

Improving Auditors' Review of Inconsistent Audit Evidence

Abstract: We examine whether a promotion focus improves audit reviewers' detection of conclusion deficiencies, which occur when auditors fail to integrate their individual audit test work into their holistic evaluation of account balances. These deficiencies are particularly concerning when the audit evidence set includes inconsistencies. A promotion focus is a frame of mind that broadens attention and big picture thinking, whereas a prevention focus narrows attention and emphasizes avoiding mistakes. In two experiments, we provide evidence that a promotion focus improves reviewers' detection of conclusion deficiencies when considering inconsistent evidence. Specifically, we find that a promotion focus increases reviewers' identification of inconsistencies among evidence items and the extent to which they link these items to their evaluations of the overall sufficiency of evidence. Moreover, promotion focused reviewers are more likely to communicate inconsistent evidence to the engagement manager. Communicating issues for follow-up is important because quality control necessitates either conducting additional procedures or proposing an adjustment. In sum, we find that an intervention encouraging a promotion focus improves auditors' review of complex evidence sets containing items that are inconsistent with client assertions.

Keywords: Audit review; inconsistent evidence; red flags; nonfinancial measures; regulatory focus

I. INTRODUCTION

Auditors' consideration of inconsistent evidence is a bedrock of their professional judgment and skepticism (PCAOB 2011; Glover and Prawitt 2013; CAQ 2014). Auditors increasingly confront complex evidence that appears inconsistent with client assertions, but they struggle to evaluate this inconsistent evidence appropriately and holistically (Griffith, Hammersley, and Kadous 2015a; Griffith, Hammersley, Kadous, and Young 2015b; Kadous and Zhou 2016; PCAOB 2013, 2014; Church and Shefchik 2011). For instance, auditors may test inconsistent items but not recognize that the items raise doubts about the overall reasonableness of a client assertion (Griffith, et al. 2015a). We refer to these types of judgment errors, in which preparers fail to draw appropriate inferences about the audit evidence set as a whole, as *conclusion deficiencies*. Troublingly, PCAOB inspection results suggest that conclusion deficiencies survive audit engagements' key quality control, the review process (PCAOB 2014). When the audit evidence set includes inconsistencies, undetected conclusion deficiencies can lead audit teams to neglect the need to conduct additional tests or propose adjustments.

We examine an intervention to improve audit reviewers' detection of conclusion deficiencies. In our setting, common features of auditing may constrain the broad attention and holistic thinking necessary for effective review, especially when reviewing inconsistent evidence. For instance, reviewers often confront time pressure or use highly-structured audit programs (Agoglia, Hatfield, and Lambert 2015; Hammersley and Ricci 2020; Lambert, Jones, Brazel, and Showalter 2017). Regulatory focus theory suggests that these types of contextual features influence how people focus on information and, thus, whether they attend to information broadly, whether they think holistically, and how they pursue their goals (Higgins 1997; 1998). We argue that the audit context increases the likelihood that reviewers adopt one type of

regulatory focus, a *prevention focus*, which is a defensive frame of mind motivating reviewers to avoid mistakes. This type of focus inhibits abstract thinking and reduces comfort with large information sets (Higgins 1997; Polman 2012; Friedman and Förster 2005). By contrast, the other type of regulatory focus, a *promotion focus*, broadens attention and promotes holistic thinking (Higgins 1997; 1998; Higgins, Shah, and Friedman 1997; Friedman and Förster 2001; Förster and Higgins 2005; Kuschel, Förster, and Denzler 2010). We expect a broader focus helps reviewers consider audit evidence holistically, which is critical for recognizing inconsistencies among individual evidence items. Thus, we argue that reviewers with a promotion focus are less likely to “miss the forest for the trees” and, as such, will do a better job detecting conclusion deficiencies.

Further, we examine whether a promotion focus affects how reviewers respond to emerging changes in the structure of review tasks. Firms have changed audit support systems to allow subdividing review tasks (Dowling and Leech 2014). In this segmented review structure, reviewers individually assess each discrete test that comprises the body of evidence, whereas the more traditional unsegmented review structure frames the portfolio of tests as a single task. For example, in a segmented review structure, reviewers may separately assess (1) whether a sample of sales transactions exists, (2) whether a sample of sales transactions were included in the correct accounting period and, (3) whether any sales transactions were improperly excluded from the sales balance by separating these tasks and fitting them in among their other work, creating time gaps and distractions between tasks. In an unsegmented review structure, reviewers would assess all three of these tests at the same time, as a portfolio. Segmenting tasks likely narrows reviewers’ attention and inhibits holistic information processing, which in turn likely decreases the detection of conclusion deficiencies (see Driver and Baylis 1998; Driver et al. 2001). We

argue that a promotion focus mitigates the negative effects of segmented, as opposed to unsegmented, review.

We test our research questions in two experiments with professional auditors. In Experiment 1 we test the effect of promotion focus, and in Experiment 2 we test the interactive effect of promotion focus and segmented review. In both experiments, the case involves reviewing six tests for a sales balance, with some tests over financial information that supports the client's balance and some tests over nonfinancial information that is inconsistent with the client's balance. We seed deficiencies in all conditions, as the audit workpaper preparer (hereafter, "preparer") concludes that the sales balance is reasonable without adequately considering inconsistent evidence (i.e., a conclusion deficiency). The dependent measure is reviewers' assessment of the sufficiency of evidence to support the client's reported sales balance. In both experiments, we manipulate regulatory focus by priming participants with either a promotion focus or a prevention focus (Freitas and Higgins 2002). In Experiment 2, we also manipulate review structure by having some reviewers use segmented review, assessing one discrete test at a time with distractor tasks in between, and having the others use unsegmented review, assessing the portfolio of tests all at once.

Our findings support the prediction that a promotion focus improves reviewers' holistic evidence evaluation and detection of conclusion deficiencies. In Experiments 1 and 2, we find that promotion focused reviewers rate the sufficiency of evidence to support the client's sales balance lower than prevention focused reviewers. A supplemental analysis shows that promotion, as opposed to prevention, focused reviewers who give lower evidence sufficiency ratings to *individual tests* with inconsistent evidence also give lower evidence sufficiency ratings to the *holistic test* of the client's sales balance. A path model provides evidence consistent with

the predicted underlying judgment processes. That is, promotion focused reviewers think more holistically and thus better recognize evidence as inconsistent, leading these reviewers to evaluate the audit evidence over the sales balance as less sufficient than prevention focused reviewers. Finally, we find that promotion focused reviewers are more likely than prevention focused reviewers to communicate inconsistent evidence to their supervisor, which is a key engagement quality control. Contrary to our prediction, we find no main or interactive effects of segmented review. Supplemental tests suggest this is because segmented review may only inhibit the most experienced reviewers in our study.

Our study contributes to both practice and theory on the review process—in particular, the potential for the review process to improve auditors’ consideration of inconsistent evidence. Preparers struggle to think holistically about inconsistent evidence (Griffith et al. 2015b; Kadous and Zhou 2016), which increases audit risk if reviewers do not detect issues related to inconsistent evidence. We find that a promotion focus helps reviewers more effectively detect and communicate such deficiencies to their supervisors. This suggests that a promotion focus may be an important, relatively easy-to-implement intervention to help reviewers adopt the type of holistic thinking that has long been viewed as critical to high quality reviewer judgment (Moeckell 1990; Gibbins and Trotman 2002).

We also provide evidence that a promotion focus benefits the manner in which reviewers process information, not just the final evidence sufficiency ratings they provide. Promotion focused reviewers engage in more holistic thinking and give greater weight to inconsistent evidence. Identifying this causal chain is important for identifying and validating weaknesses and potential improvements in how audit teams work with inconsistent evidence.

Our findings may be helpful to scholars, regulators, and practitioners focused on improving audit firms' systems of quality control (e.g., Johnson, Keune, and Winchel 2019; PCAOB 2019). Our theory and findings suggest that it is likely beneficial to identify features of the auditing context that trigger a prevention focus in reviewers (e.g., tight budgets, step-by-step audit programs). Because prevention focused reviewers particularly struggle to identify problems in audit test work over inconsistent evidence, this type of examination would refine our understanding of conditions in which the review process is most deficient.

Finally, our findings have additional important implications for research and practice. Researchers spend considerable time investigating high-risk audit areas like complex estimates (e.g., Austin, Hammersley, and Ricci 2020; Gold, Kadous, and Leiby 2020) and accounts with nonfinancial inputs (e.g., Brazel, Leiby, and Schaeffer 2020; Brazel and Schmidt 2019). Our findings are of greatest importance in these types of high-risk audit areas, which require a broad focus on interdependencies among evidence items. Further, the academy and practitioners are increasingly focusing on accounting data analytics (e.g., Austin, Carpenter, Christ, and Nielson 2020; Emmett, Kaplan, Mauldin, and Pickerd 2020; EY 2015). The potential role of a promotion focus is likely to increase in importance as audit firms more broadly adopt data analytics tools, which auditors use to process a greater volume of information, and thus inconsistent information, than they process with traditional audit tools.

II. THEORY AND HYPOTHESIS DEVELOPMENT

Supervisors perform audit workpaper review as a quality control mechanism to detect and correct preparer deficiencies and client misstatements (Trotman 1985; Trotman and Yetton 1985; Libby and Trotman 1993; Rich, Solomon, and Trotman 1997). Financial statement items are

increasingly complex due to evolving accounting standards as well as the increased use of estimates and fair values (Barth 2006). As a result, auditors must assess client assertions using evidence that is increasingly complex and often inconsistent with client assertions, and auditors often struggle with this complexity (Bell and Griffin 2012; Griffith et al. 2015a). It is critical for auditors to consider inconsistent evidence, as such evidence may necessitate revised risk assessments and audit procedures (PCAOB 2011).

This complexity presents a challenge for audit reviewers, who must address not only *testing deficiencies* in which preparers do not conduct adequate procedures, but also *conclusion deficiencies* in which preparers do not draw appropriate inferences about the evidence set as a whole.¹ Conclusion deficiencies are particularly concerning when the evidence set contains inconsistent evidence. For example, a PCAOB Part II finding identified three engagements on which ineffective reviews were a primary reason that auditors did not consider inconsistent evidence appropriately and prematurely concluded that clients' intangibles were reasonable (PCAOB 2014).² On one engagement, evidence existed to support the estimate; however, the report argued that auditors did not collect sufficient evidence to evaluate the inconsistency between the increase in projected revenue growth in the estimate's model and the issuer's recent decline in revenues (PCAOB 2014). That is, auditors tested these items individually but did not adequately consider how these inconsistent items affected the balance as a whole.³

¹ By nature, conclusion deficiencies are an important subset of conceptual deficiencies (Ramsay 1994). Conclusion deficiencies are not only subjective and unverifiable—attributes Ramsay observes are common to all conceptual deficiencies—but also require holistic evaluation to identify.

² PCAOB inspection reports have a public portion, referred to as Part I, and a nonpublic portion, referred to as Part II. Part II contains information about the inspected audit firm's quality control system. The PCAOB publicly releases the Part II portion only when the inspected audit firm fails to satisfactorily remediate identified deficiencies within one year.

³ Bamber and Ramsay (1997; 2000) find that review quality decreases when more senior and more junior reviewers specialize in detecting different subsets of deficiencies, rather than both more senior and more junior reviewers aiming to identify all types of deficiencies. These authors note that this effect may arise because specialization disrupts integrated, big picture thinking.

We argue that audit reviewers' identification of conclusion deficiencies is likely to improve when reviewers have a promotion, as opposed to a prevention, focus. Regulatory focus theory argues that these foci affect how people perceive information and pursue their goals (Higgins 1997; 1998). In brief, a prevention focus is defensive and emphasizes avoiding mistakes, whereas a promotion focus emphasizes reaching the best possible outcome (Higgins 1997; Shah, Higgins, and Friedman 1998). Whether a reviewer adopts a prevention or promotion focus likely depends on pressures in the auditing setting. For example, audit firms confront pervasive pressures to reduce fees and meet aggressive filing deadlines (Agoglia, Hatfield, and Lambert 2015; Lambert, Jones, Brazel, and Showalter 2017), which leads firms to set tight time budgets and adopt highly structured audit programs (Hammersley and Ricci 2020). We expect time pressure and task rigidity to make reviewers more prevention focused.⁴

When these pressures result in reviewers adopting a prevention focus, it likely worsens their review of complex evidence sets containing inconsistent information. Specifically, a prevention focus narrows people's attention such that they are susceptible to selective attention (Memmert, Unkelbach, and Ganns 2010).⁵ Selective attention occurs when people unconsciously fail to attend to information—sometimes even obvious information—that is not relevant to their

⁴ We do not argue that a prevention focus is necessarily bad, because it can help people meet explicit performance standards and minimize distractions (Wallace et al. 2009; 2010; Knechel and Leiby 2016). Indeed, we have conducted an initial experiment to examine whether a prevention focus *improves* the detection and correction of deficiencies such as missing procedures, procedures not conducted according to policy, or procedures that do not conform to explicit standards of due care. In an experiment with professional auditors ($n = 36$), we find that reviewers with a prevention focus are superior to those with a promotion focus at detecting this type of deficiency. Moreover, Liang et al. (2007) find that Master's students in China identify a greater number of proofreading errors in the financial statements (e.g., misspellings, subject-verb agreement problems) when they are prevention focused. Our study focuses on a different construct from Liang et al. (2007); namely, we focus on conclusion deficiencies.

⁵ A famous example is the “invisible gorilla task” in which people watch a video of two groups passing basketballs in a circle and are instructed to count the number of times one of the groups passes the ball. During the video, a person in a gorilla costume walks conspicuously through the middle of the groups, yet half of people fail to notice! See “The Monkey Business Illusion” at https://youtu.be/IGQmdoK_ZfY. This video illustrates how people become “blind” to seemingly obvious information that is not directly relevant to their goals. Even when people become aware that they have missed this information, they subsequently fail to notice other obvious information.

current goal (Simons and Chabris 1999; Horstmann and Ansorge 2016). While promotion focused people are more likely to notice abstract relationships among seemingly unrelated items, prevention focused people are more likely to attend to the specific attributes of an item, independent of other items (Zhu and Meyers-Levy 2007). This likely negatively affects reviewers' consideration of inconsistent evidence, as reviewers may carefully evaluate preparers' tests of these items but not these tests' holistic implications. For example, in a test of a sales balance, a sales growth trend and a production capacity trend may each individually be supported by evidence, but the link between trends must be reasonable in order to conclude that the sales balance is reasonable. Selective attention inhibits drawing these types of connections.

By contrast, theory and evidence suggests that a promotion focus may help reviewers detect conclusion deficiencies. In organizations, a promotion focus helps people reconcile conflicting information and focus more on information's substance than on its stylistic elements (Graham et al. 2015). Promotion, as opposed to prevention, focused people are more comfortable with large information sets, exercise greater resilience in completing complex tasks, and think more abstractly and creatively (Polman 2012; Shah, Higgins, and Friedman 1998; Friedman and Förster 2005). Identifying conclusion deficiencies requires understanding the connections among evidence, a holistic view that is likely to benefit from a promotion focus. Thus, we predict that reviewers with a promotion focus will better detect preparers' conclusion deficiencies. Stated formally:

H1: Reviewers with a promotion focus, as opposed to a prevention focus, will better detect conclusion deficiencies in preparers' consideration of inconsistent evidence.

Review Task Structure

We also expect regulatory focus to affect how reviewers respond to an emerging change in how review tasks are structured in practice. Specifically, segmented review occurs when the audit support system subdivides workpaper review for an audit section into several smaller segments (e.g., discrete “sign offs” in Dowling and Leech 2014). Reviewers assess and approve each completed segment on its own rather than assessing all completed segments together as one task. This approach intends to encourage more frequent and timely reviewer/preparer interaction.

However, segmented review likely also inhibits reviewers’ identification of conclusion deficiencies. Segmented review frames review as a collection of individual segments rather than a cohesive whole. Individuals who process information sets in segments struggle to understand the relations among segments (Driver and Baylis 1998; Driver et al. 2001). Moreover, segmented review likely causes reviewers to define their information processing goals more narrowly, which worsens selective attention by causing a larger set of information to seem irrelevant to those goals. Indeed, attentional goals are the cognitive mechanism underlying selective attention. For instance, as compared to people with a goal to count how many times black shapes “jump” on a screen, people focused on white shapes are far less likely to attend to a black shape that unexpectedly enters the screen, since this new shape is irrelevant to their goal (Most, Scholl, Clifford, and Simons 2005). Since segmented review narrows reviewers’ attentional goals, it likely undermines the integrated, holistic thinking that benefits reviewers when considering inconsistent evidence and improves their detection of conclusion deficiencies.⁶

⁶ Bonner, Libby, and Nelson (1996) observe that in practice, audit tasks are commonly decomposed by breaking a task into steps using a checklist or audit plan. These authors find that when audit risk assessment tasks are decomposed, mechanical formulas aggregate holistic risk more accurately than auditors judge holistic risk. Zimbelman (1997) finds that this type of task decomposition can result in auditors paying more attention to decomposed steps, such that auditors spend more time reading fraud red flags and budget more time for fraud detection when fraud risk assessment tasks are decomposed, rather than holistic. Raiffa (1968) argues that the advantage of decomposition is to “decompose a complex problem into simpler problems” (271). In our study, when reviewers use either a segmented or unsegmented review structure, they review six individual tests. That is, in both structures, our review task is decomposed into simpler problems. We do so in order to hold the level of

We expect the negative effects of segmented review to be conditional on whether reviewers have a promotion or prevention focus. Specifically, if reviewers with a promotion focus are less susceptible to selective attention, then we expect them to be less negatively affected by segmented review than those with a prevention focus. This idea is consistent with evidence in psychology that promotion focused people prefer to interpret information holistically, as opposed to piecemeal, and perform better than prevention focused people at tasks that require a holistic approach (Higgins, Shah, and Friedman 1997; Friedman and Förster 2001; Förster and Higgins 2005; Kuschel, Förster, and Denzler 2010; Lee, Choi, and Li 2014; Förster, Friedman, Özelsel, and Denzler 2006).

In sum, when using a segmented review structure, promotion focused reviewers are more likely than prevention focused reviewers to recognize the implications of inconsistent evidence on the sufficiency of the evidence set for a client assertion. As a result, promotion focused reviewers are more likely to identify conclusion deficiencies when preparers inadequately consider inconsistent evidence. Therefore, we expect that reviewers will struggle to detect conclusion deficiencies when they use segmented, as opposed to unsegmented, review, and that a promotion focus will attenuate this effect. Stated formally:

H2: Reviewers will better detect conclusion deficiencies in preparers' consideration of inconsistent evidence when the review task is unsegmented, as opposed to segmented, but this difference will be smaller when reviewers have a promotion, as opposed to prevention, focus.

III. RESEARCH METHOD

We conduct two experiments to investigate the effect of regulatory focus on reviewers' identification of conclusion deficiencies. As discussed below, the task content is nearly identical

decomposition constant across review structures. Rather, we vary review structure by altering the extent to which reviewers perceive their decomposed review task holistically, i.e., as part of a larger whole.

across the two experiments. Experiment 1 is a 1 x 2 (regulatory focus: prevention, promotion) design that we use to test H1.⁷ Experiment 2 is a 2 (regulatory focus: prevention, promotion) X 2 (review structure: segmented, unsegmented) design that we use to test H2. We administered both experiments online using Qualtrics.

Participants

For both experiments, we focused our recruiting on audit seniors because they are commonly tasked with reviewing audit workpapers (Rich et al. 1997). For Experiment 1, 49 auditors participated from one office of a Big Four firm. The office managing partner invited all seniors and recently-promoted managers to a one-hour conference call. On the call, the partner introduced one author-team member, who described the research study in broad terms and emailed a study link to call participants. The partner expressed support for the study and asked everyone to spend the rest of the hour completing the study. For Experiment 2, 94 auditors participated from seven public accounting firms, including Big Four firms. We received recruiting assistance from the Center for Audit Quality. Most auditors received an invitation to participate in our online study via email, though auditors from one firm participated during their national training sessions, for which one author was present.

We screen auditors to ensure that they both had the necessary experience to interpret the materials and exerted sufficient effort on the task.⁸ We use the same screening criteria for both studies. First, we exclude all auditors with less than two years of audit experience, as these auditors are unlikely to have experience reviewing subordinates' work. This excludes zero

⁷ Experiment 1 also includes a third condition in which prevention-focused reviewers view seeded testing deficiencies, i.e., documentation omits some of the preparer's test work. We include this condition because we conducted an initial experiment in which prevention focused reviewers were relatively adept at detecting this type of objectively poor quality work, even when they miss conclusion deficiencies. We find no differences across prevention conditions and thus collapse across prevention conditions.

⁸ Dropout rates, i.e., individuals who clicked the link to the experiment but exited before viewing our manipulations, were 5% (3 of 59) in Experiment 1 and 15% (26 of 178) in Experiment 2.

auditors in Experiment 1 and 21 auditors in Experiment 2. Second, we screen for insufficient effort by excluding auditors who fulfill two or more of the following criteria: (1) skip an important open-ended dependent measure, (2) spend over five hours on the study, which likely indicates quitting in the middle (median duration of included auditors is 40 minutes), and (3) skip the regulatory focus reminder (see the independent variable section for a description). This excludes seven auditors in Experiment 1 and 37 auditors in Experiment 2. In total, we exclude seven auditors in Experiment 1 and 58 auditors in Experiment 2.

The auditors in our experiment are primarily seniors (90%, managers 9%, staff 1%). All have at least 24 months of audit experience, and on average, participants have three to four years of experience; 83% are Certified Public Accountants (90% in Experiment 1, 80% in Experiment 2); and 68% have used nonfinancial measures in their substantive testing (75% in Experiment 1, 65% in Experiment 2). On average, 10-25% of our participants' chargeable audit hours have been spent on manufacturing clients in the preceding three years. There are no significant differences between auditors in Experiments 1 and 2 for any of these demographic variables (all p -values > 0.1). Further, there are no significant differences between auditors in Experiments 1 and 2 for either their reported effort on the task ($t_{141} = 1.06$, $p = 0.29$, two-tailed) or their reported difficulty of the task ($t_{141} = 0.06$, $p = 0.95$, two-tailed). See Table 1 for demographic information on auditors in our first and second experiments.

Insert Table 1 here

Experimental Task (Experiments 1 & 2)

The task asked auditors to review a substantive analytical procedure over sales for an operating unit of a publicly traded manufacturing client. The client had no previous restatements and the audit revealed no control issues. To emphasize time pressures that would exist in reality,

the case informed auditors that additional testing of sales would likely make them go over budget and that client management would react negatively to unwarranted additional tests. In all conditions, the preparer concluded that the sales balance was reasonable and did not warrant further investigation.

Auditors first viewed a workplan of six tests completed by the preparer and supporting documentation for each of the tests. We held the order of tests constant, and the general evidence pattern showed the operating unit reporting positive sales growth. The first five tests each focused on a different measure or ratio that could influence the expected sales balance, and the sixth test was the holistic substantive analytic over the sales balance. The first three tests focused on financial measures and ratios, all of which were *positive* and support sales. The fourth and fifth tests focused on nonfinancial measures that were *negative*, inconsistent with reported sales.⁹

For each of the first five tests, the preparer conducted client inquiries and concluded that the test produced sufficient evidence that the change examined was reasonable and should not change the overall expectation for sales. However, the preparer rarely described diagnostic tests to corroborate client representations. Immediately after reviewing the preparers' evidence for each of these individual procedures, the audit reviewers evaluated the sufficiency of evidence for the tests.

⁹ The tests were (1) a three-year trend analysis of operating unit sales, (2) comparing sales growth to the industry average, (3) a budget-to-actual sales comparison, (4) an analysis of year-over-year changes in product offerings, (5) an analysis of year-over-year changes in production space, and (6) the overall substantive analytic. We held order constant because varying order would have limited the types of analyses available to us. For example, if we had varied order, participants would have identified inconsistencies between evidence after viewing different tests. Those viewing financial tests that support the client's proposed sales balance first would have identified the inconsistency after viewing the subsequent nonfinancial, inconsistent tests. Those viewing nonfinancial tests first would have identified the inconsistency after viewing the subsequent financial tests. The point at which participants realized the seeded inconsistency could have affected their evidence sufficiency judgments for the individual tests, which we use in our analyses.

In the sixth test, the preparer developed a holistic substantive analytic over the sales balance based on the prior year's balance, with an adjustment for inflation. Given the evidence from the prior five tests, this method of evaluating sales ignores important changes in sales during the year. The preparer concluded that the sales balance was reasonable in all conditions. Immediately after reviewing the preparers' evidence for this holistic procedure over the sales balance, the audit reviewers evaluated the sufficiency of evidence for this test.

Regulatory Focus Manipulation (Experiments 1 and 2)

Prior to reviewing the analytic, auditors were randomly assigned to complete one of two primes that we used to vary regulatory focus. Following prior literature, the promotion condition instructed auditors "In three to four sentences, write a brief description of your current aspirations, ideals, and hopes." The prevention condition instructed auditors "In three to four sentences, write a brief description of your current obligations, duties, and responsibilities." To increase the longevity of the prime, we included a reminder of the manipulation after auditors reviewed the first five tests and before they reviewed the holistic substantive analytical procedure. We asked promotion and prevention condition participants to list three of their hopes and aspirations or duties and obligations, respectively.

Review Structure Manipulation (Experiment 2)

We manipulated review structure at two levels between-participants, as segmented and unsegmented. Auditors in the segmented review condition viewed a workplan with a sign-off required for each individual test. They then reviewed the preparer's discrete workpapers in segments, one substantive procedure at a time. After reviewing each segment, auditors worked for 30 seconds decoding as many four-letter anagrams as possible before reviewing the next

segment.¹⁰ Those in the unsegmented review condition viewed a workplan with one sign-off at the completion of all the review tasks. These auditors received the portfolio of workpapers (i.e., all six segments) and reviewed it as one task.

Dependent Measures (Experiments 1 and 2)

Our primary dependent measure is the sufficiency of evidence to support the client's sales balance. We measure this as auditors' assessments of the sufficiency of evidence for the holistic substantive analytic procedure over the sales balance (i.e., the sixth procedure). Auditors responded to the question "In your judgment, is the documented audit evidence sufficient to support [the client's sales]?" on an 11-point Likert scale anchored by 0 = "Definitely no" and 10 = "Definitely yes." Theory suggests this measure should cleanly reflect differences across regulatory focus conditions, because it captures how well the individual pieces of evidence form a cohesive whole, i.e., forest for the trees. Lower values indicate lower sufficiency.

IV. RESULTS

Experiment 1 Results

Manipulation Checks

For regulatory focus, all of our included auditors followed the instructions to write about their hopes and aspirations (promotion focus) or obligations (prevention focus). We also included a reminder for our auditors during the task to re-write three attributes that they listed in response to the regulatory focus manipulation. In total, 45 of 49 auditors (92%) wrote all three attributes, and this percentage did not significantly differ between conditions. This suggests that auditors attended to the manipulation.

¹⁰ We included the 30 second delay and anagram task to create a brief time lag between the review of individual segments in order to clear auditors' short-term memories. Anagram-like tasks or Sudoku puzzles are commonly used as distractor tasks (see Barrett and Kensinger 2010).

Tests of H1

H1 predicts that reviewers with a promotion, as opposed to a prevention, focus will better detect conclusion deficiencies in preparers' consideration of inconsistent evidence. In experiment 1, sufficiency of evidence to support the client's sales balance ratings are significantly lower in the promotion focus condition (4.65 vs. 3.17, $t_{47} = 1.96$, $p = 0.03$, one-tailed), as shown in Figure 1, Panel A and Table 2, Panels A and B. For robustness, we also test a broader proxy of auditors' overall evaluation of the preparer's work by extracting a principal component from ratings of evidence sufficiency, reasonableness of the client's sales balance, the preparer's performance, and the appropriateness of how the preparer constructed their expectation.¹¹ The overall evaluation component explains 83% of shared variance across these measures. Figure 1, Panel B shows that reviewers with a promotion focus have lower scores on this overall evaluation measure as expected (-0.41 vs. 0.24, $t_{47} = 2.25$, $p = 0.01$, one-tailed). Thus, H1 is supported.

Insert Table 2 here

Insert Figure 1 here

We expect that reviewers with a promotion focus will more strongly link their evidence sufficiency assessments for individual, inconsistent evidence items to their holistic evidence sufficiency assessment for the sales balance, as compared to those with a prevention focus. Sufficiency ratings for the inconsistent items capture whether reviewers believe the auditor's test work has provided enough evidence that the inconsistency does not reflect heightened

¹¹ Promotion focused reviewers are lower on all four individual measures, as well. Auditors rate these measures on 11-point Likert scales: "Please assess the reasonableness of the sporting good sales balance" with anchors of 0 = "Not at all reasonable" and 10 = "Extremely reasonable," "The preparer concluded that the sporting good sales balance was reasonable. How would you assess the preparer's performance?" with anchors of 0 = "Extremely low quality" and 10 = "Extremely high quality," and "The preparer used prior year sporting goods sales and an inflation adjustment as the bases for the current year sporting good sales expectation. In your judgment, is this expectation reasonable for developing an estimate of current year sporting good sales?" with anchors of 0 = "Definitely, no" and 10 = "Definitely, yes."

misstatement risk. Lower sufficiency ratings for the inconsistent evidence should translate into lower holistic sufficiency ratings for the sales balance. To test this, we run a General Linear Model that crosses regulatory focus with reviewers' average sufficiency evaluation rating for the two inconsistent, nonfinancial evidence items. Following Aiken, West and Reno's (1991) guidance, we mean center the average ratings in order to reduce the risk of multicollinearity.¹² The dependent measure is reviewers' holistic sufficiency ratings for the sales balance.

Consistent with our theory, there is a significantly positive interaction between regulatory focus and the average sufficiency rating for inconsistent evidence ($t_{48} = 2.11$, $p = 0.04$, two-tailed). We probe this interaction using a simple slopes approach that estimates the effect of regulatory focus at one standard deviation above and one standard deviation below the mean inconsistent evidence sufficiency rating (Aiken et al. 1991). This probe allows us to estimate the effect of regulatory focus on holistic evidence sufficiency evaluations when participants rated sufficiency for inconsistent evidence as high versus when they rated sufficiency for inconsistent evidence as low. That is, this analysis tests the effect of regulatory focus conditional on reviewers' ratings of the evidential support for inconsistent evidence.

As shown in Figure 1, Panel C, when giving low sufficiency ratings for inconsistent evidence, promotion-focused reviewers assess lower holistic sufficiency ratings for the sales balance than prevention-focused reviewers ($\beta = -1.98$, $t_{48} = -2.51$, $p = 0.02$, two-tailed). When giving high sufficiency ratings for inconsistent evidence, there is no effect of regulatory focus. In other words, when reviewers note insufficient work for inconsistent evidence, reviewers are

¹² We compute each participant's mean inconsistent evidence item rating and the mean inconsistent evidence item rating for the entire sample, then subtract the sample mean from each participant's mean inconsistent evidence item rating.

more likely to interpret this as a negative signal for holistic evidence sufficiency if they have a promotion focus than if they have a prevention focus. This supports our theory.

Audit Quality Implications

Communicating significant issues to a manager is an important engagement quality control (Lightle et al. 2017). Accordingly, we test whether a promotion focus increases reviewers' willingness to communicate significant issues to their supervisor. Immediately after evaluating evidence sufficiency for the holistic sales balance test, we instructed auditors to write the concerns that they would share with their manager. We code auditors' responses to this open-ended question into two categories for the communication variable:¹³ (1) no concerns or only concerns about evidence other than inconsistent evidence, and (2) concerns about the seeded inconsistencies between financial and nonfinancial measures.^{14, 15}

Items in the first coding category include concerns about documentation and non-specific concerns that the preparer had not performed sufficient testing. For example, two items in this first category are, "No concerns, just to add a bit of additional documentation," and "... Based on what I see, we will need to follow-up with Management to understand the nature of the sales amounts, I understand that Madison is sensitive to this information but we need a deeper understanding to ensure our substantive analytic makes sense." Items in the second category

¹³ An author and two research assistants coded participants' responses while blind to experimental condition; the research assistants were also blind to our research question. The author coded participants' responses from both studies, while the research assistants coded responses for only one study each. For both studies, the author and research assistant reconciled any disagreements, jointly agreeing on the final coding that we use in our analyses.

¹⁴ Fourteen participants in Experiment 1 and eight participants in Experiment 2 did not write anything in this open-ended question. We coded these participants in category 1 rather than coding the data as missing because in each of these cases the participant answered all three of the previous questions in the experiment as well as the subsequent question in the experiment. This pattern of behavior indicates that the participants had no concerns, not that the participant skipped the question without reason.

¹⁵ For Experiment 1, the coders initially agreed at a rate of 87.8%, with a Cohen's Kappa of 0.75 ($p < 0.01$). For Experiment 2, the coders initially agreed at a rate of 97.3%, with a Cohen's Kappa of 0.94 ($p < 0.01$). These Cohen's Kappas suggest substantial agreement between the coders in both Experiment 1 and Experiment 2.

either note inconsistent evidence concerns explicitly or implicitly link the seeded inconsistencies to sales. For example, two items in this second category are, “We have encountered contradictory evidence during our testing of the [sales balance]...,” and “...based on the support that we have received so far, it appears that the evidence is contradictory and we would expect sales decreases. However, not only are sales not decreasing, they are increasing at a large rate....”

As shown in Figure 2, Panel C, as well as Table 3, Panels A and B, our results support the idea that promotion focus increases the communication of significant issues to supervisors. We run a Binomial Logit regression because the analysis uses a dichotomous dependent measure, and parameter estimates indicate that communication is greater in the promotion, as opposed to prevention, focus condition (0.96 vs. -0.19, $\chi^2_1 = 3.46$, two-tailed $p = 0.06$). Thus, a promotion focus not only improves reviewers’ detection of conclusion deficiencies but also increases their willingness to bring significant, inconsistent evidence to the attention of their supervisors, increasing the chances that the audit team will, for example, resolve the issue with the client or propose an adjustment.

Insert Figure 2 here

Insert Table 3 here

Test of Theorized Judgment Process

Our theory predicts that promotion focus causes reviewers to think holistically and, thus, draw connections from inconsistent evidence items to the holistic evidence set. In turn, this likely leads to lower assessments of evidence sufficiency for the holistic sales balance test and then more issues communicated to managers. As shown in Figure 3, we use PROCESS to test this path (Hayes 2013). To proxy for holistic thinking, we use a coded variable that captures whether

participants' review notes on the overall substantive analytic focus on the implications of the inconsistent evidence about sporting goods revenue (2), on the implications of the audit preparer's substantive analytic test more narrowly (0), or in between (1).¹⁶ To measure extent of reliance on inconsistent evidence, we use a coded variable, inconsistent evidence, that captures the number of times the participant describes information as inconsistent with or contradicting sales in his or her review notes.¹⁷ We use our primary dependent measure to capture how participants rate the sufficiency of evidence for the holistic test over the client's sales balance. We use the communication variable to measure issues communicated to managers.

Our results show that promotion focus increases reviewers' tendency to think holistically (link 1 = 0.41, $t = 1.64$, $df = 43$, one-tailed $p = 0.05$), which leads them to identify more inconsistent evidence (link 2 = 1.02, $t = 4.30$, $df = 43$, one-tailed $p < 0.01$). Further, identification of more inconsistent evidence leads to lower assessments of evidence sufficiency for the aggressive sales account balance (link 3 = -0.56, $t = -1.89$, $df = 43$, one-tailed $p = 0.03$). Finally, as expected, evidence sufficiency is inversely related to upward communication of issues within the audit team (link 4 = -0.23, $z = -1.36$, $df = 43$, one-tailed $p = 0.09$). That is, promotion focused reviewers think more holistically, which enables them to better understand the implications of inconsistent evidence, ultimately leading to lower assessments of the sufficiency of evidence and greater upward communication of audit issues.¹⁸ These results illustrate the judgment process

¹⁶ Items are coded as too vague if they are not specific enough to fit into category 0, 1, or 2 and are thus excluded from this analysis. The coders initially agreed at a rate of 70.5%, with a Cohen's Kappa of 0.57 ($p < 0.01$). This Cohen's Kappa suggests agreement between the coders.

¹⁷ The coders initially agreed at a rate of 87.7%, with a Cohen's Kappa of 0.74 ($p < 0.01$). This Cohen's Kappa suggests substantial agreement between the coders.

¹⁸ For observable variable models like ours (i.e., those with no latent variables), differences between PROCESS models and Structural Equation Models (SEM) are generally trivial (Hayes, Montoya, and Rockwood 2017). However, SEM is a large sample technique (e.g., 100+ participants per condition) because it relies on maximum likelihood estimation, whereas PROCESS uses ordinary least squares regression (OLS) and is, thus, more appropriate for smaller samples like ours. Unlike for SEM, there are no model fit statistics for PROCESS models (Hayes et al. 2017).

underlying H1 and suggest that promotion focus improves the detection and communication of inconsistent evidence by improving *how* reviewers think about the holistic evidence set, i.e., conclusion deficiencies.

Insert Figure 3 here

Experiment 2 Results

Manipulation Checks

For regulatory focus, all of our included auditors followed the instructions to write about their hopes and aspirations (promotion focus) or obligations (prevention focus). We also included a reminder for our auditors during the task to re-write three attributes that they listed in response to the regulatory focus manipulation. In total, 97 of 109 (89%) auditors wrote all three attributes and this percentage did not significantly differ between conditions. This suggests that included auditors attended to the manipulation. For review structure, auditors in the segmented review condition experienced a 30 second delay between reviewing each segment, and we asked them the complete four-word anagrams during this delay. We measure the number of anagrams that auditors completed in order to measure the degree of additional distraction between reviews of individual tests. On average, participants completed 14.7 anagrams, and a difference contrast indicates that this number increased as the task went along, consistent with participants exerting effort and attending to the manipulation's instructions.

Replication of H1

The Experiment 2 data are consistent with H1 (see Figure 1, Panel A), and we follow a similar plan for data analysis as we did for Experiment 1. In Table 2, Panel D, we run a 2 (regulatory focus: promotion, prevention) x 2 (review structure: segmented, unsegmented) ANOVA with sufficiency of evidence as the dependent variable. Then, in Panel E, we run a

pairwise comparison across regulatory focus conditions and find that auditors in the promotion condition assessed evidence sufficiency lower than those in the prevention condition (4.45 vs. 3.68, $t_{92} = 1.56$, $p = 0.06$, one-tailed).¹⁹ We also examine the overall evaluation principal component, which explains 75% of shared variance across Experiment 2's measures. However, reviewers with a promotion focus (mean = -0.11) do not evaluate evidence significantly lower than those with a prevention focus using this measure (0.08, $t_{91} = 0.91$, $p = 0.18$, one-tailed); see Figure 1, Panel B.

We also run a GLM testing the link between reviewers' average evidence sufficiency ratings for the individual, inconsistent evidence items and evidence sufficiency for the holistic sales balance test. As in Experiment 1, we find that promotion focus significantly decreases holistic evidence sufficiency assessments ($\beta = -0.79$, $t_{89} = -1.82$, $p = 0.04$, one-tailed). Though the interaction between regulatory focus and average sufficiency rating for inconsistent evidence is not statistically significant ($p = 0.24$, two-tailed), we find a similar pattern to that in Experiment 1. That is, promotion focus decreases holistic evidence sufficiency ratings when reviewers rate evidence sufficiency for inconsistent evidence as low ($\beta = -1.31$, $t_{89} = -2.23$, $p = 0.03$, two-tailed), and there is no effect when inconsistent evidence sufficiency ratings are high. However, unlike Experiment 1, we find no evidence in Experiment 2 of differences in communication of audit issues to supervisors (see Table 3, Panel D).

Tests of H2

H2 predicts that a segmented review structure will decrease reviewers' detection of conclusion deficiencies in preparers' consideration of inconsistent evidence, but this effect will be weaker for reviewers in a promotion, as opposed to a prevention, focus. To test H2, we run a

¹⁹ Note that our analysis in Panel E collapses over review structure, as we find no effect of review structure.

2 (regulatory focus: promotion, prevention) x 2 (review structure: segmented, unsegmented) ANOVA with sufficiency of evidence as the dependent variable. As shown in Table 2, Panel D, we do not observe the significant negative effect of review structure that we anticipated ($F = 0.23$, $p = 0.64$, two-tailed). As such, we would not expect to see our predicted review structure by regulatory focus interaction. Consistent with this idea, we do not find a significant review structure by regulatory focus interaction ($F = 0.45$, $p = 0.50$, two-tailed), indicating the data do not support H2.

Additional Analysis

We conduct supplemental tests to investigate our surprising finding in Experiment 2 that segmented review structure does not impair reviewers' evaluation of evidence sufficiency. These supplemental tests suggest that review structure affects more experienced reviewers more than less experienced reviewers, potentially explaining the lack of support for H2. Note that 92 percent of study reviewers are seniors, for whom this review is an appropriate task, thus these differences demonstrate within rank experience effects. Researchers document many ways in which experience influences auditors' performance of analytical procedures tasks (e.g., Libby and Frederick 1990; Kaplan et al. 1992; Shelton 1999). Knowledge and work habits, or routines, develop from auditors' experience (Libby and Luft 1993; Solomon et al. 1999). Related to our study, more experienced reviewers have primarily gained their experience using traditional, unsegmented review. Thus, these more experienced reviewers' review process routines are likely more ingrained, resulting in segmented review structure being most disruptive to these reviewers. By contrast, less experienced reviewers are still developing review process routines, so modifying review structure may have had a less significant effect on their judgments. Further, there are typically two levels of audit review. The first is conducted by less experienced

reviewers and is often intended to be a very detailed review, whereas the second is conducted by more experienced reviewers and is often intended to be a more holistic review (Bacsik and Rizzo 1983). If more experienced reviewers aim to conduct this latter type of holistic review, segmented review may only impair the review efforts of these more experienced reviewers. In other words, our less experienced reviewers may not have negative effects from segmented review to mitigate.

To examine this, we first compute the mean evidence sufficiency, holistic thinking, inconsistent evidence, communication of significant issues, and overall evaluation principal component ratings for each of three experience levels: 24 to 35 months, 36 to 47 months, and over 48 months in Table 4, Panels A to E. As shown in Table 4, Panels B and C, auditors in Experiment 2 with over 48 months of experience exhibited more holistic thinking and identified more inconsistent evidence in the unsegmented, as opposed to segmented, review condition. Accordingly, auditors at this experience level also evaluated evidence sufficiency for the holistic sales balance test lower in the unsegmented condition than in the segmented condition (Panel A). Further, these most experienced auditors communicated more significant issues to their supervisor in the unsegmented condition than in the segmented condition (Panel D). We do not find this same pattern of results for less experienced auditors.

Insert Table 4 here

For inferential tests in Experiment 2 (untabulated), we first examine the relation between experience and review structure using an ordinal regression with holistic thinking as the dependent measure, experience as a continuous predictor, and review structure as a binary predictor. The results reveal that holistic thinking is marginally higher when reviews are unsegmented ($\beta = -2.55$, Wald Chi Square = 3.42, $p = 0.07$, two-tailed). There is also a

significantly positive experience by review structure interaction suggesting that this difference is greater when reviewers have more experience ($\beta = 0.06$, Wald Chi Square = 4.11, $p = 0.04$, two-tailed), consistent with the univariate results above.

Next, we examine how experience is related to these reviewers' identification of inconsistent evidence. We use a Poisson regression with a loglinear link, with inconsistent evidence as the dependent measure, experience as a continuous predictor, and review structure as a binary predictor. We observe a significant experience by structure interaction ($\beta = 0.04$, Wald Chi Square = 8.61, $p < 0.01$, two-tailed), indicating that experience is positively related to identifying inconsistent evidence in the unsegmented structure but not in the segmented structure.²⁰

Then, we examine how experience is related to reviewers' evaluation of evidence sufficiency for the holistic sales balance test. We use a GLM regression with evidence sufficiency as the dependent measure, experience as a continuous predictor and review structure as a binary predictor. We observe a negative association between experience and sufficiency ratings indicating that detecting conclusion deficiencies increases with experience ($\beta = -0.07$, $t_{86} = -2.21$, $p = 0.03$, two-tailed), though not a significant experience by structure interaction ($\beta = 0.03$, $t_{86} = 0.94$, $p = 0.35$, two-tailed). However, as in Experiment 1, we also extract an overall evaluation principal component from evidence sufficiency, reasonableness, preparer performance, and expectation appropriateness, which explains 74% of the shared variance. We use this principal component as the dependent measure in a GLM regression with experience as a continuous predictor and review structure as a binary predictor. We find a significant experience

²⁰ In our Experiment 2 analyses we find no interactions between segmented review and manufacturing industry experience or experience using nonfinancial measures. We also find no interaction between experience and regulatory focus. Suppressing the effect of review structure appears to be the only experience effect in our studies.

by review structure interaction ($\beta = -0.04$, $t_{86} = -2.39$, $p = 0.02$, two-tailed). That is, reviewers detect conclusion deficiencies as they gain experience but segmented review nullifies this experience effect.

Finally, we examine how experience is related to reviewers' propensity to communicate significant issues to their supervisors. We use a binomial regression, with communication as the dependent measure, experience as a continuous predictor, and review structure as a binary predictor. We observe a marginally significant experience by structure interaction ($\beta = 0.01$, Wald Chi Square = 2.82, $p = 0.09$, two-tailed), indicating that experience is positively related to identifying inconsistent evidence in the unsegmented structure but not in the segmented structure.

Based on these results, we examine the judgment process underlying H2, including experience as an explanatory variable. Specifically, we use PROCESS to evaluate whether promotion focus and non-segmented review structure interact with experience to improve more experienced reviewers' holistic thinking, identification of inconsistent evidence, evidence sufficiency evaluations, and finally, communication of significant issues to their supervisors. As shown in Figure 4, our results show that promotion focus, non-segmented review, and greater experience interact to increase reviewers' tendency to think holistically (interaction 1, $F = 1.95$, $df = 84$, one-tailed $p = 0.07$), which then leads reviewers to identify more inconsistent evidence (link 2 = 1.26, $t = 8.60$, $df = 89$, one-tailed $p < 0.01$). Further, identification of more inconsistent evidence leads to lower assessments of evidence sufficiency for the aggressive sales account balance (link 3 = -0.34, $t = -1.55$, $df = 89$, one-tailed $p = 0.06$). However, we do not find evidence in this model that evidence sufficiency is significantly, inversely related to communication of significant issues to the supervisor (link 4 = -0.04, $Z = -0.29$, $df = 89$, one-

tailed $p = 0.39$). Overall, these results illustrate that promotion focus interacts with non-segmented review and greater experience to improve the detection of inconsistent evidence by improving *how* more experienced reviewers think about conclusion deficiencies.

Insert Figure 4 here

These additional analyses are consistent with experience effects at least partly explaining the lack of support for H2. Consistent with our theory, segmented review decreases experienced reviewers' tendency to exhibit holistic thinking, identify inconsistent evidence, identify conclusion deficiencies, and communicate conclusion deficiencies to their supervisor. However, segmented review did not affect less experienced reviewers in these ways. Thus, it is possible that more review experience is necessary to observe the negative effects of review structure that we anticipated. That is, it is possible that H2 would be supported among highly experienced reviewers, such as those with four or more years of experience. Consistent with this idea, we find that promotion focus, non-segmented review, and greater review experience interact, leading to more holistic thinking, which then leads to improvements in identifying inconsistent evidence and evaluating evidence sufficiency. However, we draw no causal inferences from our analyses because a multitude of factors could explain a non-result in an experimental setting.

V. CONCLUSION

This study examines reviewers' detection of conclusion deficiencies, which are particularly concerning when they relate to preparers' consideration of inconsistent audit evidence. Conclusion deficiencies are the type of judgment errors in which auditors review individual evidence items but do not recognize that holistically, the items raise doubts about the reasonableness of a client assertion. We draw on regulatory focus theory to investigate how reviewers' promotion focus (i.e., a positive frame of mind that motivates thinking holistically) or

prevention focus (i.e., a defensive frame of mind that motivates avoiding mistakes) affect reviewers' detection of conclusion deficiencies.

In two experiments, we find that reviewers with a promotion focus, as opposed to a prevention focus, (1) better detect conclusion deficiencies, (2) more ably draw connections between inconsistent evidence and the sufficiency of the holistic evidence set, and (3) are more likely to communicate inconsistent evidence up the engagement team hierarchy, increasing the chances of ultimate resolution. We also examine whether reviewers' detection of conclusion deficiencies is influenced by whether they assess discrete audit tests individually (i.e., segmented review) or as a portfolio (i.e., unsegmented review). Contrary to our expectations, we find no negative effects of segmented review structure on reviewers' judgments. Supplemental tests suggest these effects may be limited to more experienced reviewers, but future research should further examine this possibility.

Our findings contribute to the review process, both to theory and in practice. We contribute to review process theory by highlighting the central importance of holistic thinking to appropriately reviewing tests over inconsistent evidence. We provide theory and evidence that promotion focused reviewers think more holistically, note more inconsistent evidence, and consequently evaluate the set of audit evidence as less sufficient and communicate this issue up the engagement hierarchy. We also find that a promotion focus causes reviewers to weight inconsistent evidence more heavily. That is, promotion focused reviewers focus on inconsistent evidence, link evidence items together, and act on their judgments more effectively than prevention focused reviewers.

Our findings suggest that a promotion focus is a possible tool to improve auditors' well-documented struggles with evaluating the implications of inconsistent evidence (see Austin et al.

2020a; Griffith et al. 2015a). Such a tool is important in the review process, as this process is the primary quality control mechanism for detecting and correcting preparers' deficiencies and client misstatements (Trotman 1985; Trotman and Yetton 1985; Libby and Trotman 1993; Rich, Solomon, and Trotman 1997). Conclusion deficiencies, which can result from inadequate regard for inconsistent evidence, are concerning since financial statement items are increasingly complex (e.g., Barth 2006), increasing the chances that these types of deficiencies exist. Moreover, auditors' increased use of data analytics, which enable auditors to process a greater volume of information and thus inconsistent information, can also lead to information overload (Austin et al. 2020b). Information overload could lead to conclusion deficiencies becoming increasingly common, thus increasing the need for corrective tools.

Because a promotion focus can be activated quickly and unobtrusively, it is worthy of practical and scholarly inquiry to examine the most effective and efficient means to implement such interventions in the review process. For instance, it would be possible to embed a promotion focus intervention into a routine workflow, though questions such as the intervention's strength and duration will need to be resolved. Our theory suggests that such an intervention would be most helpful in high-risk audit areas, such as complex estimates and accounts with nonfinancial inputs, which require a broad focus on interdependencies among evidence items. Moreover, the quality control that effective workpaper review provides is particularly critical in these high-risk areas that are more susceptible to misstatement.

We must caveat that we hold the preparer's conclusions constant in our experiments. Thus, we cannot make claims about how different levels of preparer work quality or workpaper stylization affect reviewers' judgments. This design choice allows us to investigate our research question and theory with precision. Also note that theory suggests that the effects of promotion

focus are likely strongest when there is a high degree of regulatory fit, which is the idea that goal pursuit is more effective when the goal's contextual demands align with the specific regulatory focus (Unkelbach and Ganns 2010). For instance, it is possible that promotion focused reviewers would be particularly effective when the review task is framed as an opportunity (e.g., to improve reporting, to prevent fraud, to coach subordinates, to learn, etc.), since opportunity-seeking is aligned with a promotion focus. Future research could examine whether preparer work quality, preparer stylization attempts, or review task framing interact with regulatory focus to affect review performance.

REFERENCES

- Agoglia, C. P., Hatfield, R. C., & Lambert, T. A. 2015. Audit team time reporting: An agency theory perspective. *Accounting, Organizations and Society*, 44, 1-14.
- Aiken, L.S., West, S.G. and Reno, R.R. 1991. *Multiple regression: Testing and interpreting interactions*. sage.
- Austin, A. A., Hammersley, J. S., & Ricci, M. A. 2020a. Improving Auditors' Consideration of Evidence Contradicting Management's Complex Estimate Assumptions. *Contemporary Accounting Research* 37 (2): 696-716.
- Austin, A. A., Carpenter, T., Christ, M. H., & Nielson, C. 2020b. The Data Analytics Journey: Interactions among Auditors, Managers, Regulation, and Technology. Working paper.
- Bamber, E. M., & Ramsay, R. J. 1997. An investigation of the effects of specialization in audit workpaper review. *Contemporary Accounting Research*, 14(3), 501-513.
- Bamber, E. M., & Ramsay, R. J. 2000. The effects of specialization in audit workpaper review on review efficiency and reviewers' confidence. *Auditing: A Journal of Practice & Theory*, 19(2), 147-157.
- Barrett, L. F., & Kensinger, E. A. 2010. Context is routinely encoded during emotion perception. *Psychological Science*, 21(4), 595-599.
- Barth, M. E., 2006. Including estimates of the future in today's financial statements. *Accounting Horizons*, 20(3), 271-285.
- Bell, T. B., & Griffin, J. B. 2012. Commentary on auditing high-uncertainty fair value estimates. *Auditing: A Journal of Practice & Theory*, 31(1), 147-155.
- Bonner, S.E., Libby, R. and Nelson, M.W. 1996. Using decision aids to improve auditors' conditional probability judgments. *Accounting Review*, 221-240.
- Brazel, J., Leiby, J., & Schaeffer, T. 2020. Do Rewards Encourage Professional Skepticism? Working paper.
- Brazel, J., & Schmidt, J. 2019. Do Auditors and Audit Committees Constrain Inconsistencies between Financial and Nonfinancial Measures? *Auditing: A Journal of Practice & Theory*, 38(1), 103-122.
- Center for Audit Quality (CAQ). 2014. *Professional Judgment Resource*. Washington, DC. Available at: <https://www.thecaq.org/wp-content/uploads/2019/03/professional-judgment-resource.pdf>
- Church, B. K., & Shefchik, L. B. 2011. PCAOB inspections and large accounting firms. *Accounting Horizons*, 26(1), 43-63.

- Dowling, C., & Leech, S.A. 2014. A Big 4 firm's use of information technology to control the audit process: How an audit support system is changing auditor behavior. *Contemporary Accounting Research*, 31(1), 230-252.
- Driver, J., Davis, G., Russell, C., Turatto, M., & Freeman, E. 2001. Segmentation, attention and phenomenal visual objects. *Cognition*, 80(1), 61-95.
- Driver, J., & Baylis, G. C. 1998. *Attention and visual object segmentation*.
- Emett, S. A., Kaplan, S. E., Mauldin, E. G., & Pickerd, J. S. 2020. Auditing with Data and Analytics: External Reviewers' Judgments of Audit Quality and Effort. Working paper.
- EY. 2015. How big data and analytics are transforming the audit.
https://www.ey.com/en_us/assurance/how-big-data-and-analytics-are-transforming-the-audit
- Förster, J., & Higgins, E. T. 2005. How global versus local perception fits regulatory focus. *Psychological Science*, 16(8), 631-636.
- Förster, J., Friedman, R. S., Özelsel, A., & Denzler, M. 2006. Enactment of approach and avoidance behavior influences the scope of perceptual and conceptual attention. *Journal of Experimental Social Psychology*, 42(2), 133-146.
- Friedman, R. S., & Förster, J. 2001. The effects of promotion and prevention cues on creativity. *Journal of personality and social psychology*, 81(6), 1001.
- Friedman, R. S., & Förster, J. 2005. The influence of approach and avoidance cues on attentional flexibility. *Motivation and Emotion*, 29(2), 69-81.
- Freitas, A. L., & Higgins, E. T. 2002. Enjoying goal-directed action: the role of regulatory fit. *Psychological Science* 13(1), 1–7.
- Gibbins, M., & Trotman, K. T. 2002. Audit review: Managers' interpersonal expectations and conduct of the review. *Contemporary Accounting Research*, 19(3), 411-444.
- Glover, S. M., & Prawitt, D. F. 2013. *Enhancing Auditor Professional Skepticism*. Available at: https://www.iaasb.org/system/files/meetings/files/20150615-iaasb-agenda_item_10-b-gloverprawitt_enhancing_auditor_professional_skepticism-final.pdf
- Gold, A., Kadous, K., & Leiby, J. 2020. Does Status Equal Substance? The Effects of Specialist Status on Auditor Assessments of Complex Estimates. Working paper.
- Graham, K., J. Ziegert, and J. Capitano. 2015. The effect of leadership style, framing, and promotion regulatory focus on unethical pro-organizational behavior. *Journal of Business Ethics* 126: 423 – 436.

- Griffith, E. E., J. S. Hammersley, and K. Kadous. 2015a. Audits of complex estimates as verification of management numbers: How institutional pressures shape practice. *Contemporary Accounting Research* 32 (3): 833-863.
- Griffith, E. E., Hammersley, J. S., Kadous, K., and Young, D. 2015b. Auditor mindsets and audits of complex estimates. *Journal of Accounting Research*, 53(1), 49-77.
- Hammersley, J. S., & Ricci, M. A. 2020. Using Audit Programs to Improve Auditor Evidence Collection. *The Accounting Review*, forthcoming.
- Harding, N., & Trotman, K. T. 1999. Hierarchical differences in audit workpaper review performance. *Contemporary Accounting Research*, 16(4), 671-684.
- Hayes, A. 2013. *Introduction to Mediation, Moderation, and Conditional Process Analysis: A Regression-Based Approach* New York, NY: Guilford Press.
- Hayes, A. F., Montoya, A. K., & Rockwood, N. J. 2017. The analysis of mechanisms and their contingencies: PROCESS versus structural equation modeling. *Australasian Marketing Journal* 25 (1): 76-81.
- Higgins, E. T. 1997. Beyond pleasure and pain. *American psychologist*, 52(12), 1280.
- Higgins, E. T. 1998. Promotion and prevention: Regulatory focus as a motivational principle. *Advances in experimental social psychology*, 30, 1-46.
- Higgins, E. T., Roney, C. J., Crowe, E., & Hymes, C. 1994. Ideal versus ought predilections for approach and avoidance distinct self-regulatory systems. *Journal of personality and social psychology*, 66(2), 276.
- Higgins, E. T., Shah, J., & Friedman, R. 1997. Emotional responses to goal attainment: strength of regulatory focus as moderator. *Journal of personality and social psychology*, 72(3), 515.
- Horstmann, G., & Ansorge, U. 2016. Surprise capture and inattention blindness. *Cognition*, 157, 237-249.
- Johnson, L., M. Keune, and J. Winchel. 2019. U.S. auditors' perception of the PCAOB inspection process: A behavioral examination. *Contemporary Accounting Research* 36(3): 1540-1574.
- Kadous, K., & Zhou, Y. D. 2019. How Does Intrinsic Motivation Improve Auditor Judgment in Complex Audit Tasks? *Contemporary Accounting Research*, 36(1): 108-131.
- Kaplan, S., C. Moeckel, and J. Williams. 1992. Auditors' hypothesis plausibility assessments in analytical review settings. *Auditing: A Journal of Practice & Theory* 9(Spring): 50 – 65.

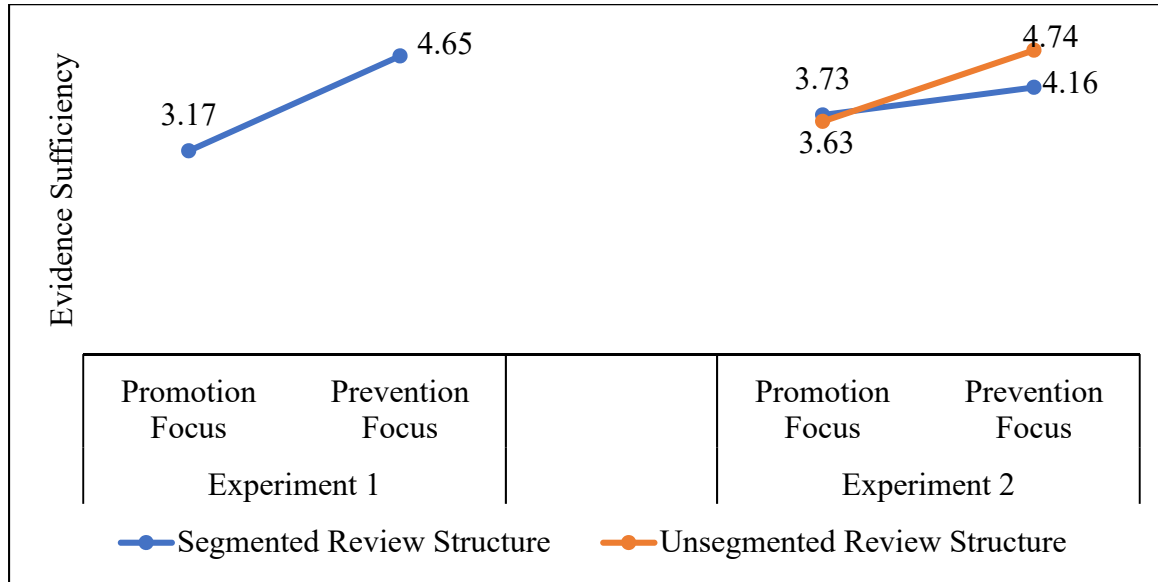
- Knechel, W.R., and J. Leiby. 2016. If You Want My Advice: Status Motives and Audit Consultations about Accounting Estimates. *Journal of Accounting Research* 54(5): 1331 – 1363.
- Kuschel, S., Förster, J., & Denzler, M. 2010. Going beyond information given: How approach versus avoidance cues influence access to higher order information. *Social Psychological and Personality Science*, 1(1), 4-11.
- Lee, K., Choi, J. and Li, Y. J., 2014. Regulatory focus as a predictor of attitudes toward partitioned and combined pricing. *Journal of Consumer Psychology*, 24(3): 355-362.
- Lambert, T. A., Jones, K. L., Brazel, J. F., & Showalter, D. S. 2017. Audit time pressure and earnings quality: An examination of accelerated filings. *Accounting, Organizations and Society*, 58, 50-66.
- Liang, H., Y., Xue, and L. Wu. 2007. Auditor effectiveness and efficiency in workpaper review: The impact of regulatory focus. *Academy of Management Proceedings*. Memmert, D., C.
- Libby, R., and D. Frederick. 1990. Experience and the ability to explain audit findings. *Journal of Accounting Research* 28 (Autumn): 348 – 367.
- Libby, R., and J. Luft. 1993. Determinants of judgment performance in accounting settings: Ability, knowledge, motivation, and environment. *Accounting, Organizations, and Society* 18 (6): 559 – 574.
- Libby, R., & Trotman, K. T. 1993. The review process as a control for differential recall of evidence in auditor judgments. *Accounting, Organizations and Society*, 18(6), 559-574.
- Lightle, S., Castellano, J.F. & Baker, B. 2017. Why audit teams need the confidence to speak up. *Journal of Accountancy*.
- Moeckell, C. 1990. The effect of experience on auditors' memory errors. *Journal of Accounting Research*, 368-387.
- Most, S.B., Scholl, B.J., Clifford, E.R. & Simons, D.J., 2005. What you see is what you set: sustained inattention blindness and the capture of awareness. *Psychological review*, 112(1), 217.
- PCAOB. 2011. *Staff Audit Practice Alert No. 9 – Assessing and Responding to Risk In the Current Economic Environment*. Washington, DC.
- PCAOB. 2013. *In the Matter of Deloitte & Touche LLPs Quality Control Remediation Submission*. PCAOB Release No. 104-2013-191: Washington DC.
- PCAOB. 2014. *In the Matter of Grant Thornton LLPs Quality Control Remediation Submission*. PCAOB Release No. 104-2014-101: Washington DC.

- Polman, E. 2012. Effects of self–other decision making on regulatory focus and choice overload. *Journal of personality and social psychology*, 102(5), 980.
- Raiffa, H., 1968. Decision analysis: introductory lectures on choices under uncertainty.
- Ramsay, R. J. 1994. Senior/manager differences in audit workpaper review performance. *Journal of Accounting Research*, 32(1), 127-135.
- Rich, J. S., Solomon, I., & Trotman, K. T. 1997. The audit review process: A characterization from the persuasion perspective. *Accounting, Organizations and Society*, 22(5), 481-505.
- Shah, J., Higgins, T., & Friedman, R. S. 1998. Performance incentives and means: how regulatory focus influences goal attainment. *Journal of personality and social psychology*, 74(2), 285.
- Shelten, S. 1999. The effect of experience on the use of irrelevant evidence in auditor judgment. *The Accounting Review* 74 (April): 217 – 224.
- Simons, D. J., & Chabris, C. F. 1999. Gorillas in our midst: Sustained inattention blindness for dynamic events. *Perception*, 28(9), 1059-1074.
- Solomon, I., M. Shields, and R. Whittington. 1999. What do industry-specialist auditors know? *Journal of Accounting Research* 37 (Spring): 191 – 208.
- Trotman, K. T. 1985. The review process and the accuracy of auditor judgments. *Journal of Accounting Research*, 740-752.
- Trotman, K. T., & Yetton, P. W. 1985. The effect of the review process on auditor judgments. *Journal of Accounting Research*, 256-267.
- Unkelbach, and S. Ganns. 2010. The impact of regulatory fit on performance in an inattentional blindness paradigm. *The Journal of General Psychology* 137(2): 129 – 139.
- Wallace, J. C., Johnson, P. D., & Frazier, M. L. 2009. An examination of the factorial, construct, and predictive validity and utility of the regulatory focus at work scale. *Journal of Organizational Behavior*, 30(6), 805-831.
- Zimbelman, M.F., 1997. The effects of SAS No. 82 on auditors' attention to fraud risk factors and audit planning decisions. *Journal of Accounting Research*, 35, 75-97.
- Zhu, R., & Meyers-Levy, J. 2007. Exploring the cognitive mechanism that underlies regulatory focus effects. *Journal of Consumer Research*, 34(1), 89-96.

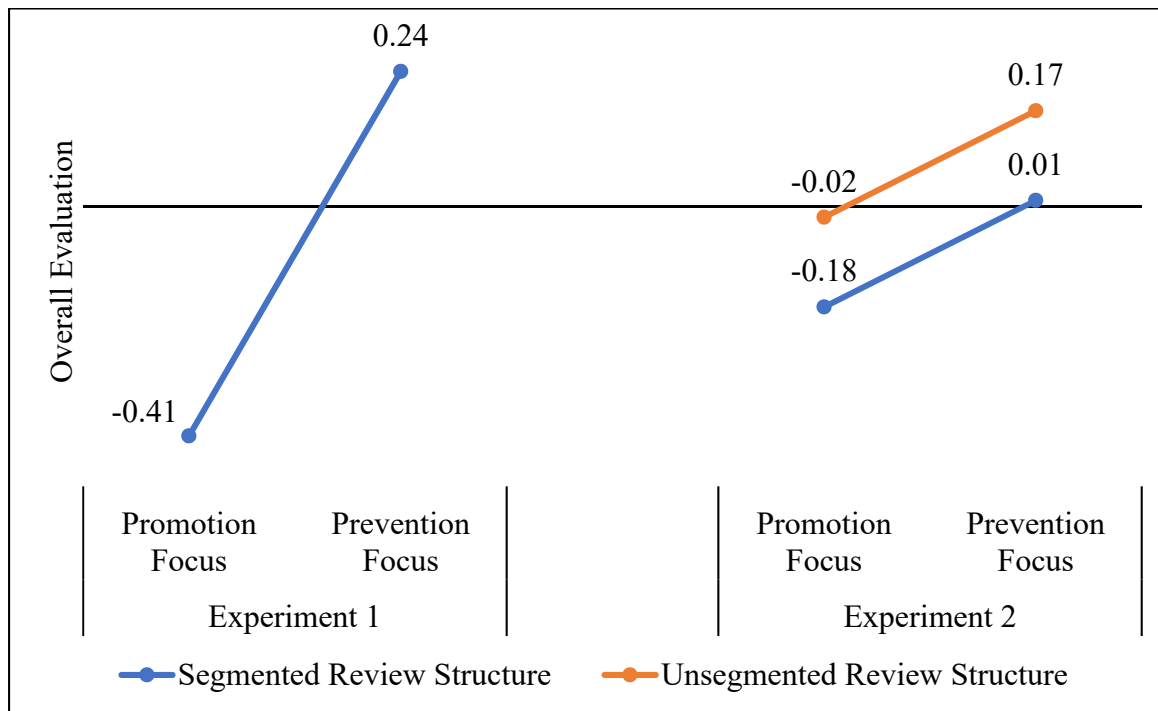
FIGURE 1

Reviewer Assessments of Evidence Sufficiency and Overall Evaluation

Panel A: Experiment 1 and 2, Evidence Sufficiency



Panel B: Experiment 1 and 2, Overall Evaluation



Panel C: Experiment 1 and 2, Sufficiency Ratings for Inconsistent Evidence, Regulatory Focus, and holistic Evidence Sufficiency for the account balance

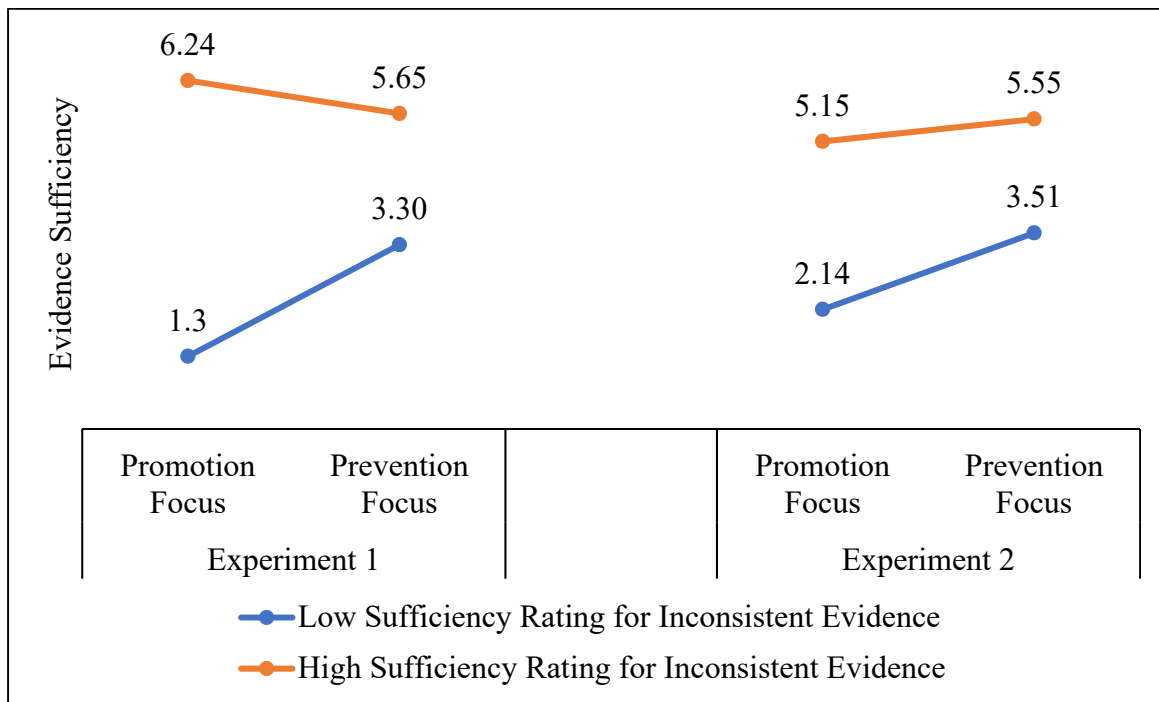


Figure 1 graphs reviewers' mean ratings of evidence sufficiency for the client's holistic test of the sales account balance on 11-point Likert scales anchored by 0 = "Definitely no" and 10 = "Definitely yes." Higher ratings are more supportive of the account balance. Panel A depicts the mean ratings across regulatory focus conditions in Experiment 1 and across regulatory focus and review structure conditions in Experiment 2. We manipulated regulatory focus with either a promotion or prevention focus. Promotion focus auditors receive a prime to describe their current aspiration, ideals and hopes, while prevention focus auditors receive a prime to describe their current obligations, duties, and responsibilities. We manipulated review structure as segmented or unsegmented. Segmented review auditors review and sign off on individual procedures and work for 30 second on an anagram task between procedure reviews, while unsegmented review auditors review six procedures as a portfolio of tests with a single signoff.

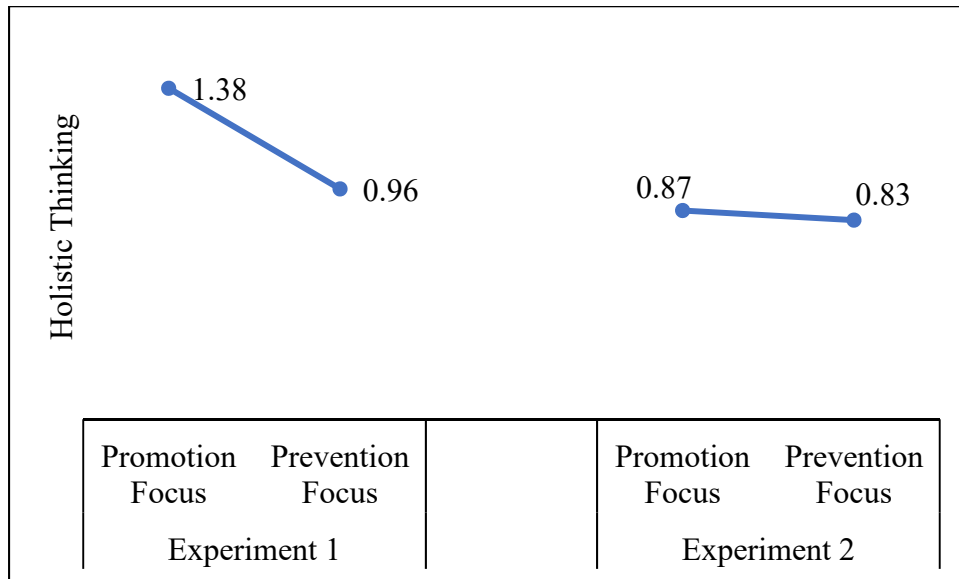
Panel B depicts the mean overall evaluation ratings across regulatory focus conditions in Experiment 1 and across regulatory focus and review structure conditions in Experiment 2. Overall evaluation is a principal component of evidence sufficiency, reasonableness of the client's sales balance, the preparer's performance, and the appropriateness of how the preparer constructed their expectation. Each of these four measures is captured on an 11-point Likert scale.

Panel C depicts the mean evidence sufficiency ratings for the holistic test of the sales account balance across regulatory focus conditions when reviewers rate the evidence sufficiency for individual, inconsistent evidence items as low (i.e., one standard deviation below the mean sufficiency rating for inconsistent evidence) or high (i.e., one standard deviation above the mean sufficiency rating for inconsistent evidence).

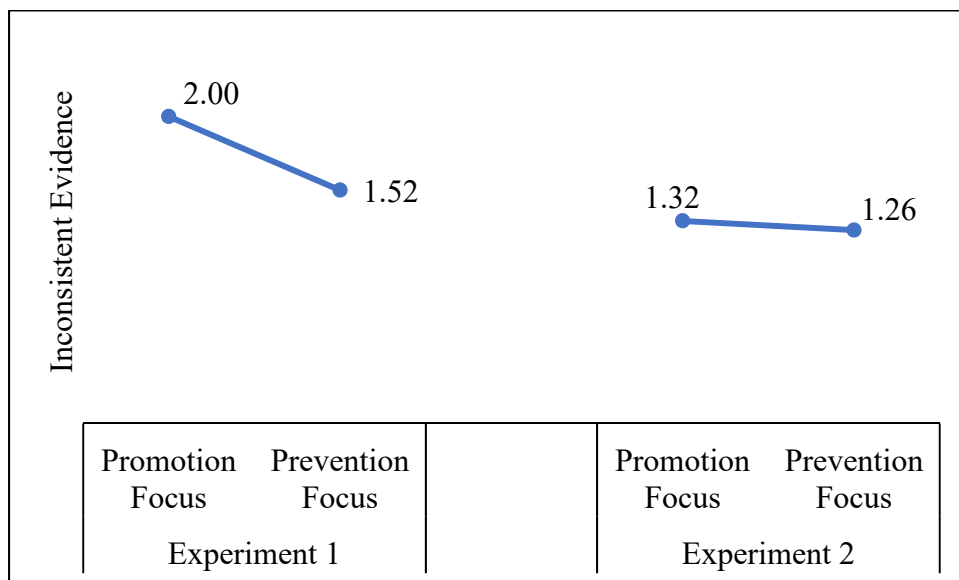
FIGURE 2

Reviewer Assessments of Holistic Thinking, Inconsistent Evidence, and Communication

Panel A: Experiment 1 and 2, Holistic thinking



Panel B: Experiment 1 and 2, Inconsistent Evidence



Panel C: Experiment 1 and 2, Communication of Significant Issues

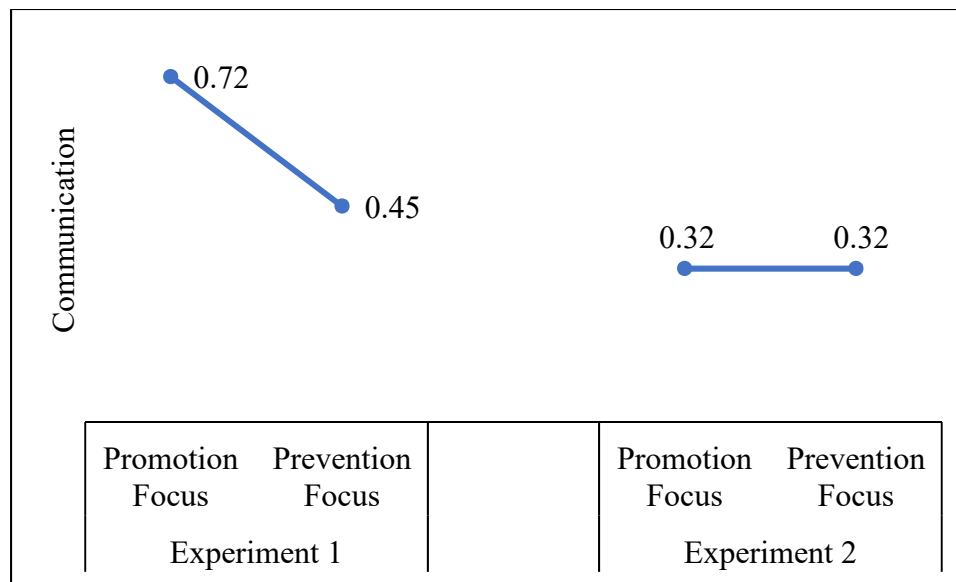


Figure 2 graphs reviewers' mean ratings of holistic thinking, inconsistent evidence, communication of significant issues, and overall evaluation. See Figure 1 for a description of the regulatory focus manipulation. Panel A depicts the mean holistic thinking ratings across regulatory focus conditions in Experiment 1 and 2. Holistic thinking is measured using a coded variable that captures the breadth of auditors' consideration of the audit evidence's implications for the revenue balance. Panel B depicts the mean inconsistent evidence ratings across regulatory focus conditions in Experiment 1 and 2. Inconsistent evidence is measured using a coded variable that captures the extent to which auditors describe identified evidence as inconsistent with revenue. Panel C depicts the mean communication ratings across regulatory focus conditions in Experiment 1 and 2. Communication is measured using a coded variable that captures whether or not auditors raised the inconsistent evidence seeded in our case to their supervisor when given the opportunity.

FIGURE 3

Path Model of the Effect of Regulatory Focus on Holistic Thinking, Inconsistent Evidence, Evidence Sufficiency, and Communication of Significant Issues

Experiment 1

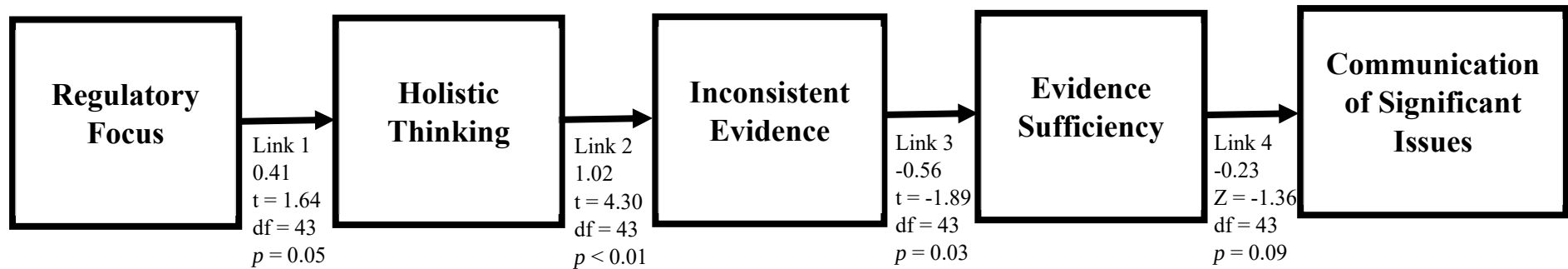


Figure 3 shows a PROCESS model representing the effect of regulatory focus on auditors' holistic thinking, leading to connecting inconsistent evidence, leading to differences in auditors' evidence sufficiency judgments, and finally leading to differences in auditors' proclivity to communicate identified issues to their supervisors. See Figure 1 for a description of the regulatory focus manipulation. We measure holistic thinking using a coded variable that captures the breadth of auditors' consideration of the audit evidence's implications for the revenue balance. We measure inconsistent evidence using a coded variable that captures the extent to which auditors describe identified evidence as inconsistent with or contradicting revenue. We measure evidence sufficiency using a Likert scale which captures how participants rate the sufficiency of evidence to support the client's holistic sales balance. We measure communication of significant issues using a coded variable that captures whether or not auditors raised the inconsistent evidence seeded in our case to their supervisor when given the opportunity. Our model supports our theory that promotion focus improves auditors holistic thinking and then inconsistent evidence connection, appropriately reducing their judgments of evidence sufficiency, and increasing their proclivity to communicate identified issues. All *p*-values are one-sided, reflecting directional predictions.

FIGURE 4

Path Model of the Effect of Regulatory Focus, Review Structure, and Experience on Holistic Thinking, Inconsistent Evidence, Evidence Sufficiency, and Communication of Significant Issues

Experiment 2

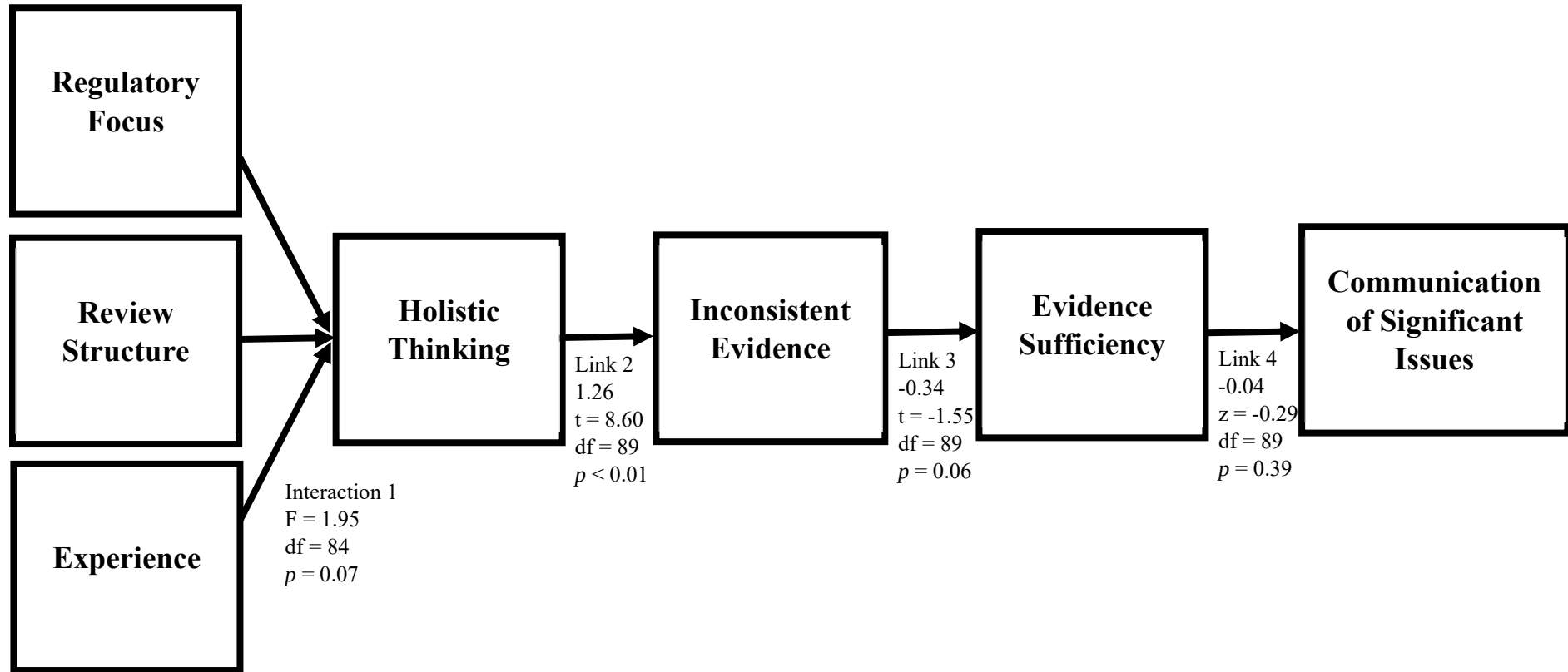


Figure 4 shows a PROCESS model representing the effect of regulatory focus, review structure, and experience on auditors' holistic thinking, leading to connecting inconsistent evidence, leading to differences in auditors' evidence sufficiency judgments, and finally leading to differences in auditors' communication of identified issues to their supervisors. See Figure 1 for a description of the regulatory focus and review structure manipulations. We measure

experience as a continuous variable, using participants' months of audit experience. See Figure 3 for a description of how we measure holistic thinking, inconsistent evidence, evidence sufficiency, and communication of significant issues. Our model suggests that for more experienced reviewers, promotion focus and non-segmented review interact to improve reviewers' holistic thinking, which leads these reviewers to identify more inconsistent evidence and appropriately reduce their judgments of evidence sufficiency. That is, our theory is supported among reviewers with more experience. However, this model does not provide evidence that reviewers who appropriately reduce their judgments of evidence sufficiency are more likely to then communicate significant issues to their supervisors. All *p*-values are one-sided, reflecting directional predictions.

TABLE 1
Participant Demographic Data
Experiments 1 and 2

Panel A: *Demographic Data Points* - Mean (SD) [n]

	Experiment 1	Experiment 2	Overall Mean
CPA	0.90 (0.31) [49]	0.80 (0.40) [94]	0.83 (0.38) [143]
Use of Nonfinancial Measures	0.75 (0.44) [48]	0.65 (0.48) [94]	0.68 (0.467) [142]
Effort on Task	5.47 (1.60) [49]	5.12 (2.02) [94]	5.24 (1.89) [143]
Difficulty of Task	3.76 (1.74) [49]	3.73 (1.91) [94]	3.74 (1.85) [143]

Panel B: *Demographic Data Ranges* – n (%)

	Experiment 1	Experiment 2	Overall Mean
Title	Staff Seniors Managers	1 (1%) 86 (92%) 7 (7%)	2 (1%) 128 (90%) 13 (9%)
Audit Experience	2-3 Years 3-4 Years >4 Years	17 (35%) 17 (35%) 15 (30%)	25 (27%) 45 (48%) 23 (25%)
Manufacturing Clients	< 10% 10-25% 26-50% >50%	39 (43%) 17 (19%) 14 (15%) 21 (23%)	65 (47%) 24 (17%) 17 (12%) 34 (24%)

Table 1 reports demographic data for our auditor participants in our first and second experiments. We code *CPA* as 0 for auditors without a CPA license and as 1 for auditors with a CPA license. We code *Use of Nonfinancial Measures* as 0 for auditors who have not used nonfinancial measures in substantive testing and as 1 for auditors who have used nonfinancial measures in substantive testing. *Effort on Task* is auditors' assessment of how hard they worked on the task in our study, where 0 indicates not hard at all and 10 indicates extremely hard. *Difficulty of Task* is auditors' assessment of the difficulty of the task in our study, where 0 indicates not at all difficult and 10 indicates extremely difficult. We code *Title* as 0 for staff auditors, as 1 for senior auditors, and as 2 for manager auditors. We code *Audit Experience* as 1 for auditors with between two and three years of experience, as 2 for auditors with between 3 and 4 years of experience, and as 3 for more than 4 years of experience. We code *Manufacturing Clients* as 0 for auditors who have spent less than 10% of their chargeable audit hours on manufacturing clients during the previous three years, indicated as 1 for 10 – 25%, indicated as 2 for 26 – 50%, and indicated as 3 for more than 50%. Where *n* varies across variables, it results from participants skipping over our variable measure questions.

TABLE 2

**Sufficiency of Evidence
Experiments 1 and 2**

Panel A: *Experiment 1 Sufficiency of Evidence* - Mean (SE) [n]

Regulatory Focus

Prevention Focus	4.65 (0.46) [31]
Promotion Focus	3.17 (0.60) [18]

Panel B: *Experiment 1 Sufficiency of Evidence* – Pairwise Comparison

Source of Variation	t₄₇	one-tailed p-value
Regulatory Focus	1.96	0.03

Panel C: *Experiment 2 Sufficiency of Evidence* - Mean (SE) [n]

Review Structure

<u>Regulatory Focus</u>	Segmented	Unsegmented	Overall Mean
Prevention Focus	4.16 (0.48) [25]	4.74 (0.46) [27]	4.45 (0.33) [52]
Promotion Focus	3.73 (0.51) [22]	3.63 (0.55) [19]	3.68 (0.38) [41]
Overall Mean	3.94 (0.35) [47]	4.19 (0.36) [46]	

Panel D: *Experiment 2 Sufficiency of Evidence* – ANOVA

Source of Variation	df	MS	F	p-value
Regulatory Focus	1	13.58	2.34	0.14
Review Structure	1	1.34	0.23	0.64
Regulatory Focus x Review Structure	1	2.61	0.45	0.50
Error	89	5.79		

Panel E: *Experiment 2 Sufficiency of Evidence* – Pairwise Comparison

Source of Variation	t₉₁	one-tailed p-value
Regulatory Focus	1.56	0.06

Table 2 reports the results for auditors' assessments of the sufficiency of audit evidence. Regulatory focus is manipulated at two levels. Participants in the promotion focus condition describe their hopes and aspirations at the beginning of the task, priming them to think more holistically. Participants in the prevention focus condition describe their fears and obligations at the beginning of the task, priming them to think more narrowly. Review structure is manipulated at two levels. Participants in the segmented condition review audit tests one at a time while participants in the unsegmented condition review audit tests as a portfolio, all at the same time. Sufficiency of Evidence is measured using a Likert scale which captures how participants rate the sufficiency of evidence to support the client's holistic sales balance.

Note that our analysis in Panel E collapses over review structure, as we find no effect of review structure. Also, one participant from Experiment 2 did not answer our sufficiency of evidence dependent variable question. Thus, there is one fewer participants included in our Table 2 analyses for Panels C – E, as compared to the total number of participants included in Experiment 2. Finally, all p-values are two-tailed unless otherwise indicated.

TABLE 3
Communication of Significant Issues
Experiments 1 and 2

Panel A: *Experiment 1 Communication of Significant Issues* - LS Mean (SE) [n]

<u>Regulatory Focus</u>	
Prevention Focus	-0.19 (0.36) [31]
Promotion Focus	0.96 (0.53) [18]

Panel B: *Experiment 1 Communication of Significant Issues* – Binomial Logit Regression

Source of Variation	df	χ^2	two-tailed p-value
Regulatory focus	1	3.46	0.06

Panel C: Experiment 2 Communication of Significant Issues - LS Mean (SE) [n]

Review Structure

<u>Regulatory Focus</u>	Segmented	Unsegmented	Overall Mean
Prevention Focus	-1.00 (0.44) [26]	-0.53 (0.40) [27]	-0.77 (0.30) [53]
Promotion Focus	-1.22 (0.51) [22]	-0.32 (0.47) [19]	-0.77 (0.35) [41]
Overall Mean	-1.11 (0.34) [48]	-0.43 (0.31) [46]	

Panel D: Experiment 2 Communication of Significant Issues – Binomial Logit Regression

Source of Variation	df	χ^2	two-tailed p-value
Regulatory focus	1	0.00	0.97

Table 3 reports auditors' propensity to communicate significant issues to their supervisors. See Table 2 for an explanation of independent variables. Communication of Significant Issues represents whether auditors raised the inconsistent evidence seeded in our case to their supervisor when given the opportunity.

Note that our analysis in Panel D collapses over review structure, as we find no effect of review structure. Also, all p-values are two-tailed.

TABLE 4

**Experience Effects
Experiment 2**

Panel A: *Sufficiency of Evidence* – LS Mean (SE) [n]

Experience (in months)	Segmented Review Structure	Unsegmented Review Structure	Overall Mean
24-35	4.71 (0.45) [14]	4.72 (0.83) [11]	4.72 (0.43) [25]
36-47	3.64 (0.62) [14]	4.82 (0.56) [22]	4.36 (0.42) [36]
48 +	3.56 (0.63) [18]	3.00 (0.49) [13]	3.32 (0.42) [31]

Panel B: *Holistic Thinking* - LS Mean (SE) [n]

Experience (in months)	Segmented Review Structure	Unsegmented Review Structure	Overall Mean
24-35	0.86 (0.23) [14]	0.64 (0.24) [11]	0.76 (0.17) [25]
36-47	0.86 (0.28) [14]	0.73 (0.16) [22]	0.78 (0.14) [36]
48 +	0.88 (0.23) [17]	1.25 (0.25) [12]	1.03 (0.17) [31]

Panel C: *Inconsistent Evidence* - LS Mean (SE) [n]

Experience (in months)	Segmented Review Structure	Unsegmented Review Structure	Overall Mean
24-35	1.21 (0.55) [14]	1.27 (0.38) [11]	1.24 (0.34) [25]
36-47	1.13 (0.41) [15]	0.82 (0.30) [22]	0.95 (0.24) [37]
48 +	1.28 (0.33) [18]	2.38 (0.47) [13]	1.74 (0.29) [31]

Panel D: *Communication of Significant Issues* - LS Mean (SE) [n]

Experience (in months)	Segmented Review Structure	Unsegmented Review Structure	Overall Mean
24-35	3.29 (0.43) [14]	3.45 (0.49) [11]	3.36 (0.32) [25]
36-47	3.47 (0.42) [15]	2.91 (0.31) [22]	3.14 (0.25) [37]
48 +	2.56 (0.34) [18]	3.62 (0.40) [13]	3.00 (0.27) [31]

Panel E: Overall Evaluation - LS Mean (SE) [n]

Experience (in months)	Segmented Review Structure	Unsegmented Review Structure	Overall Mean
24-35	-0.03 (0.31) [14]	-0.15 (0.25) [11]	-0.08 (0.21) [25]
36-47	-0.03 (0.28) [14]	-0.25 (0.18) [22]	-0.16 (0.16) [36]
48 +	0.03 (0.24) [17]	0.68 (0.29) [12]	0.30 (0.19) [29]

Table 4 reports experience effects in Experiment 2. See Table 2 for an explanation of the review structure independent variable. Note that we do not include regulatory focus in our model because of power constraints. Experience is participants' months of audit experience, categorized into (1) 24 to 35 months of experience, (2) 36 to 47 months of experience, or (3) 48 or months of experience. Sufficiency of evidence is measured using a Likert scale which captures how participants rate the sufficiency of evidence to support the client's holistic sales balance. We measure holistic thinking using a coded variable that captures the breadth of auditors' consideration of the audit evidence's implications for the revenue balance. We measure inconsistent evidence using a coded variable that captures the extent to which auditors describe identified evidence as inconsistent with or contradicting revenue. Communication of significant issues represents whether auditors raised the inconsistent evidence seeded in our case to their supervisor when given the opportunity. Overall evaluation is a principal component of evidence sufficiency, reasonableness of the client's sales balance, the preparer's performance, and the appropriateness of how the preparer constructed their expectation. Each of these four measures is captured on an 11-point Likert scale.