

# **Do Auditors Expect to be Rewarded for Inaction?**

Jonathan Gay  
University of South Carolina  
jonathan.gay@grad.moore.sc.edu

Scott B. Jackson  
University of South Carolina  
scott.jackson@moore.sc.edu

Aaron Zimbelman  
University of South Carolina  
aaron.zimbelman@moore.sc.edu

April 2024

We are grateful to Chelsea Rae Austin, Joe Brazel, Susan Brinson, Owen Brown, Niki Bruno, Sanjeev Chandwani, Leah Conrad, Jared Eutsler, Alaina Gay, Scott C. Jackson, Steven Jackson, Macy Knutson, Phil Lamoreaux, Ethan LaMothe, Blair Marquardt, Elaine Mauldin, Mark Nelson, Chad Proell, Pradeep Sapkota, Casey Schwab, Spenser Seifert, Laura Self, Josh Shoulders, Chad Stefaniak, Kristen Thompson, Scott Vandervelde, Nate Waddoups, Grazia Xiong, and workshop participants at Baylor University, the University of North Texas, and the University of South Carolina for insightful comments and/or assistance obtaining participants. We are also grateful to the audit professionals who generously participated in our experiments or who supported our research by assisting us obtain participants from their firms.

## **Do Auditors Expect to be Rewarded for Inaction?**

### ***Abstract***

We examine whether audit staff expect to receive higher performance evaluations for inaction than skeptical action. Inaction is efficient in the absence of a material misstatement, and staff may expect to be rewarded for efficiency-enhancing decisions. However, inaction can degrade audit quality by leaving material misstatements undetected. We focus on expected evaluations because they reflect a lens through which audit staff view their professional incentives. When a misstatement is absent (the most common outcome of skeptical action), we find that (1) audit staff expect inaction to be rewarded, but supervisors do not reward inaction and (2) audit staff expect skeptical action to be penalized, but supervisors instead reward skeptical action. Taken together, these results suggest that audit staff evaluation expectancies are miscalibrated in ways that supervisors likely view as undesirable.

## 1. Introduction

In this study, we examine (1) whether audit staff expect to receive higher performance evaluations for inaction than skeptical action and (2) how supervisors evaluate audit staff who choose inaction over skeptical action.<sup>1</sup> Skeptical action can improve audit quality by revealing financial statement errors. Accounting firms encourage skeptical action because such behavior may increase audit effectiveness and prevent audit failures (Glover and Prawitt 2014; Arens et al. 2020; Messier et al. 2022). At the same time, accounting firms also encourage inaction because it increases efficiency and engagement profitability. Audit staff often face a consequential choice between skeptical action and inaction under conditions of ambiguity and uncertainty. While there are obvious benefits of skeptical action, theory suggests that audit staff often favor inaction over skeptical action, and they often expect to be rewarded for inaction decisions. However, inaction could degrade audit quality by leaving material misstatements undetected. Efforts to understand factors affecting audit staff action/inaction decisions are of considerable practical importance as those efforts could improve audit quality.

To understand preferences for action versus inaction among audit staff, we focus on expected evaluations because those expectations reflect a lens through which audit staff view their professional incentives. If staff expect inaction to result in higher performance evaluations, they are more likely to choose inaction. Similarly, if audit staff expect skeptical action to result in lower evaluations, they are less likely to take such action. Incentives faced by staff, including those created by performance evaluations, may critically influence the action/inaction decisions

---

<sup>1</sup> Consistent with Nelson (2009) and with the design of our experiments, we use the phrase “skeptical action” to refer to situations in which an auditor encounters cues or red flags suggesting that unplanned audit procedures may be warranted, and the auditor elects to perform those procedures and collect additional audit evidence. Conversely, we use the word “inaction” to refer to situations in which an auditor encounters cues or red flags suggesting that unplanned audit procedures may be warranted, but the auditor elects not to perform those procedures or collect additional audit evidence.

made by audit staff. To help assess whether staff correctly calibrate their evaluation expectations, we elicit supervisor evaluations. Comparisons between audit staff and supervisors could reveal expectations gaps that have implications for audit effectiveness and audit efficiency.

Although action is typically viewed in a favorable light, auditors cannot exhaustively examine every issue, claim, inconsistency, or concern that comes to their attention during the conduct of an audit. In other words, auditors must eventually decide not to collect additional audit evidence. This decision is borne out of economic necessity. Audit firms face competitive pressure (O’Keefe et al. 1994; Bills and Stephens 2016), partners face client retention pressure (Chang and Hwang 2003; Hatfield et al. 2008), managers face efficiency pressure (Houston 1999), and audit staff face budget pressure (DeZoort and Lord 1997; Weber and Stefaniak 2018). More generally, the societal costs of audits should not exceed their benefits (Arens et al. 2020; Messier et al. 2022).

Structural features of auditing may encourage inaction by audit staff for several reasons. First, audits normally involve time budgets which encourage auditors to carefully consider how and where to spend their time (Asare et al. 2000). Inaction saves time and effort, which could allow auditors to avoid going over budget and increase audit efficiency. Second, although there is certainty associated with the cost of taking skeptical action (e.g., incremental time and client disruptions), there is usually uncertainty associated with the benefits of taking skeptical action (e.g., discovery of an error or a control deviation). Asymmetry between the costs and benefits of skeptical action could make inaction appealing to audit staff, especially staff who have limited experience with financial statement errors (Ashton 1991; Durney et al. 2014).

Third, evidence suggests that the status quo exerts a strong influence on decision makers (Eidelman and Crandall 2012; Samuelson and Zechhauser 1988), and audit staff may view audit

programs, and prior year audit procedures that they often use as a reference, as the status quo. Unplanned procedures frequently associated with skeptical action may be viewed as a departure from the comfort and security of the status quo. Fourth, audits are often conducted in hierarchical teams, so there may be diffusion of responsibility, particularly among auditors at lower levels of the engagement hierarchy. Audit staff may reason that any misstatement that goes undetected due to their inaction will instead be detected by other auditors on the engagement team, thereby muting any impetus that the staff might otherwise feel to take skeptical action.<sup>2</sup> Conversely, audit staff may reason that if a misstatement is not detected by other engagement team members that the misstatement will never be detected by anyone, which avoids consequences to both the audit staff and the firm.

Although inaction represents a possible cause of audit failures (Beasley et al. 2001) and audit defects (Boyle et al. 2013), audit staff may judge their inaction decision to be validated if a misstatement is absent (i.e., when skeptical action would not have revealed a misstatement). At the same time, audit staff may judge their skeptical action decision to be invalidated when a misstatement is absent. As such, we expect audit staff to display outcome-dependent evaluation expectancies that overemphasize efficiency and underemphasize effectiveness.<sup>3</sup> Specifically, we hypothesize that audit staff expect higher evaluations for inaction when a misstatement is absent. Conversely, we hypothesize that staff expect lower evaluations for inaction when a misstatement is present. Panel A of Figure 1 summarizes our audit staff hypotheses. Turning to supervisors, we expect them to evaluate audit staff in a way that promotes effectiveness and places less emphasis

---

<sup>2</sup> An example of diffusion of responsibility comes from the “bystander effect” which refers to emergency situations in which an individual is less likely to respond when other individuals are present (Darley and Latane 1968; Latane and Darley 1968).

<sup>3</sup> We discuss research on outcome effects in Section 2, but we note here that outcome effect studies generally focus on outcomes related to some form of action, while our setting naturally involves both action and inaction.

on efficiency. As such, we hypothesize that supervisors evaluate skeptical action more favorably than inaction regardless of whether a misstatement is present. Panel B of Figure 1 summarizes our supervisor hypotheses.

Our design fully crosses outcomes and skeptical action/inaction, while measuring both expected and actual performance evaluations. While skeptical action naturally reveals outcomes, an apparent difficulty arises when examining outcomes related to inaction. Inaction normally limits outcome information available to decision makers and other observers. For example, an auditor may identify a misstatement through skeptical action, but the auditor would not identify a misstatement through inaction. However, there is an important structural feature present in audit settings that is lacking in most other settings. Specifically, our research design relies on account relatedness, which is an ecologically valid mechanism through which auditors learn outcomes even when inaction is chosen (Arens et al. 2020; Vandervelde 2006).<sup>4</sup>

We test our hypotheses in two experiments involving audit staff and audit supervisors. Each experiment involves a 2 (skeptical action versus inaction)  $\times$  2 (misstatement absent versus misstatement present) fully factorial between-subjects design. Experiment 1 elicits audit staff expected evaluations, while Experiment 2 elicits supervisor evaluations of audit staff. In each experiment, the instrument appropriately documents the stimulus for skeptical action even in the inaction condition.<sup>5</sup> Aside from specifying different dependent variables, the experiments are nearly identical.

---

<sup>4</sup> The principal implication of account relatedness is that errors and misstatements may be identified by different audit procedures in different areas of an audit. Often, a misstatement discovered in one area of the audit has implications for other business cycles/processes (Messier et al. 2022). For example, the discovery of an unrecorded payable would often reveal a misstatement in inventory (or cost of goods sold, if the inventory balance has already been corrected).

<sup>5</sup> It is possible that audit staff do not document the stimulus for skeptical action in the audit workpapers to avoid drawing attention to their inaction decision. However, Bennett and Hatfield (2013) find that most auditors appropriately document the circumstances in working papers even when they have chosen inaction.

The results of Experiment 1 support our audit staff hypotheses. Specifically, we find that staff expect higher evaluations for inaction than skeptical action when a misstatement is absent, while they expect higher evaluations for skeptical action than inaction when a misstatement is present. These findings should be interpreted in the context of misstatement frequencies. In a survey of audit partners, Maksymov et al. (2023) find that material misstatements are identified in only nine percent of public company audit engagements. As such, when audit staff deliberate between skeptical action and inaction, they may favor inaction because that choice is perceived to maximize their performance evaluation (assuming staff are aware that material misstatements are generally infrequent).

The results of Experiment 2 support our audit supervisor hypotheses, and those results provide a useful backdrop to understand audit staff expected evaluations. Specifically, we find that supervisors evaluate audit staff more favorably for skeptical action than inaction regardless of whether a misstatement is present. However, we also find that the presence of a misstatement accentuates the extent to which supervisors reward audit staff for skeptical action. Overall, our evidence not only suggests that supervisor evaluations of audit staff encourage skeptical action, but that there is also no circumstance in which those evaluations discourage skeptical action.

In comparing audit staff expected evaluations to supervisor evaluations, we look at potential divergence when a misstatement is absent (this is the most common situation) and when a misstatement is present (this is the least common situation). Beginning with the first situation, Experiment 1 suggests that when a misstatement is absent, staff expect inaction to be rewarded and skeptical action to be penalized. However, Experiment 2 suggests that when a misstatement is absent, supervisors do not reward inaction and they nonetheless reward skeptical action. This divergence between audit staff and supervisors suggests that staff may not collect evidence

desired by supervisors out of a mistaken belief that inaction in certain situations improves their performance evaluations. An implication of this divergence is that public accounting firms can partially remediate the problem of inadequate auditor skepticism by better educating staff about the expectations of supervisors.

Turning to the second situation, we find convergence between audit staff evaluation expectancies and supervisor evaluations. Experiment 1 suggests that when a misstatement is present, staff believe skeptical action is rewarded and inaction is penalized, and Experiment 2 suggests that supervisor evaluations largely match audit staff expectations. The only difference is that audit staff underestimate the extent to which supervisors reward skeptical action when a misstatement is present.

Finally, we explore responses to open-ended questions in which participants provide an explanation for their evaluation (supervisors) or expected evaluation (audit staff). We find that supervisors almost always explain performance evaluations by referring to some type of process that audit staff should follow. Staff also explain expected performance evaluations by referring to some type of process, but the frequency of such explanations is lower for audit staff than for supervisors. Further, we find that supervisors infrequently explain performance evaluations by referring to outcomes, while audit staff often explain performance evaluations by referring to outcomes. Thus, audit staff and supervisors differentially consider process and outcomes in the context of performance evaluations.

This study contributes to the accounting literature by providing initial evidence that audit staff expect inaction to be rewarded and skeptical action to be penalized when a misstatement is absent. This evaluation expectancy is concerning not only because it diverges rather sharply from supervisor evaluations, but because it is at odds with achieving high audit quality. Related to this



finding, we find no evidence that evaluation systems in accounting firms contribute to inadequate auditor skepticism. Instead, it appears that staff misperceptions of how they will be evaluated by their supervisors potentially dissuade audit staff from taking skeptical action. Thus, our evidence suggests that public accounting firms can promote skeptical action by better educating audit staff about supervisor expectations.

## **2. Hypotheses development**

### ***2.1. Outcome effects***

Outcome effects arise when performance evaluations of decision makers are influenced by the outcomes of their decisions (Tan and Lipe 1997). Although outcome information may be informative about decision quality, evaluators often ignore the uncertainty inherent in decisions at the time decisions were made. As a result, evaluations are frequently positively or negatively skewed by outcome information (Baron and Hershey 1988; Mertins et al. 2013). The typical outcome effect study involves the evaluation of an individual who has taken some type of action, such as expanding a business (Brown and Solomon 1987), investing in a business (Tan and Lipe 1997), investigating a variance (Lipe 1993), or investigating an inconsistency (Brazel et al. 2016; Nelson and Proell 2018). A positive outcome is frequently seen as a validation of the individual's decision to act, while a negative outcome is often seen as a refutation of that decision.

The counterpart to action – inaction – is not part of most outcome effect studies because outcomes are usually unobservable in the absence of action. For example, if a manager chooses not to expand a business, then the additional returns from expansion would be unobservable. Likewise, if a manager chooses not to investigate a variance, then the presence of a systemic problem would be unobservable. In addition, some settings engender the issue of unobservable inaction. This occurs when evaluators are unaware that a manager faced a choice between action

and inaction because inaction was chosen. In these instances, the evaluator would be unaware of the need to evaluate the wisdom of the manager's inaction.

Auditing is a rather unique setting for inaction research because (1) inaction is generally documented in workpapers and (2) outcomes of inaction are often known. Bennett and Hatfield (2013) find that the decision not to engage in follow-up audit procedures (a form of inaction) is transparently documented in audit workpapers much of the time. Moreover, audits often reveal the outcome of inaction due to account relatedness (Arens et al. 2020; Vandervelde 2006). If an auditor chooses inaction, the procedures undertaken for other accounts may reveal the presence of a misstatement that would have been revealed if the auditor had chosen skeptical action (we elaborate on this point in the next section). Thus, audit settings often reveal both the presence of an action/inaction decision to supervisors and the outcome of that decision even when inaction is chosen by audit staff.

## ***2.2. Hypotheses***

In auditing, skeptical action and inaction correspond to two possible states of the world – a misstatement is absent or present. Skeptical action either reveals that a misstatement is absent (an outcome contradicting action) or present (an outcome supporting action). Although inaction does not reveal the true state of the world, that information may be revealed elsewhere. Auditing standards and academic research consider account relatedness, and it is widely recognized that errors and misstatements in one account are connected to, and potentially uncovered during, the examination of other accounts (Arens et al. 2020; Vandervelde 2006). As such, consideration of audit results in totality allows audit teams to learn about different states of the world even when inaction is chosen.

When conducting an audit, there may be an indication that something is awry (e.g., an inconsistency).<sup>6</sup> Audit staff may prefer inaction over skeptical action for various reasons, and budgetary concerns are likely to play a central role in that preference. Budgets are used in nearly every type of organization, including professional services firms. However, concerns are often expressed that budgets induce dysfunctional behaviors (Collins et al. 1987; Bart 1988; Jensen 2001; Van der Stede 2000; Hope and Fraser 2003). In audit contexts, time budgets have been linked to, for example, reduced testing (Asare et al. 2000), premature sign-offs (Alderman and Dietrick 1982), quality reducing acts (Kelley and Margheim 1990), and underreporting of time (Ponemon 1992). If attention is paid to an audit staff's hours relative to budget, then staff may adopt the goal of staying within the time budget.<sup>7</sup> One way to accomplish that goal is to make efficient skeptical action decisions (i.e., avoid skeptical action when a misstatement is absent).

Skeptical action and inaction are related to the concepts of effectiveness and efficiency (Messier et al. 2022). Figure 1 describes the relations in the present context. If an audit staff has nothing to show for taking skeptical action (i.e., a misstatement is absent), then the staff is likely to have a relatively low evaluation expectancy because their skeptical action decision seems to be inefficient (Cell [A] of Figure 1, which is a Type I error). Similarly, if the audit staff opts for inaction but a misstatement is present, then the staff is likely to have a relatively low evaluation expectancy because their decision seems to be ineffective (Cell [D] of Figure 1, which is a Type II error). In both situations, outcome information shapes staff expectancies, which is consistent

---

<sup>6</sup> Our discussion contemplates an inconsistency that is somewhat subtle, which affords audit staff the ability to justify skeptical action or inaction.

<sup>7</sup> Although overt emphasis on time budgets may have declined over time (Buchheit et al. 2003), evidence suggests that budget considerations may indirectly influence both performance evaluations and staffing decisions (Agoglia et al. 2015).

with research indicating that outcomes have an inordinate influence on performance appraisals and, by extension, expected appraisals (see previous section).

Choosing skeptical action when a misstatement is present is an effective decision (Cell [C] of Figure 1), and audit staff have something beneficial to match with the budgetary effects of their decision. In effect, audit staff can rationalize their budgetary performance by pointing to the beneficial outcome. Even if the skeptical action decision was unwise ex ante, it was proven to be the correct decision. In this circumstance, audit staff are likely to have relatively high evaluation expectancies. Similar reasoning applies to choosing inaction when a misstatement is absent (Cell [B] of Figure 1). Although inaction might be frowned upon by some evaluators, audit staff may reason that their decision was correct, given the outcome. In effect, the audit staff avoided unnecessary time without causing any obvious harm. This circumstance is likely to result in staff having relatively high evaluation expectancies. Our audit staff hypotheses are shown in Panel A of Figure 2. Hypotheses 1a through 1d, stated in alternative form, are as follows:

**Hypothesis 1a:** When a misstatement is absent, audit staff have relatively low evaluation expectancies when they choose skeptical action compared to when they choose inaction (Cell [A] < Cell [B] in Panel A of Figure 1).

**Hypothesis 1b:** When a misstatement is present, audit staff have relatively high evaluation expectancies when they choose skeptical action compared to when they choose inaction (Cell [C] > Cell [D] in Panel A of Figure 1).

**Hypothesis 1c:** When audit staff choose skeptical action, they have relatively low evaluation expectancies when a misstatement is absent compared to when a misstatement is present (Cell [A] < Cell [C] in Panel A of Figure 1).

**Hypothesis 1d:** When audit staff choose inaction, they have relatively high evaluation expectancies when a misstatement is present compared to when a misstatement is absent (Cell [B] > Cell [D] in Panel A of Figure 1).

Supervisor evaluations of audit staff are likely to be markedly different from audit staff expected evaluations in two cells of Figure 1. Although we expect that audit staff have relatively

low evaluation expectancies when they choose skeptical action and a misstatement is absent, we expect supervisors to have a relatively favorable assessment of that decision (Cell [A] of Figure 1). Skeptical action is an essential ingredient of high-quality audits, and failure to take skeptical action may be viewed as a potential audit deficiency (Beasley et al. 2001). Thus, supervisors are likely to subordinate the desire to achieve efficiency to the desire to achieve audit effectiveness. In turn, we expect supervisors to award relatively high evaluations to staff who choose skeptical action even when a misstatement is absent.

Although we hypothesize that audit staff expect to be positively evaluated when they opt for inaction and a misstatement is absent, we expect supervisors to be critical of that decision for two main reasons (Cell [B] of Figure 1). First, the fact that a misstatement was absent does not absolve an audit staff of being appropriately skeptical. The absence of a misstatement could be viewed as luck rather than good decision making. Second, the risk of an ineffective audit looms large to supervisors as the consequences of an audit failure could be catastrophic for the firm and audit supervisors. As a result, we expect supervisors to give staff relatively low evaluations for choosing inaction even when a misstatement is absent. Our supervisor hypotheses are shown in Panel B of Figure 2. Hypotheses 2a through 2d, stated in alternative form, are as follows:

**Hypothesis 2a:** When a misstatement is absent, supervisors award audit staff higher evaluations for choosing skeptical action than inaction (Cell [A] > Cell [B] in Panel B of Figure 1).

**Hypothesis 2b:** When a misstatement is present, supervisors award audit staff higher evaluations for choosing skeptical action than inaction (Cell [C] > Cell [D] in Panel B of Figure 1).

**Hypothesis 2c:** When audit staff choose skeptical action, supervisors award lower evaluations when a misstatement is absent than when a misstatement is present (Cell [A] < Cell [C] in Panel B of Figure 1).

**Hypothesis 2d:** When audit staff choose inaction, supervisors award higher evaluations when a misstatement is absent than when a misstatement is present (Cell [B] > Cell [D] in Panel B of Figure 1).

An outcome effect for skeptical action has been shown previously in Brazel et al. (2016, 2019), so support for Hypothesis 2c would represent a conceptual replication of prior research.

### **3. Experiment 1**

#### **3.1. Purpose**

The purpose of this experiment is to test Hypotheses 1a through 1d, which focus on audit staff expected evaluations.<sup>8</sup>

#### **3.2. Participant recruitment and demographics**

We recruited 164 participants from seven Master of Accountancy programs. Depending on the preference of the instructor, some students were offered extra credit for their participation. Of the 164 participants, 91 have had an internship. We include all participants in our analyses, but our inferences and conclusions are unchanged if we exclude those participants who have not had an internship.<sup>9</sup> Participants average about 24 years of age, and about 48 percent are female.

#### **3.3. Instrument**

We develop an instrument that is modeled after Brazel et al. (2016) and Nelson and Proell (2018).<sup>10</sup> The instrument briefly describes a hypothetical public company that has multiple divisions. Participants learn that the firm has audited the company for the past ten years and has

---

<sup>8</sup> The experiments reported in this paper received Institutional Review Board approval.

<sup>9</sup> We consider Master of Accountancy students to be reasonable proxies for audit staff for several reasons. First, the majority of our participants have internship experience (about 55 percent), so vetting by firms in terms of suitability to function as audit staff has already occurred. Second, Saunders et al. (2023) observe that novice auditors (i.e., interns and new audit staff) increasingly perform tasks that were once performed by more experienced auditors. Third, studies often reveal that students with internship experience versus students without internship experience display judgments that are indistinguishable from one another (e.g., Brazel et al. 2016). Finally, studies often reveal that distinguishing between audit interns and audit staff in statistical analyses does not change inferences (Proell et al. 2022).

<sup>10</sup> We thank the co-authorship teams for sharing their research instruments.

issued an unqualified opinion each year. The instrument also indicates that (1) there have been few audit adjustments, (2) the engagement team has assessed management to be competent and trustworthy, (3) the audit partner emphasized to the audit team that the client is low risk, and the partner is concerned about audit costs, and (4) control testing reveals that there were no material discrepancies.

Participants take the role of an audit staff who is conducting analytical procedures over revenues (and related accounts) of the Sporting Goods division of the company. The instrument informs participants that all calculations and analyses called for in the audit program have been performed. The procedures revealed “negligible” differences between expectations and reported account balances. A short summary schedule of the relevant calculations is also provided.

Next, the instrument informs participants about non-financial measures, which reveals an inconsistency between division revenues and the number of division employees and square feet of division production space. Division employees declined from 158 to 152 employees (a decline of about four percent) and the production space declined from 148,000 to 143,000 square feet (a decline of about three percent). The direction of these changes is in the opposite direction of the year-to-year change in revenues and expense accounts, which all increased by a little more than one percent. Participants learn that they must decide whether to investigate the inconsistency between the financial and non-financial measures. In the prior year, the non-financial measures were not considered, and there is no requirement to consider the non-financial measures in the current year.

The experiment is a 2 (skeptical action versus inaction)  $\times$  2 (misstatement absent versus present) fully factorial between-subjects design with random assignment. Participants learn about the two possible courses of action:

You decide to investigate the inconsistency. This decision would place greater emphasis on the non-financial measures, which would deviate from what was done in the prior year and what is called for in the current year audit program. Making this decision would also cause both you and client management to incur a substantial amount of additional time. As a result, you will go over budget and cause some friction between client management and the audit team.

You decide NOT to investigate the inconsistency. This decision would place greater emphasis on traditional financial measures, which would follow what is called for in the current year audit program. Making this decision would also avoid causing both you and client management to incur a substantial amount of additional time. As a result, you will stay within budget and avoid causing any friction between client management and the audit team.

Participants then learn which course of action was chosen and the outcome of that action.<sup>11</sup> There are four separate conditions. In the investigate the inconsistency condition combined with the misstatement absent condition, participants learn the following outcome:

You found that the inconsistency described above was the result of the Sporting Goods division outsourcing some operations overseas. With this knowledge, you made several inquiries into the matter and collected additional audit evidence, which eventually led to the conclusion that NO significant overstatement existed in the Sporting Goods division revenue account.

In the investigate the inconsistency condition combined with the misstatement present condition, participants learn the following outcome:

You found that the inconsistency described above was the result of the Sporting Goods division outsourcing some operations overseas. With this knowledge, you made several inquiries into the matter and collected additional audit evidence, which eventually led to the conclusion that a significant overstatement existed in the Sporting Goods division revenue account as revenues were being recognized prematurely at the overseas operation.

In the inaction condition, participants do not investigate the inconsistency. However, the outcome is revealed by procedures in a separate area of the audit (Arens et al. 2017; Vandervelde

---

<sup>11</sup> We do not allow participants to choose whether to investigate the inconsistency as that would introduce endogeneity and limit the inferences that can be drawn from the experiment due to the lack of random assignment.



2006). In the do not investigate the inconsistency condition combined with the misstatement absent condition, participants learn the following outcome:

In a different area of the audit (property, plant, and equipment), another staff member unintentionally discovered that the inconsistency described above was the result of the Sporting Goods division outsourcing some operations overseas. By chance, this unintentional and unplanned discovery also provided evidence that NO significant overstatement existed in the Sporting Goods division revenue account.

In the do not investigate the inconsistency condition combined with the misstatement present condition, participants learn the following outcome:

In a different area of the audit (property, plant, and equipment), another staff member unintentionally discovered that the inconsistency described above was the result of the Sporting Goods division outsourcing some operations overseas. By chance, this unintentional and unplanned discovery also provided evidence that a significant overstatement existed in the Sporting Goods division revenue account, as revenues were being recognized prematurely at the overseas operation.

Notice that the overall audit effectiveness is held constant across the action/inaction conditions as the audit team (1) documented an explanation for the inconsistency and (2) obtained evidence of the presence or absence of a misstatement (either the participant or another staff member).<sup>12</sup>

Participants then respond to the main dependent variable, which states “Given the information summarized above, how do you think that your supervisor would evaluate your overall performance?” The 101-point response scale ranges from -50 (labeled “Significantly below expectations”) to 50 (labeled “Significantly above expectations”). There are two intermediate points -25 (labeled “Somewhat below expectations”) and 25 (labeled “Somewhat above expectations”) and a midpoint 0 (labeled “Met expectations”). We divide responses by ten for ease of presentation. We then ask an open-ended question, which states “In 2 or 3 sentences, please explain why you believe that your supervisor would give you that evaluation.” Finally,

---

<sup>12</sup> We acknowledge that a misstatement could exist but go undetected. However, the true state of the world is revealed with high confidence in our experiments. This allows us to reduce noise as we investigate outcome effects, especially in the inaction domain.

participants respond to manipulation checks and a few demographic questions.<sup>13</sup> On average, participants take about 16 minutes to complete the instrument.

### **3.4. Results**

Panel A of Figure 2 graphs cell means. Table 1 provides means by condition (Panel A), ANOVA results (Panel B), simple effect comparisons (Panel C), and follow-up tests (Panel D). Visual inspection of Panel A of Figure 2 reveals obvious similarities between our hypothesized results and the actual results.

Table 1 reports the  $2 \times 2$  ANOVA results in Panel B.<sup>14</sup> The dependent variable is the expected performance evaluation, and the two manipulated variables are ACT (skeptical action versus inaction) and MIS (misstatement absent versus misstatement present). We find that ACT is insignificant ( $F_{1, 160} = 0.07$ ,  $p\text{-value} = 0.79$ ) and so is MIS ( $F_{1, 160} = 0.97$ ,  $p\text{-value} = 0.33$ ). The lack of significant main effects for ACT and MIS is explained by the presence of a significant disordinal interaction ( $F_{1, 160} = 52.08$ ,  $p\text{-value} < 0.01$ ).

Panel C of Table 1 reports simple effect comparisons to test Hypotheses 1a through 1d. Hypothesis 1a predicts that when audit staff choose skeptical action, they have relatively low evaluation expectancies when a misstatement is absent compared to when a misstatement is present (Cell [A] < Cell [B] in Panel A of Figure 1). We find support for Hypothesis 1a as Cell [A] (mean = -1.34) is significantly less than Cell [B] (mean = 1.30) ( $F_{1, 160} = 26.69$ ,  $p\text{-value} < 0.01$ ). Hypothesis 1b predicts that when a misstatement is present, audit staff have relatively high evaluation expectancies when they choose skeptical action compared to when they choose inaction (Cell [C] > Cell [D] in Panel A of Figure 1). We also find support for Hypothesis 1b as

---

<sup>13</sup> Participants responded to the skeptical action/inaction manipulation check question correctly 97.89 percent of the time, and they responded to the no misstatement/misstatement manipulation check question correctly 95.79 percent of the time.

<sup>14</sup> We report conservative two-tailed  $p$ -values throughout this study.

Cell [C] (mean = 1.55) is significantly greater than Cell [D] (mean = -0.90) ( $F_{1, 160} = 25.40$ ,  $p\text{-value} < 0.01$ ).

Hypothesis 1c predicts that when audit staff choose skeptical action, they have relatively low evaluation expectancies when a misstatement is absent compared to when a misstatement is present (Cell [A] < Cell [C] in Panel A of Figure 1). We find support for Hypothesis 1c as Cell [A] (mean = -1.34) is significantly less than Cell [C] (mean = 1.55) ( $F_{1, 160} = 32.40$ ,  $p\text{-value} < 0.01$ ). Finally, Hypothesis 1d predicts that when audit staff choose inaction, they have relatively high evaluation expectancies when a misstatement is present compared to when a misstatement is absent (Cell [B] > Cell [D] in Panel A of Figure 1). We find support for Hypothesis 1d as Cell [B] (mean = 1.30) is significantly greater than Cell [D] (mean = -0.90) ( $F_{1, 160} = 20.19$ ,  $p\text{-value} < 0.01$ ). Thus, each of our audit staff hypotheses is strongly supported.

In follow-up tests, we also examine whether the means of Cells [A] through [D] in Panel A of Table 1 differ from the midpoint of the scale which is labeled “Met expectations.” We find that Cells [A] and [D] are significantly below “Met expectations” ( $t_{36} = -4.12$ ,  $p\text{-value} < 0.01$  and  $t_{43} = -2.78$ ,  $p\text{-value} < 0.01$ , respectively), which suggests that audit staff expect to be penalized when they choose skeptical action (inaction) and a misstatement is absent (present). Conversely, we find that Cells [B] and [C] are significantly above “Met expectations” ( $t_{40} = 4.01$ ,  $p\text{-value} < 0.01$  and  $t_{41} = 3.70$ ,  $p\text{-value} < 0.01$ , respectively), which suggests that staff expect to be rewarded when they choose inaction (skeptical action) and a misstatement is absent (present).

### **3.5. Discussion**

To place the experimental results in proper context, consider evidence on the frequency of material misstatements discovered during audits. Maksymov et al. (2023) find that material misstatements are identified in only nine percent of public company audit engagements, which

comports with an earlier conclusion by Ashton (1991) that auditors generally have limited direct experience with financial statement errors. Given these findings, the most common state of the world encountered by audit staff is that a misstatement is absent. As such, Cells [A] and [B] are the most likely evaluation outcomes. If audit staff choose skeptical action, their evaluation expectancy is to receive a negative performance evaluation. If audit staff choose inaction, their evaluation expectancy is to receive a positive performance evaluation. Collectively, audit staff evaluation expectancies suggest that inaction is judged to be the optimal decision.

Interestingly, the results of Experiment 1 suggest that audit staff expect that supervisors will violate one of the cornerstone principles of management accounting theory and practice, the controllability principle. This principle states that evaluators should hold evaluatees responsible only for outcomes that evaluatees can control (or influence) (Antle and Demski 1988; Arya et al 2007; Merchant and Van der Stede 2012). However, audit staff evaluation expectancies suggest that audit staff expect to be evaluated as if they can control the outcome of their skeptical action decision. Of course, audit staff have no control over the presence or absence of a misstatement. Experiment 2 will provide evidence about whether supervisors reward inaction and penalize skeptical action when a misstatement is absent and whether supervisors evaluate audit staff as though they can control outcomes.

## **4. Experiment 2**

### ***4.1. Purpose***

The purpose of this experiment is to test Hypotheses 2a through 2d which focus on audit supervisor evaluations of audit staff. In addition, we want to identify any divergence between the expected evaluations of audit staff and the evaluations of supervisors, which could help reveal why auditors sometimes display inadequate skepticism.

#### ***4.2. Participant recruitment***

We recruited participants in two ways. First, we recruited participants from a separate audit study (all of whom were audit seniors and above) who indicated a willingness to participate in a subsequent audit study.<sup>15</sup> Of the 113 individuals who indicated a willingness, 48 participated in the present study, which is a response rate of 42.48 percent. Second, we used a commercially available web scraping tool to capture the name, email, location, and employer of individuals who have professional titles indicating they are in public audit practice (e.g., audit senior, audit manager, audit partner). We sent an email to 4,460 audit professionals seeking their participation (607 emails were returned as undeliverable).<sup>16</sup> A second email was sent about two weeks later to the resulting 3,853 email addresses. Of the 3,853 audit professionals, 47 participated in the study, yielding a response rate of 1.22 percent. This experiment has 95 total participants.

#### ***4.3. Participant demographics***

Table 2 shows that participants are primarily senior managers (about 46 percent) and partners (about 34 percent). About 94 percent of participants are CPAs, and about 61 percent have experience with an international or a national accounting firm. On average, participants have given about 71 performance evaluations. Most participants have experience with public companies (about 61 percent). Participants average nearly 39 years of age and have about 14 years of accounting experience. Females comprise approximately 41 percent of our participants.

#### ***4.4. Instrument***

The instrument used in this experiment is similar to the one used in Experiment 1 with one primary difference. Rather than elicit audit staff expected evaluations, we elicit supervisor

---

<sup>15</sup> The study to which we refer is unrelated and was administered approximately six months prior.

<sup>16</sup> We did not send requests to audit professionals from one of the Big Four accounting firms because we were in discussions with that firm to obtain participants. Ultimately, the firm declined to provide support.

evaluations of a hypothetical audit staff member named Sam. It is clear in Experiment 1 that the subordinate has decision rights over whether to investigate the inconsistency. To create parity across experiments in terms of decision authority, we inform participants in Experiment 2 that Sam is expected to exercise professional judgment. On average, participants take about 16 minutes to complete the instrument.<sup>17</sup>

#### **4.5. Results**

Panel B of Figure 1 graphs cell means. Table 3 provides means by condition (Panel A), ANOVA results (Panel B), simple effect comparisons (Panel C), and follow-up tests (Panel D). Visual inspection of Panel B of Figure 1 reveals obvious similarities between our hypothesized results and the actual results. Notice that there appears to be outcome effects for both skeptical action (an upward sloping line) and inaction (a downward sloping line), along with an apparent preference for skeptical action over inaction. Further, the location of divergence between audit staff expected evaluations (Panel A) and supervisor evaluations (Panel B) seems to occur most noticeably when a misstatement is absent (Cells [A] and [B]).

Table 3 reports the  $2 \times 2$  ANOVA results in Panel B. The dependent variable is the performance evaluation, and the two independent variables are ACT (skeptical action versus inaction) and MIS (misstatement absent versus misstatement present). We find that ACT is significant ( $F_{1, 91} = 93.41$ ,  $p\text{-value} < 0.01$ ), which indicates that performance evaluations are higher when subordinates engage in skeptical action than when they engage in inaction. We also find that MIS is insignificant ( $F_{1, 91} = 0.95$ ,  $p\text{-value} = 0.33$ ), but there is a significant interaction

---

<sup>17</sup> Participants responded to the skeptical action/inaction manipulation check question correctly 97.89 percent of the time, and they responded to the no misstatement/misstatement manipulation check question correctly 95.79 percent of the time.

between ACT and MIS ( $F_{1, 91} = 21.25$ ,  $p\text{-value} < 0.01$ ), so the main effects should be interpreted with caution.

Panel C of Table 3 reports simple effect comparisons to test Hypotheses 2a through 2d. Hypothesis 2a predicts that when a misstatement is absent, supervisors award audit staff higher evaluations for choosing skeptical action than inaction (Cell [A] > Cell [B] in Panel B of Figure 1). We find support for Hypothesis 2a as Cell [A] (mean = 1.62) is significantly greater than Cell [B] (mean = -0.15) ( $F_{1, 91} = 11.80$ ,  $p\text{-value} < 0.01$ ). Similarly, Hypothesis 2b predicts that when a misstatement is present, supervisors award audit staff higher evaluations for choosing skeptical action than inaction (Cell [C] > Cell [D] in Panel B of Figure 1). We also find strong support for Hypothesis 2b as Cell [C] (mean = 3.57) is significantly greater than Cell [D] (mean = -1.42) ( $F_{1, 91} = 111.09$ ,  $p\text{-value} < 0.01$ ).

Hypothesis 2c predicts that when audit staff choose skeptical action, supervisors award lower evaluations when a misstatement is absent than when a misstatement is present (Cell [A] < Cell [C] in Panel B of Figure 1). We find support for Hypothesis 2c as Cell [A] (mean = 1.62) is significantly less than Cell [C] (mean = 3.57) ( $F_{1, 91} = 16.66$ ,  $p\text{-value} < 0.01$ ), consistent with prior research. Hypothesis 2d predicts that when audit staff choose inaction, supervisors award higher evaluations when a misstatement is absent than when a misstatement is present (Cell [B] > Cell [D] in Panel B of Figure 1). We find support for Hypothesis 2d as Cell [B] (mean = -0.15) is significantly greater than Cell [D] (mean = -1.42) ( $F_{1, 91} = 6.21$ ,  $p\text{-value} = 0.01$ ).

In follow-up tests, we also examine whether the means of Cells [A] through [D] in Panel A of Table 3 differ from the midpoint of the scale which is labeled “Met expectations.” When a misstatement is absent, we find that Cell [A] is significantly above the midpoint ( $t_{24} = 4.42$ ,  $p\text{-value} < 0.01$ ), while Cell [B] is indistinguishable from the midpoint ( $t_{18} = -0.40$ ,  $p\text{-value} = 0.69$ ).

This latter finding suggests that supervisors are ambivalent about the evaluation of audit staff who choose inaction when a misstatement is absent. Finally, when a misstatement is present, we find that Cell [C] is significantly above the midpoint ( $t_{24} = 12.45$ ,  $p\text{-value} < 0.01$ ), while Cell [D] is significantly below the midpoint ( $t_{25} = -3.94$ ,  $p\text{-value} < 0.01$ ).

#### **4.6. Discussion**

We compare audit staff expected evaluations (Panel A of Figure 1) to audit supervisor evaluations (Panel B of Figure 1). In the case of skeptical action when a misstatement is absent, we find that audit staff expect significantly negative evaluations (mean = -1.34), while supervisors give significantly positive evaluations (mean = 1.62). The difference between the expected evaluation and actual evaluation is statistically significant ( $t_{60} = 5.97$ ,  $p\text{-value} < 0.01$ ). This divergence is concerning because it suggests that audit staff wrongly believe that their skeptical action will be penalized when it will instead be rewarded. In the case of inaction when a misstatement is absent, we find that audit staff expect significantly positive evaluations (mean = 1.30), while supervisors give a “met expectations” evaluations (mean = -0.15). The difference between the expected evaluation and actual evaluation is again statistically significant ( $t_{58} = 2.69$ ,  $p\text{-value} < 0.01$ ). Audit staff appear to believe that the absence of a misstatement vindicates the inaction decision, while supervisors look past the outcome and focus on the potentially harmful consequences of inaction. This mistaken belief by audit staff could result in the failure to collect sufficient evidential matter in certain circumstances.

#### **5. Open-ended responses**

The experiments have an open-ended question which states “In 2 or 3 sentences, please explain why you believe that your supervisor would give you that evaluation” (audit staff) or “In 2 or 3 sentences, please explain your evaluation of Sam’s overall performance” (supervisor). We



examine responses to determine whether participants reference processes or outcomes related to the presence or absence of a misstatement. Given our experimental results, we expect supervisors to emphasize processes and downplay outcomes. In contrast, we expect audit staff to pay more attention to outcomes and less attention to processes. We view an emphasis on processes as an indication that performance evaluations are viewed as a training device that has a developmental purpose. Due to their inexperience, we expect audit staff to be less attuned to the developmental purpose of performance evaluations.<sup>18</sup>

We find that 45.12 percent (74/164) of staff participants mention outcomes, and 21.05 percent (20/95) of supervisor participants mention outcomes. These percentages are significantly different between audit staff and supervisors ( $\chi^2$ -statistic = 15.01, p-value < 0.01). Conversely, we find that 87.20 percent (143/164) of audit staff participants mention processes and 98.95 percent (94/95) of supervisor participants mention processes. These percentages are significantly different between audit staff and supervisors ( $\chi^2$ -statistic = 10.70, p-value < 0.01). Overall, these results suggest that some audit staff misunderstand the purpose of performance evaluations and ascribe undue importance to outcomes, whereas supervisors focus on helping audit staff develop appropriate decision processes while giving limited importance to outcomes.

## **6. Summary and implications**

Theory suggests that audit staff may display outcome-dependent evaluation expectancies that overemphasize efficiency and underemphasize effectiveness. As such, we hypothesize that audit staff expect higher evaluations for inaction (versus action) when a misstatement is absent,

---

<sup>18</sup> To determine whether responses are process focused or outcome focused, a co-author on this study and a doctoral student each independently coded responses without knowledge of the condition to which participants were assigned. Some responses were entirely process focused or entirely outcome focused, while others were mixed. There was initial agreement approximately 81 percent of the time and Cohen's Kappa is 0.60, indicating moderate agreement. Disagreements were resolved without exception.

while audit staff expect lower evaluations for inaction (versus action) when a misstatement is present. Our results confirm these expectations. At the same time, we expect audit supervisors to evaluate audit staff in a way that promotes effectiveness and places less emphasis on efficiency. As such, we hypothesize that supervisors evaluate audit staff who choose skeptical action more favorably than audit staff who choose inaction regardless of whether a misstatement is present or absent. Our results also confirm these expectations.

The results highlight the fact that audit staff and supervisors tend to balance efficiency and effectiveness very differently. The differences are especially pronounced when the outcome is that a misstatement is absent (this is also the most common outcome as material misstatements only arise in nine percent of public company audits). When this common outcome arises, we find that audit staff evaluation expectations are significantly miscalibrated because supervisors do not reward inaction and they do reward skeptical action. The practical implication of this finding is that staff may choose not to collect evidence out of the mistaken belief that such a decision will improve their performance evaluation. Thus, to the extent that evaluation systems used in public accounting firms contribute to inadequate auditor skepticism, it appears that staff misperceptions are a cause, not the evaluation proclivities of supervisors. One way that public accounting firms could promote skeptical action is by better educating audit staff about supervisor expectations.

## REFERENCES

- Agoglia, C., R. Hatfield, and T. Lambert. 2015. Audit team time reporting: An agency theory perspective. *Accounting, Organizations and Society* 44: 1-14.
- Alderman, C., and J. Deitrick. 1982. Auditors' perceptions of time budget pressures and premature sign-offs: A replication and extension. *Auditing: A Journal of Practice and Theory* 1 (2): 54-68.
- Antle, R., and J. Demski. 1988. The controllability principle in responsibility accounting. *The Accounting Review* 63 (4): 700-718.
- Arens, A., R. Elder, M. Beasley, and C. Hogan. 2020. *Auditing and Assurance Services: An Integrated Approach* 17th edition. New York, NY: Pearson.
- Arya, A., J. Glover, and S. Radhakrishnan. 2007. The controllability principle in responsibility accounting: Another look. In *Essays in Accounting Theory in Honour of Joel S. Demski*, edited by R. Antle, P. Liang, and F. Gjesdal, 183-198. New York, NY: Springer.
- Asare, S., G. Trompeter, and A. Wright. 2000. The effect of accountability and time budgets on auditors' testing strategies. *Contemporary Accounting Research* 17 (4): 539-560.
- Ashton, A. 1991. Experience and error frequency knowledge as potential determinants of audit expertise. *The Accounting Review* 66 (2): 218-239.
- Baron, J., and J. Hershey. 1988. Outcome bias in decision evaluation. *Journal of Personality and Social Psychology*, 54 (4): 569-579.
- Bart, C. 1988. Budgeting gamesmanship. *Academy of Management Perspectives* 2 (4): 285-294.
- Beasley, M., J. Carcello, and D. Hermanson. 2001. Top 10 audit deficiencies. *Journal of Accountancy* 19 (1): 63-66.
- Bennett, G., and R. Hatfield. 2013. The effect of social mismatch between staff auditors and client management on the collection of audit evidence. *The Accounting Review* 88 (1): 31-50.
- Bills, K., and N. Stephens. 2016. Spatial competition at the intersection of the large and small audit firm markets. *Auditing: A Journal of Practice & Theory* 35 (1): 23-45.
- Boyle, J., D. Boyle, D. Hermanson, and R. Houston. 2013. Quality control defects in smaller firms' PCAOB inspection reports. *The CPA Journal* 83 (6): 34-39.
- Brazel, J., S. Jackson, T. Schaefer, and B. Stewart. 2016. The outcome effect and professional skepticism. *The Accounting Review* 91 (6): 1577-1599.
- Brazel, J., C. Gimbar, E. Maksymov, and T. Scheafer. 2019. The outcome effect and professional skepticism: A replication and a failed attempt at mitigation. *Behavioral Research in Accounting* 31 (2): 135-143.
- Brown, C., and I. Solomon. 1987. Effects of outcome information on evaluations of managerial decisions. *The Accounting Review* 62 (3): 564-577.
- Buchheit, S., W. Pasewark, and J. Strawser. 2003. No need to compromise: Evidence of public accounting's changing culture regarding budgetary performance. *Journal of Business Ethics* 42 (2): 151-163.
- Chang, C., and N. Hwang. 2003. The impact of retention incentives and client business risks on auditors' decisions involving aggressive reporting practices. *Auditing: A Journal of Practice & Theory* 22 (2): 207-218.
- Collins, F., P. Munter, and D. Finn. 1987. The budgeting games people play. *The Accounting Review* 62 (1): 29-49.
- Darley, J., and B. Latane. 1968. Bystander intervention in emergencies: Diffusion of responsibility. *Journal of Personality and Social Psychology* 8 (4): 377-383.

- DeZoort, F., and A. Lord. 1997. A review and synthesis of pressure effects research in accounting. *Journal of Accounting Literature* 16: 28-85.
- Durney, M., R. Elder, and S. Glover. 2014. Field data on accounting error rates and audit sampling. *Auditing: A Journal of Practice & Theory* 33 (2): 79-110.
- Eidelman, S., and C. Crandall. 2012. Bias in favor of status quo. *Social and Personality Psychology* 6 (3): 270-281.
- Glover, S., and D. Prawitt. 2014. Enhancing auditor professional skepticism: The professional skepticism continuum. *Current Issues in Auditing* 8 (2): 1-10.
- Hatfield, R., Agoglia, and M. Sanchez. 2008. Client characteristics and the negotiation tactics of auditors: Implications for financial reporting. *Journal of Accounting Research* 46 (5): 1183-1207.
- Houston, R. 1999. The effect of fee pressure and client risk on audit seniors' time budget decisions. *Auditing: A Journal of Practice & Theory* 18 (2): 70-86.
- Hope, J., and R. Fraser. 2003. New ways of setting rewards: The beyond budgeting model. *California Management Review* 45 (4): 104-119.
- Jensen, M. 2001. Corporate budgeting is broken – let's fix it. *Harvard Business Review* 70 (10): 95-101.
- Kelley, T., and L. Margheim. 1990. The impact of time budget pressure, personality, and leadership variables on dysfunctional audit behavior. *Auditing: A Journal of Practice & Theory* 9 (2): 21-42.
- Latane, B., and J. Darley. 1968. Group inhibition of bystander intervention in emergencies. *Journal of Personality and Social Psychology* 10 (3): 215-221.
- Lipe, M. 1993. Analyzing the variance investigation decision: The effects of outcomes, mental accounting, and framing. *The Accounting Review* 68 (4): 748-764.
- Maksymov, E., M. Peecher, A. Sutherland, and J. Weber. 2023. Audit partners' role in material misstatement resolution: Survey and interview evidence. *Journal of Accounting Research* (forthcoming).
- Merchant, K., and W. Van Der Stede. 2012. *Management Control Systems: Performance Measurement, Evaluation and Incentives*. 3rd edition. Harlow, England: Pearson Education.
- Mertins, L., D. Salvador, and J. Long. 2013. The outcome effect - A review and implications for future research. *Journal of Accounting Literature* 31 (1): 2-30.
- Messier, W., S. Glover, and D. Prawitt. 2022. *Auditing & Assurance Services: A Systematic Approach*. New York, NY: McGraw-Hill.
- Nelson, M. 2009. A model and literature review of professional skepticism in auditing. *Auditing: A Journal of Practice and Theory* 28 (2): 1-34.
- Nelson, M., and C. Proell. 2018. Is silence golden? Audit team leader reactions to subordinates who speak up "in the moment" and at performance appraisal. *The Accounting Review* 93 (6): 281-300.
- O'Keefe, T., D. Simunic, and M. Stein. 1994. The production of audit services: Evidence from a major public accounting firm. *Journal of Accounting Research* 32 (2): 241-261.
- Ponemon, L. 1992. Auditor underreporting of time and moral reasoning: An experimental lab study. *Contemporary Accounting Research* 9 (1): 171-189.
- Proell, C., Zhou, Y., and M. Nelson. 2022. It's not only what you say... How communication style and team culture affect audit issue follow-up and auditor performance evaluations. *The Accounting Review* 97 (2): 373-395.

- Samuelson, W., and R. Zeckhauser. 1988. Status quo bias in decision making. *Journal of Risk and Uncertainty* 1 (1): 7-59.
- Saunders, K., M. Keune, and E. Hawkins. 2023. More than making copies: Survey evidence on the work of novice auditors. *Auditing: A Journal of Practice & Theory* 42 (4): 131-156
- Tan, H., and M. Lipe. 1997. Outcome effects: The impact of decision process and outcome controllability. *Journal of Behavioral Decision Making* 10 (4): 315-325.
- Van der Stede, W. 2000. The relationship between two consequences of budgetary controls: Budgetary slack creation and managerial and short-term orientation. *Accounting, Organizations and Society* 25 (6): 609-622.
- Vandervelde, S. 2006. The importance of account relations when responding to interim audit testing results. *Contemporary Accounting Research* 23 (3): 789-821.
- Weber, J., and C. Stefaniak. 2018. Using undisclosed time budgets to reduce the magnitude of auditors' underreporting of chargeable time. *Behavioral Research in Accounting* 30 (2): 95-108.

**Exhibit 1**  
**Action, Inaction, and Outcomes**

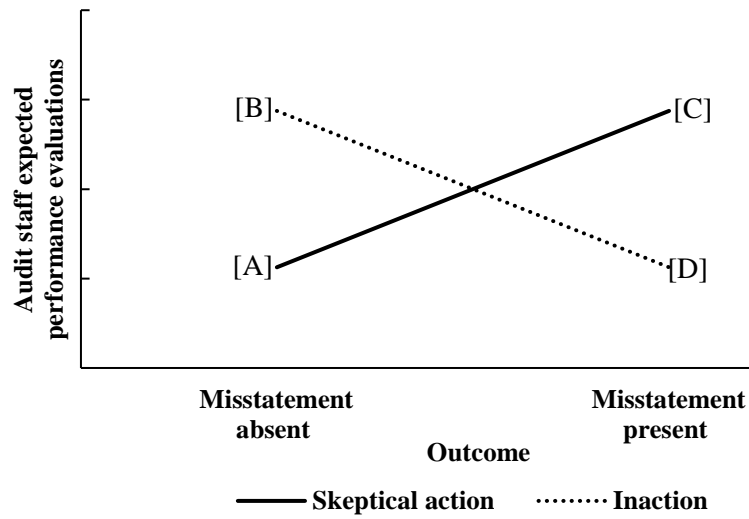
---

| <i>H</i> <sub>0</sub> : No misstatement present       | Misstatement                         |  |
|---|--------------------------------------|--|
|   | Absent ( <i>H</i> <sub>0</sub> true) | Present ( <i>H</i> <sub>0</sub> false) |
| <b>Skeptical action (reject <i>H</i><sub>0</sub>)</b> | [A] Inefficient (Type I error)       | [C] Effective decision (correct)       |
| <b>Inaction (do not reject <i>H</i><sub>0</sub>)</b>  | [B] Efficient decision (correct)     | [D] Ineffective (Type II error)        |

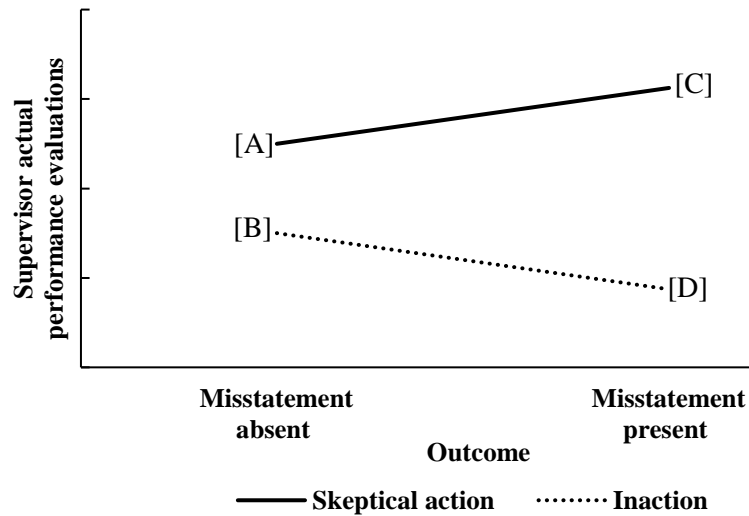
---

**Figure 1**  
**Hypothesized Results**

*Panel A: Audit staff hypotheses*



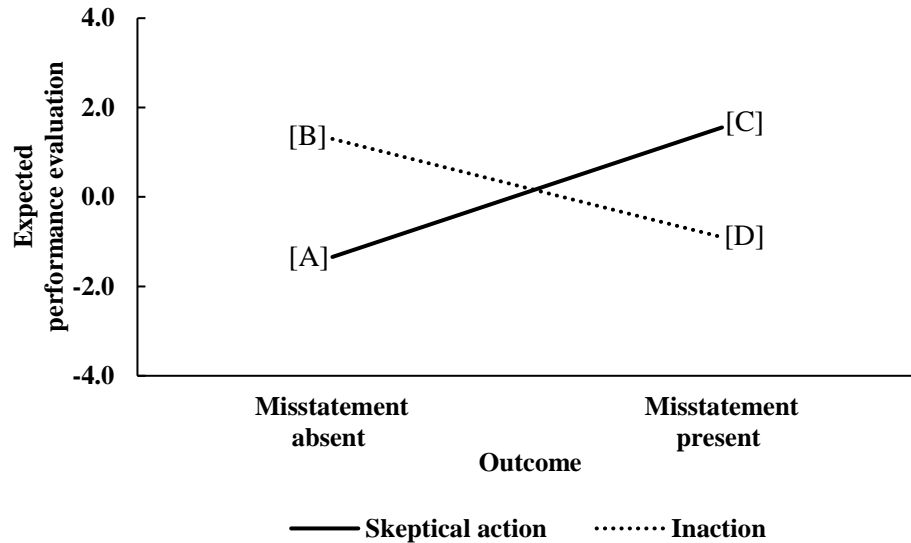
*Panel B: Supervisor hypotheses*



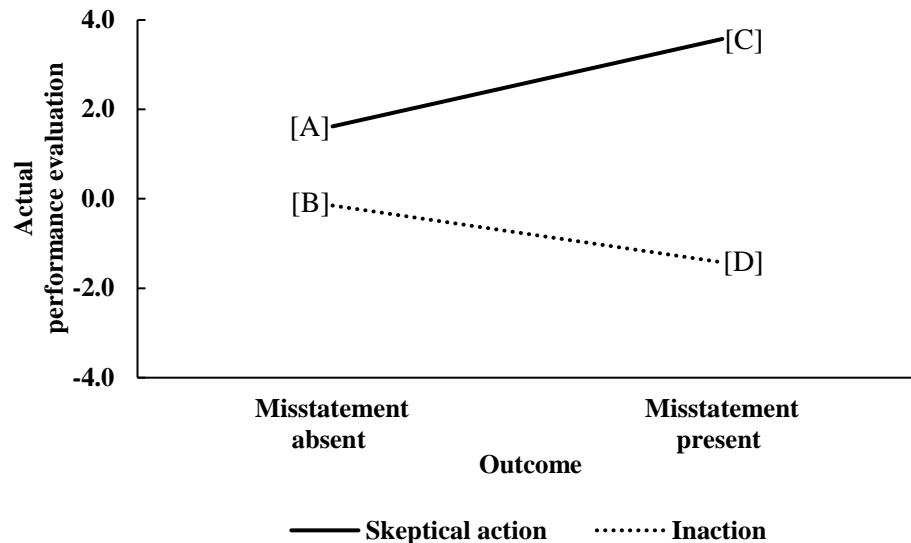
See Section 2 for the development of hypotheses.

**Figure 2**  
**Results for Experiments 1 and 2**

**Panel A: Results of Experiment 1 (audit staff)**



**Panel B: Results of Experiment 2 (audit supervisors)**



See Section 2 for development of hypotheses and Sections 3 and 4 for discussion of Experiments 1 and 2, respectively. Panel A of Table 1 provides cell means for Experiment 1, and Panel A of Table 3 provides cell means for Experiment 2. The manipulated variables are the decision whether to investigate the inconsistency (manipulated as skeptical action versus inaction) and outcome related to the decision to investigate the inconsistency (misstatement absent versus misstatement present). The dependent variable is participants' expected performance evaluation (Experiment 1) or participants' actual performance evaluation (Experiment 2). The 101-point response scale ranges from -50 (labeled "Significantly below expectations") to 50 (labeled "Significantly above expectations"). There are two intermediate points -25 (labeled "Somewhat below expectations") and 25 (labeled "Somewhat above expectations") and a midpoint 0 (labeled "Met expectations"). The dependent variable is divided by ten, so responses range from -5 to 5.



**Table 1**  
**Cell Means, ANOVA Results, and Simple Effect Comparisons for Experiment 1**  
**Panel A: Expected performance evaluations by condition – mean (standard deviation)**

| <b>Action (ACT)</b> | <b>Outcome (MIS)</b>             |                                  | <b>Overall</b>            |
|---------------------|----------------------------------|----------------------------------|---------------------------|
|                     | <b>Misstatement<br/>absent</b>   | <b>Misstatement<br/>present</b>  |                           |
| Skeptical<br>action | -1.34<br>(1.97)<br>n = 37<br>[A] | 1.55<br>(2.72)<br>n = 42<br>[C]  | 0.20<br>(2.79)<br>n = 79  |
| Inaction            | 1.30<br>(2.08)<br>n = 41<br>[B]  | -0.90<br>(2.13)<br>n = 44<br>[D] | 0.16<br>(2.37)<br>n = 85  |
| Overall             | 0.05<br>(2.42)<br>n = 78         | 0.30<br>(2.72)<br>n = 86         | 0.18<br>(2.57)<br>n = 164 |

**Panel B: Analysis of variance results**

| <b>Source</b> | <b>SS</b> | <b>df</b> | <b>MS</b> | <b>F-stat.</b> | <b>p-value</b> |
|---------------|-----------|-----------|-----------|----------------|----------------|
| ACT           | 0.37      | 1         | 0.37      | 0.07           | 0.79           |
| MIS           | 4.93      | 1         | 4.93      | 0.97           | 0.33           |
| ACT × MIS     | 264.54    | 1         | 264.54    | 52.08          | < 0.01         |
| Error         | 812.68    | 160       | 5.08      |                |                |

**Panel C: Simple effect comparisons to test Hypotheses 1a through 1d**

| <b>Comparison</b> | <b>SS</b> | <b>df</b> | <b>MS</b> | <b>F-stat.</b> | <b>p-value</b> |
|-------------------|-----------|-----------|-----------|----------------|----------------|
| H1a: [A] < [B]    | 135.58    | 1         | 135.58    | 26.69          | < 0.01         |
| H1b: [C] > [D]    | 129.01    | 1         | 129.01    | 25.40          | < 0.01         |
| H1c: [A] < [C]    | 164.59    | 1         | 164.59    | 32.40          | < 0.01         |
| H1d: [B] > [D]    | 102.53    | 1         | 102.53    | 20.19          | < 0.01         |

**Panel D: Follow-up tests of whether cell means differ from zero**

| <b>Comparison</b>                      | <b>Mean</b> | <b>t-stat</b> | <b>p-value</b> |
|--|-------------|---------------|----------------|
| Does [A] differ from met expectations? | -1.34       | -4.12         | < 0.01         |
| Does [B] differ from met expectations? | 1.30        | 4.01          | < 0.01         |
| Does [C] differ from met expectations? | 1.55        | 3.70          | < 0.01         |
| Does [D] differ from met expectations? | -0.90       | -2.78         | < 0.01         |

See Section 2 for development of the hypotheses and Section 3 for discussion of the experiment. The manipulated variables are the decision whether to investigate the inconsistency (manipulated as skeptical action versus inaction) and the outcome related to the decision to investigate the inconsistency (misstatement absent versus misstatement present). The dependent variable is participants' response to the question "Given the information summarized above, how do you think that your supervisor would evaluate your overall performance?" The 101-point response scale ranges from -50 (labeled "Significantly below expectations") to 50 (labeled "Significantly above expectations"). There are two intermediate points -25 (labeled "Somewhat below expectations") and 25 (labeled "Somewhat above expectations") and a midpoint 0 (labeled "Met expectations"). The dependent variable is divided by ten, so responses range from -5 to 5.

**Table 2**  
**Demographic Information for Participants in Experiment 2 (n = 95)**

| <b>Variable</b>                         | <b>Mean</b> | <b>S.D.</b> |
|---|-------------|-------------|
| Professional titles                     |             |             |
| Partner/principal (%)                   | 33.68       | 47.51       |
| Senior manager (%)                      | 46.32       | 50.13       |
| Manager (%)                             | 6.32        | 24.45       |
| Senior (%)                              | 13.68       | 34.55       |
| Certified Public Accountant (%)         | 93.68       | 24.45       |
| Past/current experience with firm types |             |             |
| International or national (%)           | 61.05       | 49.02       |
| Regional or local only (%)              | 38.95       | 49.02       |
| Performance evaluations given (count)   | 71.07       | 64.81       |
| Public company audit experience (%)     | 61.05       | 49.02       |
| Age (years)                             | 38.95       | 9.62        |
| Experience in accounting (years)        | 14.03       | 7.64        |
| Gender (% female)                       | 41.05       | 49.45       |

See Section 4.2 for participant recruitment procedures.

**Table 3**  
**Cell Means, ANOVA Results, and Simple Effect Comparisons for Experiment 2**  
**Panel A: Supervisor performance evaluations by condition – mean (standard deviation)**

| <b>Action (ACT)</b> | <b>Outcome (MIS)</b>             |                                  | <b>Overall</b>            |
|---------------------|----------------------------------|----------------------------------|---------------------------|
|                     | <b>Misstatement<br/>absent</b>   | <b>Misstatement<br/>present</b>  |                           |
| Skeptical<br>action | 1.62<br>(1.83)<br>n = 25<br>[A]  | 3.57<br>(1.43)<br>n = 25<br>[C]  | 2.60<br>(1.90)<br>n = 50  |
| Inaction            | -0.15<br>(1.60)<br>n = 19<br>[B] | -1.42<br>(1.84)<br>n = 26<br>[D] | -0.88<br>(1.83)<br>n = 45 |
| Overall             | 0.86<br>(1.93)<br>n = 44         | 1.03<br>(3.00)<br>n = 51         | 0.95<br>(2.55)<br>n = 95  |

**Panel B: Analysis of variance results**

| <b>Source</b> | <b>SS</b> | <b>df</b> | <b>MS</b> | <b>F-stat.</b> | <b>p-value</b> |
|---------------|-----------|-----------|-----------|----------------|----------------|
| ACT           | 266.98    | 1         | 266.98    | 93.41          | < 0.01         |
| MIS           | 2.70      | 1         | 2.70      | 0.95           | 0.33           |
| ACT × MIS     | 60.75     | 1         | 60.75     | 21.25          | < 0.01         |
| Error         | 260.10    | 91        | 2.86      |                |                |

**Panel C: Simple effect comparisons to test Hypotheses 2a through 2d**

| <b>Comparison</b> | <b>SS</b> | <b>df</b> | <b>MS</b> | <b>F-stat.</b> | <b>p-value</b> |
|-------------------|-----------|-----------|-----------|----------------|----------------|
| H2a: [A] > [B]    | 33.72     | 1         | 33.72     | 11.80          | < 0.01         |
| H2b: [C] > [D]    | 317.51    | 1         | 317.51    | 111.09         | < 0.01         |
| H2c: [A] < [C]    | 47.63     | 1         | 47.63     | 16.66          | < 0.01         |
| H2d: [B] > [D]    | 17.76     | 1         | 17.76     | 6.21           | 0.01           |

**Panel D: Follow-up tests of whether cell means differ from zero**

| <b>Comparison</b>          | <b>Mean</b> | <b>t-stat</b> | <b>p-value</b> |
|----------------------------|-------------|---------------|----------------|
| Does [A] differ from zero? | 1.62        | 4.42          | < 0.01         |
| Does [B] differ from zero? | -0.15       | -0.40         | 0.69           |
| Does [C] differ from zero? | 3.57        | 12.45         | < 0.01         |
| Does [D] differ from zero? | -1.42       | -3.94         | < 0.01         |

See Section 2 for development of the hypotheses and Section 4 for discussion of the experiment. The manipulated variables are the decision whether to investigate the inconsistency (manipulated as skeptical action versus inaction) and the outcome related to the decision to investigate the inconsistency (misstatement absent versus misstatement present). The dependent variable is participants' response to the question "Based on the information presented to you, how would you evaluate Sam's overall performance?" The 101-point response scale ranges from -50 (labeled "Significantly below expectations") to 50 (labeled "Significantly above expectations"). There are two intermediate points -25 (labeled "Somewhat below expectations") and 25 (labeled "Somewhat above expectations") and a midpoint 0 (labeled "Met expectations"). The dependent variable is divided by ten, so responses range from -5 to 5.