>
<th>% Responses above midpoint (4-6)</th>
<th>% Very Likely (6)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 = Not likely at all</td>
<td>2.56 (1.54)</td>
<td>27.7</td>
<td>77.8</td>
<td>22.2</td>
<td>5.6</td>
</tr>
<tr>
<td>6 = Very likely</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

* Participants were instructed to select all that apply. Therefore, reported percentages do not sum to 100 percent.

b Response percentages represent the relative percentage that each sampling method is used. Responses were required to sum to 100%. Reported percentages do not sum exactly to 100% due to rounding.
Figure 1 - Sample Selections

Sample Selections from Client Manager of Interest for Self to Perform Testing

Sample Selections from Client Manager of Interest for Other to Perform Testing