Improving Complex Audit Judgments: A Framework and Evidence

Emily E. Griffith
University of Wisconsin- Madison
egriffith@bus.wisc.edu

Kathryn Kadous
Emory University
kathryn.kadous@emory.edu

Donald Young
Indiana University
youngdr@iu.edu

November 2019

We thank Leo Barcellos, John Campbell, Tina Carpenter, Brian Gale, Stephanie Grant, Ryan Guggenmos, Jackie Hammersley, Karim Jamal, Tamara Lambert, Jennifer McCallen, Michael Ricci, Dan Zhou, conference participants at the 2018 University of Alberta Accounting Research Conference, the 2019 AAA Audit Mid-Year Meeting, the 2019 Contemporary Accounting Research Conference, doctoral seminar participants at the University of Wisconsin-Madison, and workshop participants at Lehigh and Rutgers Universities and the Universities of Florida, Georgia, Illinois, Waterloo, and Washington for helpful comments on prior drafts of the paper. We are grateful for access to participants provided by the Center for Audit Quality's Access to Audit Personnel Program. The views expressed in this paper are the views of the authors alone and not those of the Center for Audit Quality.

Improving Complex Audit Judgments: A Framework and Evidence

Abstract: We develop a framework for improving auditor judgment in complex audit tasks. The framework recognizes that high-quality judgment in complex tasks requires that auditors (1) possess the knowledge to perform analytical processing, (2) recognize the need for analytical (versus heuristic) processing, and (3) have the cognitive capacity to complete the processing. Based on the framework, we predict that auditors' need for cognition (NFC), an individual characteristic theoretically linked to recognizing the need for analytical processing, is associated with better judgments in a complex audit task. Analysis across 11 studies shows that NFC is reliably associated with better complex audit judgments. We further predict and find that priming auditors with an accuracy goal will improve judgments, particularly for lower NFC auditors who are less likely to spontaneously recognize the need for analytical processing. The framework facilitates systematic development of interventions to improve auditor judgment by highlighting that interventions should address the specific conditions that cause judgment problems.

Keywords: Complex estimates; Need for cognition (NFC); Accuracy goal; Goal priming; Professional skepticism; Intervention; Dual-Process Theory

1. Introduction

Auditors have difficulties making high quality judgments in audits of estimates and other complex audit judgment tasks (PCAOB 2015; IFIAR 2017; IFIAR 2019). With respect to estimates, the International Forum of Independent Audit Regulators (IFIAR) reports that, in 2018, 28 percent of audits of public entities contained at least one deficiency related to accounting estimates, including fair value measurements (IFIAR 2019). This same report shows that audit deficiencies are also common in other areas of complexity, including revenue recognition and assessing internal controls over financial reporting.

Complexity is generally defined as a set of task features that increase task difficulty or task resource demands on the decision maker (Bonner 2008; Liu and Li 2012). These features include the number of information cues, diversity of those cues, amount of ambiguity versus structure, and relationships or interconnectivity among cues (Bonner 2008; Liu and Li 2012). In order to perform complex judgment tasks well, auditors must consider a broad set of information and must consider relationships among various pieces of information (e.g., Bonner, Kadous, and Majors 2019; Brewster 2011; Griffith, Hammersley, Kadous, and Young 2015b). For example, auditors must consider incomplete fact patterns to identify possible fraud (Hammersley 2006; Simon 2012) and generate alternative hypotheses for management's explanations of unexpected fluctuations in analytical procedures (Brewster 2011; Bucaro 2019; Plumlee, Rixom, and Rosman 2014) while considering a broad set of diverse, ambiguous, and interrelated cues. Performance of complex tasks benefits from thinking broadly, objectively, and deeply about information from many sources and accurately analyzing and synthesizing the information.

In this paper, we propose a framework for improving complex audit judgments. The framework recognizes that high quality auditor performance of complex tasks is related to three

conditions. First, the auditor must have the requisite task-specific and procedural knowledge (collectively "mindware") to perform the analytical processing that the task requires. Second, the auditor must recognize the need to engage in analytical (versus heuristic) processing. Third, the auditor must have the cognitive capacity to complete the analytical processing. The framework further recognizes that these conditions are linked to individual auditor characteristics. Finally, the framework highlights that, in order to be effective, interventions should address the specific condition that reduces judgment quality.

We further develop reasoning that the second condition, recognizing the need to engage in analytical versus heuristic processing, is often problematic for complex audit tasks. Following this reasoning, we test the usefulness of our framework by using it to identify an individual characteristic associated with variability in recognizing the need for analytical processing and to develop an appropriate intervention.

Based on the framework, we propose that an auditor's trait level of need for cognition (NFC), a thinking disposition characterized by a person's tendency to engage in and enjoy thinking (Cacioppo and Petty 1982), is associated with auditor performance on complex audit tasks. People high in need for cognition are more likely to spontaneously engage in information acquisition, reasoning, and problem solving across many domains in their lives (Cacioppo, Petty, Feinstein, and Jarvis 1996; Strobel, Fleischhauer, Luong, and Strobel 2018). Importantly, NFC is reliably positively associated with cognitive processes focused on validity (i.e., accuracy) (Petty, Brinol, Loersch, and McCaslin 2009) and with a higher likelihood of recognizing the need for an analytical (versus heuristic) response (Stanovich and West 2008). In complex tasks, analytical thinking should help auditors identify relevant evidence and use it to understand and predict relationships. Accordingly, we predict that auditors higher in NFC will be better able to identify

inappropriate management assumptions underlying estimates and so will perform better on a complex audit task related to estimates. We confirm this idea in a preliminary analysis of data from 11 studies collected for other purposes in which participants made complex audit judgments and we measured NFC. NFC is reliably negatively associated with assessed reasonableness of biased assumptions across these studies.

We next develop an intervention to improve judgment for lower NFC auditors. The framework indicates a successful intervention would facilitate lower NFC auditors' recognition that an analytical response is needed without harming that recognition among higher NFC auditors. We propose that priming an accuracy goal fits these requirements. Research in psychology and accounting emphasizes the importance of goals for individuals' information processing and task performance (e.g., Weber and Johnson 2009; Griffith, Kadous, and Young 2016). Goals affect how individuals frame a problem, which aspects they focus on, and what strategies they use to solve the problem (Weber and Johnson 2009). While auditors higher in need for cognition naturally engage in processes focused on validity, auditors lower in need for cognition are less likely to frame a problem expansively, focus objectively on all of the relevant information, and choose more analytical strategies unless they are prompted to do so. We expect that priming an accuracy goal will evoke these aspects of problem solving, thereby compensating, to some extent, for low levels of need for cognition.

We examine our predictions in an experiment in which 71 senior auditor participants evaluate a client's revenue projections used as an assumption underlying a fair value estimate. Participants receive background information about the client, information gathered by the audit team, and information provided by the client. While the client projects strong revenue growth due to its planned introduction of a new product, evidence embedded in the case suggests that the

revenue projections are biased. Understanding the implications of this evidence requires analytical processing, allowing us to make inferences about participants' processing based on their use of the evidence. We measure participants' need for cognition using Cacioppo, Petty, and Kao's (1984) scale, and we manipulate goal (accuracy vs. support the client) by having participants complete a word search puzzle designed to prime one of the goals. Participants assess the reasonableness of the revenue projections and give justifications for their assessments.

We find results consistent with our hypotheses. First, higher need for cognition is associated with lower assessments of the reasonableness of the biased assumption. Process analysis shows that higher NFC auditors are more likely to identify and use embedded evidence that conflicts with management's assumption about revenue growth and less likely to heuristically support management's assumption, leading to improved judgments. Second, NFC and goal type interact such that lower NFC auditors are more sensitive to the primed goal. Specifically, lower NFC auditors make better judgments when primed with an accuracy goal, while higher NFC auditors' judgments do not differ across goal types. Process analysis indicates that this improvement comes from the accuracy goal reducing lower NFC auditors' heuristic support for management's assumption. Our use of a moderation-of-process design demonstrating that the manipulated goal affects judgments of higher and lower NFC auditors differently, as well as our mediation-by-measurement strengthen the theoretical inferences from these results (Asay, Guggenmos, Kadous, Koonce, and Libby 2019).

This paper makes several contributions. First, researchers, practitioners, and regulators have focused in recent years on developing interventions for auditors' shortcomings in making complex audit judgments. A common solution in practice has been to increase training or apply decision aids and analytics, whereas researchers often use mindset interventions and judgment

frameworks. Our framework clarifies that *these solutions address different problems*. While training and decision aids are useful in providing the mindware to solve complex judgment problems, they do not address other conditions limiting judgment quality. Our framework provides a means for more systematic, and thus, more successful development of solutions that match the specific nature of the problems. In addition, because prior frameworks were developed before dual-process theory was firmly established in the judgment and decision making literature, they do not incorporate the idea that recognition of the need for analytical processing is a key condition for high quality complex audit judgments. Our framework builds on the dual-process idea to include a broader set of potential solutions to auditor judgment problems. Researchers and practitioners should find it useful in improving audit quality.

Second, our framework and empirical findings provide further support for the idea that analytical thinking is essential for audit quality when audits include complex judgment tasks (e.g., Hoffman and Zimbelman 2009; Brewster 2011; Plumlee et al. 2014; Griffith et al. 2015b; Kadous and Zhou 2019). We contribute to this growing literature by providing evidence that an individual characteristic that is theoretically linked to analytical processing is associated with higher quality judgments in complex audit tasks. Viator, Bagley, Barnes, and Harp (2014) argue that NFC should be included in studies examining the effectiveness of feedback interventions to improve judgments because it is associated with the ability to benefit from feedback. We provide evidence that NFC is also relevant to the quality of complex audit judgments.

Finally, we provide evidence that priming an accuracy goal can temporarily compensate, at least in part, for deficits in need for cognition. Auditors juggle a number of goals. Priming an accuracy goal appears to alert those auditors who most need it to use more analytical processes, improving their judgments. In our study, priming an accuracy goal improved lower NFC

auditors' performance and did not reduce higher NFC auditors' performance, indicating that priming all auditors with an accuracy goal before certain tasks could be a simple, practical solution to problems with complex judgments.

2. Framework

In this section, we develop a theoretical framework that shows how auditor characteristics and interventions affect cognitive processes and judgments in complex tasks. The purpose of this framework is to facilitate researcher identification of interventions that will best address the underlying cause of judgment problems and are therefore most likely to improve complex judgments.

Our framework is similar to prior debiasing frameworks, including Arkes (1991) and Kennedy (1995) in that it recognizes that, to improve judgment, an intervention must address the specific conditions limiting judgment quality. It differs in two important respects. First, as we explain further below, our framework accommodates developments over the intervening period in psychologists' understanding of dual process theory and how it impacts judgment and decision making. Use of this theory refines our understanding of how and why complex judgments fail and, in doing so, expands the set of potential debiasing methods. Second, our framework specifically addresses complex audit tasks. Arkes (1991) encourages researchers to develop task-specific taxonomies that build upon the more general debiasing frameworks, and Shanteau (1989) argues that behavioral audit research should focus on problems of importance to auditors, rather than on generic judgment problems.

We draw our ideas from the dual process paradigm, which proposes that individuals use two distinct modes of processing to complete tasks and make judgments (Kahneman and Frederick 2002; Evans and Stanovich 2013). The defining features of heuristic processes (also

known as "Type 1" processes or "thinking fast") are that they operate autonomously and do not require working memory. These processes are generally characterized as quick, intuitive, heuristic processes that are used to efficiently complete a task. The defining features of analytical processes (also known as "Type 2" processes or "thinking slow") are that they require working memory and allow for hypothetical reasoning. These processes are generally characterized as deliberate, analytical processes that require cognitive effort.[1] Dual process models typically assume that individuals tend to default to heuristic processes unless they recognize, either on their own or due to some feature in the environment, that more analytical processing is required for the task at hand (Stanovich and West 2008; Evans and Stanovich 2013). In the audit setting, this corresponds to auditors defaulting to heuristic processes such as confirming separate aspects of management representations unless they recognize that a more analytical process is needed (e.g., Griffith, Hammersley, and Kadous 2015a; PCAOB 2018).[2]

Within the dual processing paradigm, the psychology literature identifies three conditions for successful completion of tasks requiring analytical processing (Stanovich and West 2008). These are that the decision maker must possess the appropriate "mindware" (i.e., the knowledge and procedural skills needed to perform the task well), detect the need for an analytical response, and possess the capacity to sustain analytical processing until the task is complete. A failure at

[1] While recent dual process paradigms provide defining features (i.e., the requirement for working memory and allowance for hypothetical reasoning) that can be used to distinguish different cognitive processes (Pennycook, De Neys, Evans, Stanovich, and Thompson 2018), some research has used these defining features to identify common correlated features (e.g., intentionality, controllability, efficiency). Evans and Stanovich (2013) note that many of the characteristics used in the extant literature to describe dual process theories are *correlated* features and should not be confused with *defining* features of Type 1 and 2 processes.

[2] We acknowledge that there is debate about the extent to which the two processes represent different *systems* of cognition or, more simply, different *modes* of reasoning (e.g., Bellini-Leite 2018; Evans and Stanovich 2013; Pennycook 2018). In the audit setting, the distinction between simpler, more heuristic processes, such as verifying management representations and more analytical processes aimed at finding the best estimate is likely one of reasoning mode. We subscribe to the view that the essential ideas that (1) humans have access to both more and less analytical types of reasoning and (2) analytical thought has to be triggered by something are commonly accepted (Pennycook 2018). These ideas are sufficient to support our framework.

any stage results in a heuristic response, and so each condition is relevant only if the prior stages are satisfied. For example, if one lacks the knowledge to perform a task properly, whether one recognizes the need for or can sustain analytical processing is irrelevant. Our framework (Figure 1) adapts and expands upon these ideas to help researchers and practitioners understand when and why auditor judgment on complex tasks fails and to identify the type of intervention needed for improved auditor performance of complex judgment tasks given particular auditor characteristics. Our framework reflects the logic that in order to improve judgment, the intervention should be matched to the underlying cause of the shortcoming.

[Insert Figure 1 here]

The first row of boxes in Figure 1 reflects that auditors must possess the right mindware to make high quality judgments in complex tasks (e.g., Bonner 2008). Mindware includes task-specific and procedural knowledge, that is, an individual's knowledge of how to perform a task (Kahneman and Frederick 2002). In auditing, relevant mindware includes general knowledge, various types of technical knowledge, and procedural knowledge (Bonner and Lewis 1990; Libby and Luft 1993) as well as tacit managerial knowledge (Tan and Libby 1997). For example, an auditor examining a complex accounting issue must know the relevant accounting rules and understand the transaction to perform well, but he or she must also know the mechanics of performing the audit task. Procedural knowledge in this case involves knowing what evidence is relevant, how to obtain the evidence, how to evaluate that evidence, what constitutes a problem with the evidence, how the problem would be resolved, and so on. An auditor lacking this knowledge would not be expected to successfully complete the task without an intervention.

Mindware is acquired through training and experience, so auditors likely possess necessary mindware for the complex tasks that they frequently encounter. Lack of necessary

mindware is more likely to be problematic for less experienced auditors and for more technical or novel tasks. Appropriate interventions for mindware failure include those that provide the relevant knowledge, such as technical training (Earley 2001) or decision aids (Butler 1985), or that substitute for the knowledge, such as data analytics tools that provide appropriate estimates without requiring the auditor to understand a model (Bell, Bedard, Johnstone, and Smith 2002).

The second row of boxes in Figure 1 reflects the idea that auditors, like all individuals, often default to simpler, heuristic processing even if they have the right mindware. They must somehow recognize that their approach is inappropriate and that more analytical processing is warranted for the task at hand so that they override the default heuristic approach (Evans and Stanovich 2013). For example, it is more natural for even knowledgeable individuals to choose evidence that confirms or matches an initial hypothesis rather than evidence that could disconfirm it (e.g., Evans 2016). In the audit setting, the piecemeal nature of audit work, standards' framing of procedures as verification, and time pressure combine to create conditions that enhance the likelihood of heuristic processing (e.g., Griffith et al. 2015a; Griffith et al. 2015b), with the result that auditors tend to engage in the simpler, heuristic default of confirming management's assertions rather than engaging in a more appropriate process (PCAOB 2018).[3] Pressures for efficiency likely further reduce auditors' tendency to evaluate whether they are using an appropriate process. Thus, failures to detect the need for a more analytical response are likely common causes of low judgment quality for complex audit tasks.

Individuals' cognitive styles (thinking dispositions) can influence their propensity to engage in analytical processing. For example, as we will discuss in the next section, the need for

[3] For example, these conditions may put auditors in an implemental mindset (Griffith et al. 2015b), or the fact that the processes involved in confirming management's assertions are straightforward and efficient and appropriate for simpler tasks could result in habitual use of heuristic processes (Bonner et al. 2019). In either case, auditors automatically engage in heuristic processing and are unlikely to recognize a need for analytical processing.

cognition (NFC), one's tendency to engage in and enjoy thinking (Cacioppo and Petty 1982), is associated with a higher likelihood of detecting the need for override of a heuristic response (Cacioppo et al. 1996). In addition, auditors may also have processing habits that prevent them from recognizing that more analytical processing is required (Bonner et al. 2019).

Appropriate interventions for failure to recognize the need for analytical processing could invoke the desired processes either directly by priming a deliberative mindset (Griffith et al. 2015b), or indirectly by priming intrinsic motivation (Kadous and Zhou 2019) or an accuracy goal (as in the current paper). Alternatively, judgment frameworks that prioritize particular issues (Backof, Bamber, and Carpenter 2016), audit programs that emphasize the overall task versus individual steps or balanced consideration of evidence (Hammersley and Ricci 2019; Austin, Hammersley, and Ricci 2019), or skepticism or fraud prompts that highlight specific information (Hammersley, Bamber, and Carpenter 2010; Hoffman and Zimbelman 2009) could alert auditors to the need to use more analytical processes.

Finally, the third row of boxes in Figure 1 reflects that, in addition to the first two conditions, individuals must also have the capacity to sustain inhibition of the default, heuristic response in order to perform well in complex tasks (Stanovich and West 2008). This is commonly referred to as cognitive capacity (Petty, Wheeler, and Tormala 2003) or ability (e.g., Bonner and Lewis 1990). Because analytical processing is more cognitively demanding than heuristic processing, even an individual who recognizes the need for analytical processing may revert to less demanding processing if capacity to execute the necessary analytical processes is lacking. For example, in auditing management's assumptions underlying a complex estimate, an auditor must evaluate each assumption individually, which includes considering contradictory evidence and alternative scenarios, and also consider the assumptions jointly for problematic

patterns or inconsistencies (e.g., all projections at the aggressive end of the range, some projections assume increasing sales while others assume decreasing sales).

Intelligence is the primary determinant of cognitive capacity. Several traditional measures of intelligence are positively associated with decision performance where analytical processing is required (Stanovich and West 2008). In the audit setting, audit partner IQ is positively associated with archival proxies for audit quality (Kallunki, Kallunki, Niemi, and Nilsson 2019) and auditors' performance on Graduate Record Examination questions is positively associated with their performance on a ratio analysis task (Bonner and Lewis 1990). That said, distraction and fatigue also influence cognitive capacity (Petty, Wegener, and Fabrigar 1997). Interventions to address shortcomings in cognitive capacity include those that use accountability or other mechanisms to increase motivation to continue working despite fatigue (e.g., Kennedy 1993), as well as those that use affect, affirmation, or other mechanisms to reduce the impact of fatigue on actions (e.g., Mullis and Hatfield 2018). In addition, decision aids can be designed to increase cognitive capacity by, for example, placing information to be compared in a single location (DeZoort, Harrison, and Taylor 2006).[4]

Our framework indicates that if all three conditions are satisfied, opportunities for an appropriate analytical response are increased. It further suggests that, for tasks that require deliberate, analytical processing, interventions should be designed to specifically compensate for the characteristic or condition that causes the heuristic response.[5] For example, a lack of relevant

[4] While the three levels of the framework and conditions for high quality judgments are theoretically distinct, some interventions may operate at more than one level. For example, salient intrinsic motivation encourages a broader and more thoughtful approach to a problem (Cerasoli, Nicklin, and Ford 2014; Kadous and Zhou 2019), indicating that its primary effect is likely to be to help in detecting when analytical processing is required (level 2); however, it is also associated with working longer on tasks (Cerasoli et al. 2014: Vansteenkiste, Simons, Lens, Sheldon, and Deci 2004), suggesting that it may also address the third level.

[5] The framework is designed to facilitate systematic development of interventions to improve auditor judgment, rather than to be the last word on the subject. In this spirit, the lists of traits and interventions included in the framework are not intended to be exhaustive.

knowledge that implies inadequate mindware could be addressed with training or decision aids that provide the necessary information; low need for cognition that prevents detection of the need for analytical processing could be addressed by priming an accuracy goal; and fatigue that impedes sustained analytical processing could be addressed with interventions prompting motivation or positive affect. Likewise, the framework suggests that interventions should avoid inhibiting the effects of characteristic and conditions that are conducive to analytical processing.[6]

3. Hypotheses Development

In the remainder of the paper, we develop and test hypotheses based on our framework. We focus our testing on the second level of the framework for two reasons. First, this level of the framework is novel in the accounting literature. As we note above, prior frameworks include decision problems and solutions related to the first (mindware) and third (capacity) levels, but do not contemplate the possibility that auditors may not recognize the need to engage in analytical processing. Second, this level of the framework is likely most relevant to auditors' observed judgment problems with complex tasks. As discussed earlier, pressures for efficiency and other features of the audit setting likely prevent detection of a need for override. In addition, complex audit judgments continue to be problematic despite that they are made by intelligent, knowledgeable professional auditors who are highly motivated to make good judgments. This suggests that issues other than mindware and capacity are limiting.

[6] As noted above, Arkes (1991) and Kennedy (1995) provide debiasing frameworks. Kennedy's (1995) framework classifies errors as resulting from problems with internal or external "data" (overlapping with level 1 in our framework) or with effort, which captures the decision maker's capacity and motivation (i.e., level 3 in our framework). Arkes (1991) classifies errors as strategy-based versus association- (or perceptual-) based, resulting in the idea that improving judgment comes from either modifying the decision maker or modifying the decision environment. His suggested solutions fit into the first and third levels of our framework. Subsequent advances in dual-process theory (e.g., Stanovich and West 2008) highlight the importance of the second level of our framework—recognition of the need for analytical processing—in addressing complex judgment tasks among knowledgeable professionals, giving rise to potentially more effective solutions in this context.

The need for cognition and complex audit judgments

Complexity in judgment tasks arises from features associated with task difficulty or capacity requirements such as the number of elements, number of information cues, diversity of information cues, extent of ambiguity versus structure, and the nature of relationships and connectivity among cues (Bonner 2008; Liu and Li 2012). Many important audit tasks, including auditors' evaluation of complex estimates and their underlying assumptions, qualify as complex in that they involve significant ambiguity and a large number of diverse, inter-related information cues. The complexity tends to trip auditors up. This is evidenced by the fact that audit deficiencies tend to cluster in areas of complexity (e.g., audits of complex estimates and revenues, analysis of internal controls) (e.g., IFIAR 2019) and that, within the area of complex estimates, specific deficiencies tend to include failures to consider all relevant information and to consider relationships among information cues (e.g., IFIAR 2017). Our framework suggests that some auditors are likely better at handling these types of complex decision tasks than others. In particular, focusing on the second level of the framework, we expect that auditors with a disposition to process information completely and reflectively (i.e., those with a higher need for cognition) will perform better in complex audit judgment tasks.

The need for cognition (NFC) is a well-established cognitive style that captures stable individual differences in the tendency to engage in and enjoy thinking (Cacioppo and Petty 1982; Fleischhauer, Enge, Brocke, and Ullruck 2010). Those who are higher in NFC tend to "seek, acquire, think about, and reflect back on information to make sense of stimuli, relationships, and events…" (Cacioppo et al. 1996, 198). Lower NFC individuals do not tend to spontaneously engage in deep thinking and are less likely to detect a need to override a heuristic response

(Stanovich and West 2008), so they tend to rely on heuristics more frequently than their higher NFC counterparts.

For example, a number of studies show that argument quality has a larger impact on higher (versus lower) NFC individuals' judgments about message persuasiveness (Cacioppo, Petty, and Morris 1983; Cacioppo et al. 1996; Luttrell, Petty, and Xu 2017).[7] Similarly, in a consumer context, more relevant cues such as product attributes influence higher NFC individuals' decisions, while less relevant cues such as spokesperson attractiveness influence lower NFC individuals' decisions (Haugtvedt, Petty, and Cacioppo 1992; Peltier and Schibrowsky 1994).[8]

While these studies do not employ complex tasks, consistent with our framework, they demonstrate that the need for cognition reliably predicts cognitive processing and judgments (Fleischhauer et al. 2010; Strobel et al. 2018).[9] In particular, higher NFC is associated with careful analysis of the quality of relevant cues and a decision maker focus on validity (Petty et al. 2009). Because higher NFC individuals are more likely to detect a need to engage in analytical processing, they tend to consider a broader set of information (e.g., Verplanken, Hazenberg, and Palenewen 1992), evaluate that information carefully to discriminate relevant from irrelevant cues (e.g., Cacioppo and Petty 1982; Peltier and Schibrowsky 1994), and develop a better

[7] In these studies, student participants read short editorials proposing a controversial policy that include supporting arguments that are either strong or weak. They are typically asked to rate the persuasiveness of the message or provide their attitude toward it.

[8] In these studies, student participants view ads for products such as calculators, sometimes for as little as 15 seconds, and they then rate the product and recall information from the ad.

[9] An exception is Nair and Ramnarayan (2000), who examine the relation between NFC and managerial decisions about production and resource allocation in a computer simulation game. Their task is complex in that it involves many interacting information cues from different areas of the simulated organization (e.g., marketing, personnel, finance). However, Nair and Ramnarayan's (2000) setting lacks the ambiguity and the conflicting goals and of complex audit tasks. That is, all information cues are perfectly reliable; there is no uncertainty about reliability as in the audit setting with management-provided information and assertions that can be and often are biased (PCAOB 2018). Ambiguity and conflicting goals and incentives could interfere with NFC's theoretical effects. In addition, Nair and Ramnarayan's (2000) task is concerned with prediction rather than evaluation of others' judgments.

understanding of relationships among cues (e.g., Nair and Ramnarayan 2000). This implies that higher NFC auditors will be more likely to consider all relevant information cues, their diagnosticity, and their inter-relationships. Therefore, we expect that higher NFC auditors will be better able than lower NFC auditors to identify unreasonable (biased) assumptions underlying a complex estimate.

> **Hypothesis 1:** Higher NFC auditors will be more likely to identify an unreasonable management assumption than will lower NFC auditors.

Audit researchers have proposed that NFC is relevant to auditor discovery of fraud (Hammersley 2011) and auditor performance in complex audit tasks (Griffith et al. 2016) for reasons similar to those above; however, evidence is limited. Prior research shows that NFC is associated with auditors' ethical judgments (Setiawan 2017) and that higher NFC auditors make better use of feedback interventions (Viator et al. 2014).[10] Based on our framework, we expect that NFC is also associated with judgment quality in complex tasks. Despite the available evidence, this hypothesis is not without tension. First, it assumes that auditors have the mindware for the task (i.e., level 1 of the framework is satisfied). Absent the relevant mindware, higher NFC would not improve auditor judgments. Second, it assumes that these knowledgeable auditors are not always engaged in analytical thinking when they encounter complex audit tasks. If they were, higher NFC would not influence auditors' judgment performance.

The need for cognition, auditor goals, and complex audit judgments

[10] Setiawan (2017) demonstrates that NFC is significantly correlated with the extent to which auditors report they would resist pressure to engage in unethical behavior. Participants rate their willingness to go along with either a supervisor's or client's unethical suggestion (e.g., to allow the client access to inventory sample selections, to remove certain accounts from a confirmation sample, to ignore a misstatement) across five scenarios. Viator et al. (2014) find that feedback interventions tested in four prior studies improve auditor performance only for auditors higher in NFC. Two of Viator et al.'s tasks are highly abstracted from the accounting context, and the other two tasks examine auditors' cue weightings for bankruptcy and gross profit predictions.

While individuals lower in NFC do not spontaneously engage in analytical processing, they will engage in it is needed and they are prompted to do so (Fleischhauer et al. 2010; Strobel et al. 2018; Petty et al. 2009, Sandra and Otto 2018). We expect that providing auditors with an appropriate goal will prompt lower NFC auditors to recognize the need for analytical processing, and that this will result in improved judgments. We employ a goal manipulation for this purpose.

Goals drive behavior, influencing all aspects of the decision making process from decision framing to strategy selection to execution (Weber and Johnson 2009). Auditors presumably have a goal of making high quality judgments; however, they simultaneously pursue other conflicting goals (e.g., Ricci 2019).[11] In particular, auditors have a number of incentives that make it attractive to support the client's position (Johnstone, Warfield, and Sutton 2001) such that, in general, auditors prefer to agree with the client (Austin et al. 2019). Auditor behavior is consistent with a support goal across a variety of tasks (e.g., Hackenbrack and Nelson 1996; Kadous, Kennedy, and Peecher 2003; Moore, Tetlock, Tanlu, and Bazerman 2006).

In the context of complex tasks, support goals can be satisfied with heuristic processing but accuracy goals tend to require more analytical processing. Interview evidence suggests that auditors tend to approach audits of complex estimates in a heuristic way; they view their task as that of collecting evidence to support aspects of management's estimate (Griffith et al. 2015a). Consistent with this, regulators argue that auditors default to a heuristic process of seeking support for individual aspects of management's estimates rather than using a more analytical approach that considers disconfirming evidence and joint implications of evidence (PCAOB

[11] These goals include completing the audit on time and within budget (McDaniel 1990; Mocadlo 2018; Bhaskar, Hopkins, and Schroeder 2019), avoiding disagreement with their supervisors (Peecher 1996; Wilks 2002), avoiding uncomfortable client interactions (Bennett and Hatfield 2013), avoiding litigation costs (Kida 1980; Hackenbrack and Nelson 1996; Blay 2005), and avoiding negative inspection findings (Christensen, Glover, Omer, and Shelley 2016; Glover, Taylor, and Wu 2018; Johnson, Keune, and Winchel 2019).

2018). This claim is backed up by results of several experimental studies (Austin et al. 2019; Griffith et al. 2015b; Kadous and Zhou 2019; Majors and Bonner 2019).[12]

In complex audit tasks, we expect that an accuracy goal can prompt auditors to use more analytical processes, leading to higher quality judgments, particularly for auditors lower in NFC. For example, Chen, Shechter, and Chaiken (1996) demonstrate that an accuracy goal causes relatively even-handed and analytical processing as compared to a goal of getting along with a partner, which results in biased processing that favors the partner's position. In a complex judgment task such as evaluating assumptions underlying estimates, the cognitive processes associated with an accuracy goal should help auditors make judgments that are less biased and that take into account more of the available relevant information.

We previously noted that higher NFC is associated with a decision maker focus on validity (Petty et al. 2009) and analytical, even-handed processing consistent with an accuracy goal. This implies that auditors higher in NFC are more likely to override the heuristic response and engage in analytical processing, independent of their goals, while auditors lower in NFC are more likely to use heuristic processing. Therefore, we expect that goals and NFC have an interactive effect on auditors' cognitive processing and the resulting judgments. Lower NFC auditors will benefit more from an accuracy goal, which encourages them to override the heuristic response of supporting the client and to engage in more analytical processing. In other

[12] Austin et al. (2019) show that an intervention for balanced evidence consideration increases attention to and elaboration of disconfirming evidence compared to a control condition, suggesting auditors in the control condition used relatively more heuristic and less analytical processing. Griffith et al. (2015b), Kadous and Zhou (2019), and Majors and Bonner (2019) each report experiments with three conditions: an intervention for more analytical processing, an intervention for more heuristic (i.e., confirmatory) processing, and a control condition. In each case, auditors' cognitive processing and judgments are very similar in the heuristic and the control conditions, but are different from that in the analytical conditions, suggesting that auditors' control condition (default) processing tends to be more heuristic than analytical.

words, we predict that providing the "right" goal will be more important for lower NFC auditors' judgment performance.

> **Hypothesis 2:** Priming auditors with an accuracy (versus support) goal will improve auditors' identification of an unreasonable management assumption to a greater extent for lower NFC auditors than for higher NFC auditors.

Prior audit research supports the idea that accuracy goals can influence auditors' judgments, but their effects appear to depend the strength and salience of additional goals, and possibly how the goal is instantiated. Accuracy goals shift judgments away from supporting the client in ambiguous scenarios where conflicting goals are weaker or are not made salient (Peecher 1996; Asare and Cianci 2009; Asare, Cianci, and Tsakumis 2009). These scenarios are brief and do not include evidence to process or embedded issues to identify. Accuracy goals are indirectly and consciously invoked via accountability to a party with a preference for objectivity (Peecher 1996) or for compliance with GAAS (Asare and Cianci 2009; Asare et al. 2009). However, when accuracy goals (manipulated by asking auditors to identify the "best" method) are layered over strong directional goals to agree with the client in ambiguous settings, they can have the effect of enhancing pro-client bias (Kadous et al. 2003; Koch and Salterio 2017). Consistent with Kunda's (1990) theory and Kadous et al.'s (2003) arguments, we speculate that when auditors with strong support goals are given conscious accuracy goals, they become more motivated to come to a correct judgment, so they apply more effort; however, they apply that effort to the same inappropriate processes that serve their support goals. This could arise from auditors mentally wrestling with how to satisfy both goals.[13] In contrast, nonconscious goal

[13] None of the cited studies examined auditors' cognitive processing. Studies that do tend to find no effect of conscious accuracy goals on processing or judgments. Austin et al. (2019) consciously and directly invoke an accuracy goal by instructing auditors to determine the most accurate values for a client's revenue projections. They find no evidence of an effect of the accuracy goal on processing or judgments relative to a control condition, and they conclude that the accuracy goal conflicted with auditors' concurrent goal to support the client's preference, and was ignored. Zhou (2019) manipulates an accuracy goal via accountability to a party with a preference for

interventions prime appropriate processes directly, avoiding conscious efforts to satisfy multiple goals. Thus, they may be more effective across a variety of settings. We expect that a nonconscious accuracy goal manipulation will be effective for lower NFC auditors, who do not already have this implicit goal.[14]

4. Preliminary Evidence on NFC and Complex Audit Judgments

As a preliminary test of our framework, we jointly analyzed data from 11 related studies collected as pilot and main studies to test various questions. The studies comprise 1287 participants who completed complex audit tasks (assessing the reasonableness of either a biased revenue projection or a biased estimate that relies on several assumptions) and Cacioppo et al.'s (1984) need for cognition scale. Participants represent three populations: auditors, students, MTurk workers. In each case, participants rated the reasonableness of a biased estimate or assumption on an 11-point scale ranging from 0 (not at all likely or completely unreasonable) to 10 (extremely likely or completely reasonable). We report the analysis in Appendix A.

Panel A of Appendix A describes the studies. Panel B reports NFC means by population. The mean (standard deviation of) NFC for auditors is 65.43 (10.32). This is significantly greater than the mean for students (M = 62.23) and is equal to the mean for MTurk workers (M = 65.19). NFC is negatively associated with reasonableness judgments for auditors and MTurk workers, indicating better judgments for higher NFC in these populations, but not for students.[15] Panel C

---

objectivity as well as intrinsic motivational orientation. He finds that intrinsic motivation improved cognitive processes but there was no effect of the accuracy goal on either processing or judgments.

[14] In the context of complex audit judgments, Majors and Bonner (2019) nonconsciously prime "rubber stamp" or "refuse to accept" goals and compare them to a control condition. Their "refuse to accept" goal encourages auditors to question, rather than support, the client, so it is not an accuracy goal, per se, but, like an accuracy goal, is designed to move auditors away from support. This goal results in greater identification of embedded fraud risks as compared with other conditions, while there are no differences between the rubber stamp goal and the control condition.

[15] We speculate that for students, inadequate mindware (particularly task-specific and procedural knowledge) is the limiting factor in their judgments, rather than a failure to recognize that analytical processing is required. Our framework implies that training or a decision aid might be a helpful intervention for this population.

shows that the mean NFC score is higher for auditors with more than three years of experience (M = 66.71) than for those with three years or less of experience (M = 63.98), but it is not significantly higher for CPA auditors (M = 66.18) than for non-CPAs (M = 64.72).[16] Most important for our purposes, Panel D shows that NFC is significantly negatively related to reasonableness judgments across these 11 studies, and that the effect is robust to inclusion of experience, college GPA, intrinsic and extrinsic motivational orientation, and other individual difference variables, where available. This analysis supports the idea that NFC and the increased recognition of a need for analytical processing that comes with it is relevant to judgment quality for complex audit judgments, and that it contributes over and above other demographic factors.

5. Experimental Method

To test our specific hypotheses, we conducted a 2 x 2 between-participants experiment in which participants evaluated revenue projections underlying a client's fair value estimation, which was part of a goodwill impairment analysis. We measured NFC and we manipulated goal as either "support the client" (hereafter, support goal) or "make an accurate conclusion" (hereafter, accuracy goal).[17] This moderation-of-process design enhances our confidence about NFC's causal role, despite our inability to randomly assign participants to lower and higher NFC

[16] The lack of significance for the CPA variable may be due to low power, as the studies include few non-CPA auditors.

[17] The full design is 2 x 2 x 2. In addition to the independent variables identified above, we manipulated auditors' accountability *to the client* at two levels (higher and lower) by telling participants in the higher client accountability condition that the controller is very involved in the audit process and would be asking for a justification of the auditor's evaluation. As a manipulation check, we asked participants to rate on an 11-point Likert scale, "How much pressure did you feel to justify your evaluation to the client?" with endpoints "Very little pressure" (0) and "A great deal of pressure" (10). Participants attended to the client accountability manipulation (M = 5.78 vs. 4.37, two-tailed p = 0.020). However, this manipulation did not significantly affect our dependent measure, reasonableness assessments (main effect, p = 0.637; all two- and three-way interactions, p > 0.51) or our process measure, identification of embedded evidence (main effect for each measure, p > 0.68; all two- and three-way interactions, p > 0.19). Thus, we pool data across both client accountability levels in the reported analyses. Further, client accountability is not correlated with NFC (r = 0.073, two-tailed p = 0.542).

conditions, as the predicted interaction with a manipulated variable allows insight into the theoretical process (e.g., Asay et al. 2019).

Participants first completed the goal manipulation. They next read background information about the client, information gathered by the audit team, and information provided by the client. The client projected strong revenue growth due to its planned introduction of a new product in the coming fiscal year. Sales forecasts and charts supported management's projections. However, evidence embedded across multiple narrative sections of the case indicated that the revenue projections were overly optimistic. Specifically, projections were not consistent with prevailing macroeconomic conditions, and close examination of past forecasts revealed that management had consistently over-projected in past years. We designed the case so that auditors using their default, heuristic processes would be more likely to focus on the supportive, easily interpreted forecasts and charts that do not suggest any bias in the client's projections. In contrast, auditors using analytical processes would be more likely to identify the disconnect between management's projections and macroeconomic conditions and link this disconfirming cue with others (e.g., management's history of over-projection) to identify bias. After reading the case information, participants completed the dependent and process measures, as well as post-experimental questions that included the need for cognition scale.

Participants

Participants are senior auditors from two Big 4 firms. Seventy-eight auditors completed the study in a paper and pencil format at firm training sessions. We omit from analyses data from five participants who reported their effort as less than or equal to one on a 0-10 scale and from two participants who skipped the goal manipulation.[18] Our final sample includes 71 participants

[18] The five participants who reported low effort spent significantly less time on the task than the rest of the sample (M = 16.2 vs. 27.0 minutes, two-tailed $p < 0.001$). The two participants who skipped the goal manipulation did not

with an average of 3.35 years of experience and who have worked on an average of 1.51 audits on which they have evaluated revenue projections.

Independent variables

We measure participants' need for cognition with Cacioppo et al.'s (1984) short-form Need for Cognition Scale. This scale includes 18 statements. Participants respond on a 1-5 scale to indicate the extent to which each statement is characteristic of them. Example items include: "I would prefer complex to simple problems," "I like to have the responsibility of handling a situation that requires a lot of thinking," and "Thinking is not my idea of fun." Some items, such as the third example, are reverse-coded. Participants' need for cognition score is their sum across the 18 items. Psychology researchers have extensively tested the reliability and validity of this scale (e.g., Cacioppo et al. 1996; Bors, Vigneau, and Lalande 2006; Fleischhauer et al. 2010). The scale reliably measures a distinct, unidimensional construct that reflects a trait invariant to situational stimuli (Bors et al. 2006).[19] Our reliance on this well-established, previously validated and reliable scale ensures that we are capturing the construct of interest and minimizes measurement error.[20] We use a median split to create the lower and higher NFC conditions used

---

make any marks on the word search puzzles used as the manipulation. Inferences do not change if we include these data. Specifically, tests of H1 and H2 remain significant at $p < 0.05$. Although the low-effort participants provided more invalid justifications (M = 0.80 vs. 0.14, two-tailed p = 0.004) and fewer valid justifications (M = 2.2 vs. 3.0, two-tailed p = 0.058) than the rest of the sample, overall SEM results do not change.

[19] Though NFC reflects a distinct construct, it is modestly correlated with some measures of intelligence and ability (e.g. Tidwell, Sadowski, and Pate 2000; Bors et al. 2006). Fleischhauer et al. (2010, Table 3) report correlations ranging from 0.06 to 0.25 between NFC and 11 intelligence measures. Correlations between NFC and intelligence are stronger for fluid versus crystallized intelligence, suggesting that any associations with intelligence are driven by reasoning tests (Fleischhauer et al. 2010). These small correlations may reflect higher NFC participants' higher enjoyment of and propensity to engage in reasoning. Overall these studies demonstrate limited overlap between NFC and general intelligence. In the studies reported in Appendix A, NFC is not significantly correlated with GPA (for students [r = 0.006, two-tailed p = 0.932], MTurk workers [r = -0.024, two-tailed p = 0.677], auditors [r = 0.042, two-tailed p = 0.478], or collectively [r = -0.010, two-tailed p = 0.798]). NFC is not significantly correlated with the goal manipulation (r = 0.056, two-tailed p = 0.638).

[20] Internal consistency of NFC scores in our study, as indicated by Cronbach's $\alpha = 0.90$, is high and consistent with other studies that use this scale ($\alpha = 0.84$ in Fleischhauer et al. (2010), $\alpha = 0.90$ in Cacioppo et al. (1984), and $0.85 \leq \alpha \leq 0.97$ in other studies that use the 18-item scale as summarized by Cacioppo et al. (1996)).

in the primary analyses, and we test the robustness of our results using alternative specifications.

We manipulate goal at two levels (support vs. accuracy) between participants with word search puzzles designed to prime either goal nonconsciously prior to starting the case. Goal pursuit is commonly thought to consist of three functions: individuals adopt goals, select strategies to achieve them, and monitor and control their progress (Custers and Aarts 2014). Although we tend to think of these functions as being consciously taken, in fact all of these steps routinely occur outside of our awareness (Custers and Aarts 2010, 2014). Thus, a goal primed at a nonconscious level can result in mental and physical actions being taken without individuals consciously selecting them.

Psychology research commonly uses word search puzzles with goal-related words to prime nonconscious goals (e.g., Bargh, Gollwitzer, Lee-Chai, Barndollar, and Trötschel 2001; Sheeran, Webb, and Gollwitzer 2005). Such nonconscious priming manipulations have the advantage of isolating the semantic content of the manipulation from the content of the experimental task, thereby improving our ability to infer that the manipulated goal construct caused the observed results rather than any information contained in the manipulation. Following this principle, we designed two word search puzzles to prime support or accuracy goals. Each puzzle includes seven target words chosen to prime the relevant goal and the same six neutral words. Each puzzle is 10 x 10, and words can appear horizontally, vertically, or diagonally. Words can be spelled forward or backward. We placed the neutral words in the same location in each version to hold difficulty constant (Payne, Duggan, and Neth 2007). A pretest confirmed that the puzzles were similar in terms of difficulty and completion time.[21]

[21] Fourteen (twelve) individuals who did not participate in the study pre-tested the support (accuracy) word search. Pre-test participants took an average of 3.36 (3.67) minutes to complete the support (accuracy) word search (t_{24} = -0.546, two-tailed p = 0.590) and rated difficulty as 3.71 (3.67) on a 0-10 scale (t_{24} = 0.097, two-tailed p = 0.923).

Word search puzzles used to prime goals typically include synonyms for the goal itself or goal-related actions or processes. For example, Bargh et al. (2001) prime a high performance goal using a puzzle with six verbs related to performance (*win, compete, succeed, strive, attain, achieve,* and *master*) as target words and seven neutral words. Similarly, Hart and Albarracín (2009) prime achievement using a word search with eight synonyms for the verb "achieve" as target words and five neutral words. To follow this approach, we first identified a key verb related to each of our goals: *validate* for the support goal and *investigate* for the accuracy goal. We chose these verbs because they epitomize actions or processes auditors would undertake in pursuit of either goal. We identified synonyms for the key verbs "validate" and "investigate" and chose seven verbs that reflect the constructs of interest, that would resonate with auditors, and that fit the parameters of our format (e.g., no more than 10 letters long if the word is to be placed horizontally or vertically). Our final word list for the support goal included *approve, confirm, support, endorse, affirm, verify,* and *validate*. Our final word list for the accuracy goal included *examine, probe, question, scrutinize, delve, dig,* and *explore*.

Dependent variable and process measure

After reading the case information described above, participants answered the dependent and process measures. For the dependent variable, participants answered "How likely is it that management's revenue projections are reasonable?" on a 0-10 scale anchored by 0, not at all likely, and 10, extremely likely. These assessments reflect participants' identification of an unreasonable management assumption, because the embedded evidence indicates that management's revenue projection is biased.

For the process measure, we coded the issues included in participants' justifications for their reasonableness assessments. Participants provided their justifications immediately after

assessing the reasonableness of management's revenue projections. The justification question read: "Briefly summarize the reasons for your evaluation in the box below. When writing your summary, be sure to include your explanations and justifications for your evaluation." The format of this question retains consistency with the natural audit task and allows us to create relatively unobtrusive measures of the evidence they attended to and how they interpreted and synthesized it to make their judgments (Asay et al. 2019; Kadous and Zhou 2018).

We coded the issues included in justifications into one of seven categories that we determined in advance, based on the case design. The first five categories represent valid reasons for assessing the revenue projections as unreasonable based on the embedded evidence. These categories include items related to (1) the client's history of overestimating revenue, (2) broader negative macroeconomic trends, (3) uncertainty about the new product upon which the client's expectations for revenue growth rely, (4) concerns about the client's discontinuation of an old product, and (5) other valid concerns about the revenue projection. The remaining two categories include items that (6) support management's projection or (7) are invalid (i.e., factually inaccurate based on the case).[22] An author and a research assistant who were both blind to experimental condition independently coded all items and achieved 91.1 percent initial agreement (Cohen's $k = 0.889$, $p < 0.001$). Coders met to resolve disagreements; analyses use resolved coding. Our analyses use the count of items coded into each category to examine the process by which NFC affects judgments, allowing us to identify which information influenced auditors' judgments.

[22] Recall that auditors using analytical processing would be more likely to attend to and understand the implications of the embedded evidence, and auditors using heuristic processing would be more likely to focus on the information supporting management's revenue projections.

6. Experimental Results

Initial analysis

We used a median split to assign participants to NFC conditions, assigning those with scores at or above the median of 65 to the higher NFC condition and those with scores below 65 to the lower NFC condition. Participants' mean NFC score is 64.30 and the standard deviation is 11.76. The mean NFC score in the lower NFC condition (M = 55.06) is significantly lower than the mean in the higher NFC condition (M = 73.28, $t_{71} = 10.366$, $p < 0.001$).

To examine adherence to our goal manipulation, we established that 69 out of 71 participants (97%) completed the word search puzzle (i.e., found all 13 words). Completion rates did not differ across conditions (97% in the accuracy goal condition and 97% in the support goal condition, $F_{1, 69} = 0.000$, two-tailed $p = 0.984$). We did not include a manipulation check because people are unaware of the effects of primed states on their behavior, so responses are unreliable (Bar-Anan, Wilson, and Hassin 2010; Hauser, Ellsworth, and Gonzalez 2018). Following Bar-Anan et al. (2010), we asked participants if they thought the puzzle had a theme and, if so, whether they thought the theme influenced their performance on the audit task. While 41 participants (58%) thought the puzzle had a theme, only six participants (8%) thought the theme influenced their performance. Neither thinking the puzzle had a theme nor thinking the puzzle theme influenced performance was correlated with reasonableness assessments ($\rho = -0.071$, two-tailed $p = 0.556$ and $\rho = -0.118$, two-tailed $p = 0.331$, respectively).

We further analyzed the responses of the 41 participants who thought the puzzle had a theme. Seventeen of these 41 participants *correctly* identified the puzzle's theme as support or accuracy. The proportion of participants who correctly identified the theme does not differ across goal conditions (8 out of 36 (22.2%) in the accuracy goal condition vs. 9 out of 35 (25.7%) in the

support goal condition, $X^2_1 = 0.119$, $p = 0.730$). Further, correctly identifying the theme is not correlated with reasonableness assessments ($\rho = -0.045$, two-tailed $p = 0.707$). Thus, any effect of the goal manipulation on the results is not likely due to conscious awareness of its influence.

Tests of hypotheses

Table 1, Panel A reports descriptive statistics for reasonableness assessments by condition. Means are graphed in Figure 2. Table 1, Panel B reports an ANOVA with goal and need for cognition as independent variables and assessed reasonableness of the revenue projections as the dependent variable. Our first hypothesis predicts that higher NFC auditors will be more likely to identify an unreasonable management assumption than will lower NFC auditors. Consistent with H1, higher NFC auditors assess the revenue projections as less reasonable ($M = 4.03$) than do lower NFC auditors ($M = 5.06$; $F_{1,67} = 12.39$, $p < 0.001$).

[Insert Table 1 and Figure 2 here]

Our second hypothesis predicts that providing auditors with an accuracy goal will be more beneficial for lower NFC auditors than for higher NFC auditors. Figure 2 shows that the predicted pattern of means (Panel A) is similar to the observed pattern (Panel B). In the ANOVA model (Table 1, Panel B), the non-directional interaction test is marginally significant ($p = 0.083$). An untabulated directional test of the interaction using the planned contrast supports H2 ($t_{67} = 1.76$, $p = 0.041$).[23] The simple effects reported in Table 1, Panel C show that the primed goal has a significant effect on the reasonableness assessments of lower NFC auditors (p =

[23] That is, contrast weights are [-1, +1, +1, -1]. Use of this contrast to test H2 avoids the need to make assumptions about the relative position of cell means and ensures that our tests of H1 and H2 are orthogonal (e.g., Myers, Well, and Lorch 2010; Guggenmos, Piercey, and Agoglia 2018).

0.022), but not on those of higher NFC auditors (p = 0.888). These results demonstrate that the effect of the primed goal is greater for lower versus higher NFC auditors.

A joint test of H1 and H2 using a planned contrast with weights of [-1, +3, -1, -1] captures the entire pattern shown in Figure 1 and is reported in Table 1, Panel D. This contrast is significant ($F_{1,67}$ = 14.61, two-tailed p < 0.001) and has low residual between-cells variation (q^2 = 13.0%; semi omnibus $F_{2,67}$ = 1.18, p = 0.314), indicating that the predicted pattern fits the observed data well. In sum, we find that higher trait levels of need for cognition are associated with better judgments about an unreasonable management assumption and that priming an accuracy goal helps lower NFC auditors make higher quality judgments in this complex task.

Our results are robust to alternative specifications.[24] Need for cognition has a significant main effect when we use a Wilcoxon rank sum (Mann-Whitney) test for differences in medians (z = 3.21, p = 0.001). Further, if we split NFC scores into terciles (highest, M = 76.12, n = 25; middle, M = 64.91, n = 22; lowest, M = 51.42, n = 24) and conduct a one-way ANOVA, the effect of NFC is significant whether modeled categorically ($F_{2,68}$ = 4.24, p = 0.018) or as ordered terciles ($F_{1,69}$ = 8.59, p = 0.005). We also obtain consistent results for the interaction (H2) test when we only use the subsample of observations in the highest and lowest terciles (M = 76.12, n = 25 and M = 51.42, n = 24, respectively) and repeat our ANOVA and planned contrast analyses (t_{45} = 1.90 and one-tailed p = 0.032). The joint test of H1 and H2 also remains significant ($F_{1,45}$ = 16.44 and p < 0.001).

Our results are also robust to the inclusion of covariates. When we include measures of work and task-specific experience as well as intrinsic and extrinsic motivation as covariates in

[24] These robustness analyses address the underlying issues related to using a dichotomous, rather than continuous, NFC measure. Specifically, the non-parametric tests reported above provide evidence that outliers do not drive our results, and the tercile analyses demonstrate that our results are not sensitive to the breakpoint used (i.e., median).

our models, only extrinsic motivation is significant ($p = 0.047$; $p > 0.5$ for intrinsic motivation, work experience, and task-specific experience). Both our separate and joint tests of H1 and H2 remain significant when we include any of the covariates in the analysis.

Process analyses

We next perform additional analyses to provide insight into how and why high trait levels of need for cognition and the accuracy goal improve auditors' judgments in our complex task.

How does NFC improve reasonableness judgments?

Our theory predicts that higher NFC auditors will make higher quality judgments because these auditors use more analytical processes that allow them to identify and use the diagnostic embedded evidence that contradicts management's assumptions, while avoiding the heuristic response of relying on information that supports the client. Because there is a match between identified issues and processing (i.e., identification of embedded contradictory evidence and understanding its implications requires analytical processes, while reliance on supportive evidence requires heuristic processes), we can examine processing by analyzing the issues participants included in their justifications.[25, 26] We estimate a structural equations model to test

[25] Coding issues included in participants' justifications allows for relatively unobtrusive measures of process, as providing justifications is a normal part of performing the audit task. Using more intrusive questions or measures could provide an inaccurate picture of decision process due to lack of participant insight or by distorting processing and judgments (Asay et al. 2019; Hauser, Ellsworth, and Gonzalez 2018; Kadous and Zhou 2018).

[26] Issues most commonly identified were broader negative macroeconomic trends ($M = 0.87$), followed by doubts about the viability of the client's new product ($M = 0.75$), items supporting management's projections ($M = 0.45$), the client's history of overestimating revenue ($M = 0.45$), other valid concerns about the revenue projections (0.23), invalid items (0.14), and concerns about the client's discontinuation of an old product (0.11). A stepwise regression ($p < 0.10$ threshold) of the reasonableness assessment on the coded justification variables shows that negative macroeconomic trends ($p < 0.01$), items supporting management's projections ($p = 0.01$), and the client's history of overestimating revenue ($p = 0.01$) significantly predict auditors' reasonableness assessments. No other predictors are significant. In preliminary modeling, we examine whether higher NFC auditors better utilize the two embedded issues and are able to avoid over-relying on the supporting justifications using a Poisson binomial regression of the number of issues identified for each justification on higher NFC. We find that higher NFC auditors better identify broader negative macroeconomic trends as an issue ($p = 0.049$, one-tailed) and they are less likely to include items supporting management's projection ($p = 0.035$, one-tailed). Although directionally consistent, higher NFC auditors are not significantly more likely to identify the client's history of overestimating revenue ($p = 0.266$, one-tailed).

whether identifying embedded evidence about broader negative macroeconomic trends and disregarding information supporting management's projections jointly mediate the effect of NFC on auditors' reasonableness assessments. The model is shown in Figure 3. Overall, the model fits the data well. The traditional Chi-square test shows a good fit ($\chi^2_1 = 0.11$, $p = 0.74$), as do other standard fit measures, including the Comparative Fit Index (CFI = 1.00), which is above the generally accepted minimum value of 0.95 (Byrne 2013), and the Root Mean Square Error of Approximation (RMSEA = 0.00), which is below the 0.05 rule of thumb indicating good fit (MacCallum, Browne, and Sugawara 1996).

Path coefficients show that the relationship between NFC and identification and use of embedded evidence (Link 1) is positive and significant ($p = 0.05$), indicating that higher NFC is associated with greater identification and incorporation of diagnostic evidence that contradicts management's assumption (i.e., negative macroeconomic trends). The relationship between higher NFC and justifications supporting the revenue projections (Link 2) is negative and significant ($p = 0.04$), indicating that higher NFC is associated with less reliance on justifications supporting management's projection. Finally, both use of the embedded evidence (Link 3) and justifications supporting the revenue projections (Link 4) are significantly associated with reasonableness assessments in the expected directions (negative, $p < 0.01$ and positive, $p < 0.01$, respectively).

In sum, higher NFC auditors make better judgments on our complex task because they are better able to identify and use embedded evidence that conflicts with management's assumptions and they are less likely to heuristically support management's position. These relationships support the framework's assertion that NFC is associated with auditors' ability to recognize the need for analytical processing, and that a failure to do so degrades auditors'

judgment quality in this task. The model also shows a significantly negative direct effect of NFC on reasonableness assessments (Link 5), indicating that NFC has additional impacts on judgment beyond those captured by these two process variables.

[Insert Figure 3 here]

How does an accuracy goal improve lower NFC auditors' judgments?

We perform a similar structural equations analysis to assess how the accuracy goal influences higher and lower NFC auditors' judgments. Our theory predicts an interactive effect of NFC and goal on auditors' cognitive processing and resulting judgments. Whereas higher NFC auditors are more likely than their lower NFC counterparts to engage in analytical processing about a client's assumptions regardless of the primed goal, lower NFC auditors will be less critical of the client's assumptions in the absence of the right goal and thus will more likely be influenced by the primed accuracy goal. We estimate a nested structural equations model to test whether priming a goal has a greater effect on lower NFC auditors than on higher NFC auditors. The model is shown in Figure 4.

[Insert Figure 4 here]

Overall, the model fits the data well. The traditional Chi-square test shows a good fit (χ^2_4= 2.82, p = 0.59), as do other standard fit measures, including the Comparative Fit Index (CFI = 1.00), and the Root Mean Square Error of Approximation (RMSEA = 0.00). Path coefficients show that the accuracy goal does not significantly affect identification and use of embedded evidence (Link 1) for either level of NFC (higher NFC: p = 0.99, lower NFC: p = 0.86). Instead, the mediating path is through justifications supporting the revenue projections (Link 2), which depends on the level of NFC (χ^2_1 = 3.42, p = 0.06). This path is negative and significant for lower NFC auditors (p < 0.01), but insignificant for higher NFC auditors (p = 0.90), indicating that the

primed accuracy goal reduces the extent to which lower NFC auditors heuristically support the client's position. Goal type does not impact higher NFC auditors' use of this process. The model further shows that both identification and use of embedded evidence (Link 3) and justifications supporting the revenue projections (Link 4) are associated with reasonableness assessments in the expected directions (negative, $p < 0.01$ and positive, $p = 0.05$ respectively). In addition, the direct effect of the accuracy goal on reasonableness assessments (Link 5) is significant and negative for lower NFC auditors ($p < 0.01$), indicating that the accuracy goal has additional helpful impacts on the judgments of lower NFC auditors that are not captured by the process mediators included in our model. The direct effect of the accuracy goal is significant only for lower NFC auditors ($\chi^2_1 = 3.80$, $p = 0.05$). Moderated mediation tests further support the interactive effect of NFC and goal on identification and use of embedded evidence (Link 1) and on justifications supporting the revenue projections (Link 2). See Figure 4 for details.

7. Conclusion

This paper develops a framework for improving complex audit judgments. The framework systematically analyzes conditions leading to higher quality judgments in complex audit tasks, thereby facilitating development of interventions that address the specific judgment problems that limit auditor performance. We expect that researchers will find the framework useful in their efforts to improve auditor judgment and audit quality. While we developed the framework in the auditing context, we expect it applies to complex tasks that require deliberate, analytical thinking across a variety of contexts. Future work can test the framework's applicability to other complex judgments.

The framework suggests that a common cause of auditors' judgment problems in complex tasks is that some auditors engage in heuristic processing aimed at supporting a client's

position and fail to recognize a need to use analytical processing. Prior debiasing frameworks omit this important condition to high-quality judgment. By including it, our framework allows the insight that an individual difference variable related to this recognition, auditors' dispositional tendency to engage in and enjoy effortful thinking (NFC), as well as interventions that support this recognition, such as accuracy goals, are associated with better judgment in complex audit tasks. We provide experimental support for this aspect of our framework by demonstrating that higher NFC auditors make better judgments in a complex audit case. Process evidence is consistent with the idea that they are more likely to identify and use disconfirming information and are less likely to heuristically support the client. We also find that NFC interacts with primed accuracy goals to influence judgment performance and processing in a complex audit task. In particular, priming lower NFC auditors with an accuracy goal improves their judgments by reducing their heuristic support for the client's position. This result demonstrates the value of designing an intervention to target the source of the judgment problem.

This study also contributes beyond the framework. First, this study establishes the importance of NFC in auditors' performance in complex judgment tasks. Need for cognition is a well-established dispositional trait that has been shown to impact persuasion and information processing outside of accounting; however, it has not been widely used in accounting or auditing research. Our framework and evidence highlight its importance in complex audit tasks. Given the importance of an analytical judgment process to professional skepticism (Nolder and Kadous 2018), our evidence suggests that NFC is an important input into auditor skepticism. Of practical relevance, NFC can be measured simply, quickly, and at low cost; firms could collect this information to use in making decisions about staffing engagements and assigning responsibility

for specific audit tasks in addition to determining who might benefit from targeted interventions such as the one in this study.

Second, this study shows that providing auditors with an appropriate goal can compensate, at least to some extent, for deficits in need for cognition. In our experiment, the judgment quality of lower NFC auditors primed with an accuracy goal approached that of their higher NFC counterparts. This convergence occurred in part because the accuracy goal caused lower NFC auditors to reduce their heuristic support for the client's position, improving auditor skepticism. This suggests that priming accuracy goals will likely improve auditor judgment in other complex areas identified as problematic by regulators, such as revenue recognition and assessing internal controls over financial reporting (IFIAR 2017), and in instances where auditors appear to over-rely on management's assertions and evidence (PCAOB 2018).

Third, this study demonstrates the value of priming, both in its ease of practical application and its contributions to internal validity and comparability of research designs. Regarding practical application, this study suggests that goal priming may be a low-cost and effective way to promote analytical thinking before auditors tackle tasks requiring it. In our study, priming an accuracy goal improved the judgment performance of lower NFC auditors without impacting that of higher NFC auditors. Thus, firms can consider using such a prime with all auditors before they work on complex judgment tasks, for example, by incorporating goal-relevant words into weekly emails or screen savers, or, to better target specific tasks, into warm up exercises such as a word search puzzle or even into passwords to access those tasks in work papers.[27] Regarding internal validity, this study demonstrates the usefulness of a nonconscious

[27] Stajkovic et al. (2018) demonstrate the effectiveness of supraliminal priming in a business setting. In their field experiment, call center employees primed with an achievement goal via key words embedded in an email message from the CEO increased efficiency by 15 percent and effectiveness by 35 percent over a 5-day work week.

prime in making causal inferences. Since the word searches do not contain information that could help participants perform the experimental task, we can more confidently attribute the effects of the manipulation to the theoretical construct invoked by the manipulation. In addition, since the manipulation is independent of the task, other researchers can use the same manipulation without modification, increasing comparability across studies aiming to examine the same theoretical construct.

Finally, this study sheds additional light on the nuanced ways in which auditors' goals affect their judgments. Our nonconscious goal manipulation is designed to prime validity-focused processing directly, and we find different results than prior studies that layer a conscious accuracy goal on top of strong directional goals to support the client (Kadous et al. 2003; Koch and Salterio 2017). Because the primed goal automatically leads to goal-directed action (e.g., Stajkovich et al. 2018), it may be that using a nonconscious goal is more effective because it bypasses the reconciliation process that occurs when auditors struggle consciously with conflicting goals. Thus, future research might examine whether nonconsciously provided goals are superior to conscious ones when conflicting goals exist.

A potential limitation of our study is that we necessarily measure, rather than manipulate, need for cognition, and this potentially reduces our ability to draw causal inferences. That said, we employ several design features to address this issue and strengthen our causal inferences. First, we use a previously validated scale to measure NFC. Extensive research demonstrates that this scale has low associations with other potentially relevant constructs, such as intelligence. We further provide evidence that NFC is not correlated with our manipulated variable or with other variables in our studies. Second, we provide process evidence consistent with theory by measuring participants' justifications for their judgments; variation in these justifications is

consistent with the processing differences in lower and higher NFC auditors predicted by our theory. Third, we predict and find support for an interaction between NFC and provision of an accuracy goal. This moderation-of-process design helps to rule out alternative explanations such as intelligence, rather than NFC, driving the results because the processes associated with an accuracy goal would not systematically improve lower intelligence auditors' justifications or judgments (e.g., Asay et al. 2019). These three factors give us high confidence that NFC, rather than some correlated factor, drives our results.

Appendix A

This appendix reports an analysis of 11 studies in which participants assessed the reasonableness of a biased revenue projection or a biased estimate including multiple assumptions and we measured NFC. The studies encompass three populations (auditors, students, MTurk workers). We collected these data across multiple experiments in which NFC was a relevant construct but not the primary focus of the experiment; as such, these data have not been previously analyzed or reported unless otherwise noted.

Panel A: Index of studies

Study	Participant type	Reasonableness DV	Other measures	N	Date of study
1	Business students at all levels	Revenue projection	NFC, GPA	128	Spring 2012
2	Auditors	Revenue projection	NFC, CPA, Numeracy, Experience, NF Closure, GPA	83	Summer 2012
3	MTurk	Revenue projection	NFC, GPA	101	Spring 2012
4	MTurk	Revenue projection	NFC, GPA	171	Spring 2012
5	MTurk	Revenue projection	NFC, GPA	169	Spring 2012
6	Auditors	Revenue projection	NFC, CPA, Numeracy, Experience, NF Closure, GPA	107	Summer 2012
7	Auditors	Overall estimate	NFC, CPA, Numeracy, Experience, NF Closure, GPA	31	Spring 2013
8[†]	Auditors	Overall estimate	NFC, CPA, Numeracy, Experience, NF Closure, GPA	108	Summer 2013
9	Fifth-year accounting students	Revenue projection	NFC, GPA	77	Spring 2016
10	Auditors	Revenue projection	NFC, Experience, Intrinsic and Extrinsic Motivational Orientation	242	Summer 2016
11[$]	Auditors	Revenue projection	NFC, Experience, Intrinsic and Extrinsic Motivational Orientation	78	Summer 2017

Panel B: NFC by type of participant (using all studies)

	Participant			
	Auditor	Student	MTurk	Combined
Mean NFC	65.43	62.23	65.19	64.84
SD	10.32	10.22	12.93	11.32
n	642	205	441	1288
Mean Difference with Auditor		-3.20	-0.24	
two-tailed *p*-value		< 0.01	0.73	
NFC→Reasonableness DV *t*-stat	(-2.81)	(-0.24)	(-2.19)	(-3.75)
one-tailed *p*-value	< 0.01	0.41	0.01	< 0.01

Panel C: Auditor NFC by experience and CPA (using all auditor studies)

	Experience			
	<= 3 years	> 3 years	Difference	one-tailed *p*-value
Mean NFC	63.98	66.71	-2.73	< 0.01
SD	10.68	9.83		
n	300	342		

	CPA			
	Non-CPA	CPA	Difference	one-tailed *p*-value
Mean NFC	64.72	66.18	-1.46	0.15
SD	10.63	9.79		
n	58	261		

Panel D: Regression analysis with reasonableness as the DV

	Model			
Variable	*(1)*	*(2)*	*(3)*	*(4)*
MTurk	1.334***			
	(10.93)			
Student	1.100***			
	(6.91)			
NFC	**-0.017*** **	**-0.021*** **	**-0.028***	**-0.034*** **
	(-3.45)	**(-2.82)**	**(-1.93)**	**(-3.14)**
Experience		0.011	-.254	0.099
		(0.16)	(-1.64)	(1.25)
CPA			-0.320	
			(-0.90)	
Numeracy			-0.180	
			(-1.53)	
NF Closure			-0.012	
			(-0.92)	
GPA			0.255	
			(-0.61)	
Intrinsic Motivation				0.262
				(1.11)
Extrinsic Motivation				0.008
				(0.12)
Observations	1,287	640	280	319
R-squared	0.104	0.012	0.049	0.033
Studies included	All participants	All auditors	Auditors in Studies 2, 6, 7, 8	Auditors in Studies 10, 11

Variable definitions:
NFC is the score from Cacioppo et al.'s (1984) short-form Need for Cognition Scale (18 items, 1-5 on each item, score is sum); scores can range from 18-90. *Reasonableness* is an assessment on a 0-10 scale of how reasonable the revenue projection or overall estimate is, depending on the study. *MTurk* is an indicator variable equal to one if the participant is an MTurk worker. *Student* is an indicator variable equal to one if the participant is a business student. *Experience* is audit work experience in years. *CPA* is an indicator variable equal to one if the participant has his/her CPA license. *GPA* is self-reported college GPA. *Intrinsic Motivation* and *Extrinsic Motivation* are separate scores from Amabile et al.'s (1994) Work Preference Inventory Scale; 15 items related to intrinsic motivation on a 1-4 scale and 15 items related to extrinsic motivation on a 1-4 scale; scores range from 0 to 4 as score is the average of the 15 responses. *Numeracy* is the score on Cokely et al.'s (2012) Berlin Numeracy Test (four items, 0 or 1 on each item); scores can range from 0-4. *NF Closure* is the score from Roets and Van Hiel's (2011) brief Need for Closure scale (15 items, 1-6 on each item, score is sum); scores can range from 15 to 90.

Additional notes on Panel A:
We thank Jackie Hammersley for permission to use data she was involved in collecting (Studies 1-8).
[†] denotes previously unreported results in data collected as part of a published study (Griffith et al. 2015b).
[$] denotes data collected as part of the current study.

Additional notes on Panel D:
This table presents estimation results from OLS regressions with *Reasonableness* as the dependent variable. We report the coefficient estimate and the t-values in parentheses below. Model (1) uses all studies. Model (2) uses all studies with auditor participants. The additional variables in Models (3) and (4) were never measured in the same study, therefore the models cannot be combined.
***, **, and * denote two-tailed statistical significance of coefficient estimates at the 1, 5, and 10 percent levels, respectively.

References

Amabile, T. M., K. G. Hill, B. A. Hennessey, and E. M. Tighe. 1994. The work preference inventory: Assessing intrinsic and extrinsic motivational orientations. *Journal of Personality and Social Psychology* 66 (5):950-967.

Arkes, H. 1991. Costs and benefits of judgment errors: Implications for debiasing. *Psychological Bulletin* 110 (3): 486-498.

Asare, S. K., and A. M. Cianci. 2009. The effect of goals on auditors' judgments and their perceptions of and conformity to other auditors' judgments. *Managerial Auditing Journal* 24 (8): 724-742.

Asare, S. K., A. M. Cianci, and G. T. Tsakumis. 2009. The impact of competing goals, experience, and litigation consciousness on auditors' judgments. *International Journal of Auditing* 13 (3): 223-236.

Asay, S., R. D. Guggenmos, K. Kadous, L. Koonce, and R. Libby. 2019. Theory Testing and Process Evidence in Accounting Experiments. Working Paper: University of Iowa, Cornell University, Emory University, The University of Texas at Austin.

Austin, A. A., J. S. Hammersley, and M. A. Ricci. 2019. Improving auditors' consideration of evidence contradicting management's estimate assumptions. *Contemporary Accounting Research*, forthcoming.

Backof, A. G., E. M. Bamber, and T. D. Carpenter. 2016. Do auditor judgment frameworks help in constraining aggressive reporting? Evidence under more precise and less precise accounting standards. *Accounting, Organizations and Society* 51: 1-11.

Bar-Anan, Y., T. D. Wilson, and R. R. Hassin. 2010. Inaccurate self-knoweldge formation as a result of automatic behavior. *Journal of Experimental Social Psychology* 46 (6): 884-894.

Bargh, J.A., P. M. Gollwitzer, A. Lee-Chai, K. Barndollar, and R. Trötschel. 2001. The automated will: Nonconscious activation and pursuit of behavioral goals. *Journal of personality and social psychology* 81 (6): 1014-1027.

Bell, T. B., J. C. Bedard, K. M. Johnstone, and E. F. Smith. 2002. KRiskSM: A computerized decision aid for client acceptance and continuance risk assessments. *Auditing: A Journal of Practice and Theory* 21 (2): 97-113.

Bellini-Leite, S. C. 2018. Dual process theory: Systems, types, minds, modes, kinds or metaphors? A critical review. *Review of Philosophy and Psychology* 9: 213-225.

Bennett, G. B., and R. C. Hatfield. 2013. The effect of the social mismatch between staff auditors and client management on the collection of audit evidence. *The Accounting Review* 88 (1): 31-5.

Bhaskar, L. S., P. E. Hopkins, and J. H. Schroeder. 2019. An investigation of auditors' judgments when companies release earnings before audit completion. *Journal of Accounting Research* 57 (2): 355-390.

Blay, A. D. 2005. Independence threats, litigation risk, and the auditor's decision process. *Contemporary Accounting Research* 22 (4): 759-789.

Bonner, S. E. 2008. *Judgment and Decision Making in Accounting*. University of Michigan: Pearson/Prentice Hall.

Bonner, S. E., K. Kadous, and T. Majors. 2019. How strength of habits developed as staff influences senior auditors' evaluation of assumptions underlying an estimate. Working paper, University of Southern California and Emory University.

Bonner, S. E., and B. L. Lewis. 1990. Determinants of auditor expertise. *Journal of Accounting Research* 28 (Supplement): 1-20.

Bors, D. A., F. Vigneau, and F. Lalande. 2006. Measuring the need for cognition: Item polarity, dimensionality, and the relation with ability. *Personality and Individual Differences* 40 (4): 819-828.

Brewster, B. E. 2011. How a systems perspective improves knowledge acquisition and performance in analytical procedures. *The Accounting Review* 86 (3):915-943.

Bucaro, A. C. 2019. Enhancing auditors' critical thinking in audits of complex estimates. *Accounting, Organizations, and Society* 73: 35-49.

Buckless, F. A., and S. P. Ravenscroft. 1990. Contrast coding: A refinement of ANOVA in behavioral analysis. *The Accounting Review* 65 (4): 933-945.

Butler, S. 1985. Application of a decision aid in the judgmental evaluation of substantive test of details samples. *Journal of Accounting Research* 23 (2): 513-526.

Byrne, B. M. 2013. *Structural Equation Modeling with Amos: Basic Concepts, Applications, and Programming*. Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

Cacioppo, J. T. and R. E. Petty. 1982. The need for cognition. *Journal of Personality and Social Psychology* 42 (1): 116-131.

Cacioppo, J. T., R. E. Petty, J. A. Feinstein, and W. B. G. Jarvis. 1996. Dispositional differences in cognitive motivation: The life and times of individuals varying in need for cognition. *Psychological Bulletin* 119 (2): 197 – 253.

Cacioppo, J. T., R. E. Petty, and C. F. Kao. 1984. The efficient assessment of need for cognition. *Journal of Personality Assessment 48 (3):* 306–307.

Cacioppo, J. T., P. E. Petty, and K. J. Morris. 1983. Effects of need for cognition on message evaluation, recall, and persuasion. *Journal of Personality and Social Psychology* 45 (4): 805-818.

Cerasoli, C. P., J. M. Nicklin, and M. T. Ford. 2014. Intrinsic motivation and extrinsic incentives jointly predict performance: A 40-year meta-analysis. *Psychological Bulletin* 140 (4): 980-1008.

Chen, S., D. Shechter, and S. Chaiken. 1996. Getting at the truth or getting along: Accuracy- versus impression-motivated heuristic and systematic processing. *Journal of Personality and Social Psychology* 71 (2): 262-275.

Christensen, B. E., S. M. Glover, T. C. Omer, and M. K. Shelley. 2016. Understanding audit quality: Insights from audit professionals and investors. *Contemporary Accounting Research* 33 (4): 1648-1684.

Cokely, E. T., Galesic, M., Schulz, E., Ghazal, S., & Garcia-Retamero, R. 2012. Measuring risk literacy: The Berlin Numeracy Test. *Judgment and Decision Making, 7*(1): 25-47.

Custers, R., and H. Aarts. 2010. The unconscious will: How the pursuit of goals operates outside of conscious awareness. *Science* 329 (5987): 47-50.

Custers, R., and H. Aarts. 2014. Conscious and unconscious goal pursuit. In *Dual-Process Theories of the Social Mind,* edited by Sherman, J., B. Gawronski, and Y. Trope, 386-399. New York, NY: Guilford Press.

DeZoort, T., P. Harrison, and M. Taylor. 2006. Accountability and auditors' materiality judgments: The effects of differential pressure strength on conservatism, variability, and effort. *Accounting, Organizations and Society* 31 (4-5): 373-390.

Earley, C. E. 2001. Knowledge acquisition in auditing: Training novice auditors to recognize cue relationships in real estate valuation. *The Accounting Review* 76 (1): 81-97.

Evans, J. S. B. T. 2016. Reasoning, biases and dual processes: The lasting impact of Wason (1960). *The Quarterly Journal of Experimental Psychology* 69(10): 2076-2092. DOI: 10.1080/17470218.2014.914547

Evans, J. S. B. T., and K. E. Stanovich. 2013. Dual-process theories of higher cognition: Advancing the debate. *Perspectives on Psychological Science* 8 (3): 223-241.

Fleischhauer, M., S. Enge, B. Brocke, J. Ullruck, A. Strobel, and A. Strobel. 2010. Same or different? Clarifying the relationship of need for cognition to personality and intelligence. *Personality and Social Psychology Bulletin* 36 (1): 82-96.

Glover, S. M., M. H. Taylor, and Y. Wu. 2018. Mind the gap: Why do experts have differences of opinion regarding the sufficiency of audit evidence supporting complex fair value measurements? *Contemporary Accounting Research* (forthcoming).

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., J. S. Hammersley, K. Kadous, and D. Young. 2015b. Auditor mindsets and audits of complex estimates. *Journal of Accounting Research* 53 (1): 49-77.

Griffith, E. E., K. Kadous, and D. Young. 2016. How insights from the 'new' JDM research can improve auditor judgment: Fundamental research questions and methodological advice. *Auditing: A Journal of Practice & Theory* 35 (2): 1-22.

Guggenmos, R. D., M. D. Piercey, and C. P. Agoglia. 2018. Custom contrast testing: Current trends and a new approach. *The Accounting Review* 93 (5): 223-244.

Hackenbrack, K. and M.W. Nelson. 1996. Auditors' incentives and their application of financial accounting standards. *The Accounting Review* 71 (1): 43–59.

Hammersley, J. S. 2006. Pattern identification and industry-specialist auditors. *The Accounting Review* 81 (2):309-336.

Hammersley, J. 2011. A review and model of auditor judgments in fraud-related planning tasks. *Auditing: A Journal of Practice and Theory* 30 (4): 101-128.

Hammersley, J. S., E. M. Bamber, and T. D. Carpenter. 2010. The influence of documentation specificity and priming on auditors' fraud risk assessments and evidence evaluation decisions. *The Accounting Review* 85 (2): 547-571.

Hammersley, J. S., and M. A. Ricci. 2019. Using audit programs to improve auditor evidence collection. Working paper, University of Georgia and University of Florida.

Hart, W., and D. Albarracín. 2009. The effects of chronic achievement motivation and achievement primes on the activation of achievement and fun goals. *Journal of Personality and Social Psychology* 97 (6): 1129-1141.

Haugtvedt, C. P., R. E. Petty, and J. T. Cacioppo. 1992. Need for cognition and advertising: Understanding the role of personality variables in consumer behavior. *Journal of Consumer Psychology* 1 (3): 239-260.

Hauser, D. R., P. C. Ellsworth, and R. Gonzalez. 2018. Are manipulation checks necessary? *Frontiers in Psychology* 9: 998.

Hoffman, V. B., and M. F. Zimbelman. 2009. Do strategic reasoning and fraud brainstorming help auditors change standard audit procedures in response to fraud risk? The Accounting Review 84 (3): 811-837.

Johnstone, K. M., T. D. Warfield, and M. H. Sutton 2001. Antecedents and Consequences of Independence Risk: Framework for Analysis. *Accounting Horizons* 15 (1): 1-18.

International Forum of Independent Audit Regulators (IFIAR). 2017. Report on 2016 survey of inspection findings. Available at: https://www.ifiar.org/?wpdmdl=2055. Accessed 9/7/2019.

International Forum of Independent Audit Regulators (IFIAR). 2019. Report on 2018 survey of inspection findings. Available at: https://www.ifiar.org/?wpdmdl=9603. Accessed 9/7/2019.

Johnson, L. M., M. B. Keune, and J. Winchel. 2019. U.S. auditors' perceptions of the PCAOB inspection process. *Contemporary Accounting Research* (forthcoming).

Kadous, K., S.J. Kennedy, and M.E. Peecher. 2003. The effect of quality assessment and directional goal commitment on auditors' acceptance of client-preferred accounting methods. *The Accounting Review* 78 (3): 759–778.

Kadous, K., and Y. (D.) Zhou 2018. Maximizing the contribution of JDM-style experiments. In T. Libby and L. Thorne (Eds.), *The Routledge Companion to Behavioural Research in Accounting.* Abingdon (UK): Routledge.

Kadous, K., and Y. (D.) Zhou. 2019. How does intrinsic motivation improve auditor judgment in complex audit tasks? *Contemporary Accounting Research* 36 (1): 108-131.

Kahneman, D., and S. Frederick. 2002. Representativeness revisited: Attribute substitution in intuitive judgment. In T. Gilovich, D. Griffin, and D. Kahneman (Eds.), Heuristics and Biases: The psychology of intuitive judgment. Cambridge: Cambridge University Press.

Kallunki, J., J. Kallunki, L. Niemi, and H. Nilsson. 2019. IQ and audit quality: Do smarter auditors deliver better audits? *Contemporary Accounting Research* 36 (3): 1371-1416.

Kennedy, J. 1993. Debiasing audit judgment with accountability: A framework and experimental results. *Journal of Accounting Research* 31 (2): 231-245.

Kennedy, J. 1995. Debiasing the curse of knowledge in audit judgment. *The Accounting Review* 70 (2): 249-273.

Kida, T. 1980. An investigation into auditors' continuity and related qualification judgments. *Journal of Accounting Research* 18 (2): 506-523.

Koch, C., and S. E. Salterio. 2017. The effects of auditor affinity for client and perceived client pressure on auditor proposed judgments. *The Accounting Review* 95 (2): 117-142.

Libby, R., and J. Luft. 1993. Determinants of judgment performance in accounting settings: Ability, knowledge, motivation, and environment. *Accounting, Organizations and Society* 18 (5): 425-450.

Liu, P. and Z. Li. 2012. Task complexity: A review and conceptualization framework. *International Journal of Industrial Ergonomics* 42 (6): 553-568.

Luttrell, A., R.E. Petty, and M. Xu. 2017. Replicating and fixing failed replications: The case of need for cognition and argument quality. *Journal of Experimental Social Psychology* 69: 178-183.

MacCallum, R.C., M. W. Browne, and H. M. Sugawara. 1996. Power analysis and determination of sample size for covariance structure modeling. *Psychological Methods* 1 (2): 130-149.

Majors, T., and S. E. Bonner. 2019. The effects of a refuse to accept goal versus a rubber stamp goal on auditors' fraud risk identification and fraud action. Working paper, University of Southern California.

McDaniel, L. S. 1990. The effects of time pressure and audit program structure on audit performance. *Journal of Accounting Research* 28 (2): 267-285.

Mocadlo, R. 2018. How do auditors order their tasks, and how does task ordering affect performance? Working paper, University of North Dakota.

Moore, D. A., P. E. Tetlock, L. Tanlu, and M. E. Bazerman. 2006. Conflicts of interest and the case of auditor independence: Moral seduction and strategic issue cycling. *Academy of Management Review* 31 (1): 10-29.

Mullis, C. E., and R. C. Hatfield. 2018. The effects of multitasking on auditors' judgment quality. *Contemporary Accounting Research* 35 (1): 314-333.

Myers, J. L., A. D. Well, A. D., and R. F. Lorch, Jr. 2010. *Research Design and Statistical Analysis, 3rd edition.* New York, NY: Routledge.

Nair, U. K., and S. Ramnarayan. 2000. Individual differences in need for cognition and complex problem solving. *Journal of Research in Personality* 34 (3): 305-328.

Nolder, C. J., and K. Kadous. 2018. Grounding the professional skepticism construct in mindset and attitude theory: A way forward. *Accounting, Organizations and Society* 67: 1-14.

Payne, S. J., G. B. Duggan, and H. Neth. 2007. Discretionary task interleaving: Heuristics for time allocation in cognitive foraging. *Journal of Experimental Psychology* 136 (3): 370-388.

Peecher, M. E. 1996. The influence of auditor's justification processes on their decisions: A cognitive model and experimental evidence. *Journal of Accounting Research* 34 (1): 125–140.

Peltier, J. W., and J. A. Schibrowsky. 1994. Need for cognition, advertisement viewing time and memory for advertising stimuli. *Advances in Consumer Research* 21 (1): 244-250.

Pennycook, G. 2018. A perspective on the theoretical foundations of dual-process models. In W, De Neys (Ed.), *Dual Process Theory 2.0.* London: Routledge.

Pennycook, G., W. De Neys, J. Evans, K. Stanovich, and V. Thompson. 2018. The mythical dual-process typology. *Trends in Cognitive Sciences*, 22 (8): 667-668.

Petty, R. E., D. T. Wegener, and L. R. Fabrigar. 1997. Attitudes and attitude change. *Annual Review of Psychology* 48: 609-647.

Petty, R. E., S. C. Wheeler, and Z. L. Tormala. 2003. Persuasion and attitude change. In T. Millon & M. J. Lerner (Eds.), *Handbook of psychology: Volume 5: Personality and social psychology* (pp. 353-382). Hoboken, NJ: John Wiley & Sons.

Petty, R. E., P. Brinol, C. Loersch, and M. J. McCaslin. 2009. The need for cognition. In *Handbook of Individual Differences in Social Behavior*, edited by M. R. Leary and R. H. Hoyle, 318-329. New York, NY: The Guilford Press.

Plumlee, R. D., B. A. Rixom, and A. J. Rosman. 2014. Training auditors to perform analytical procedures using metacognitive skills. *The Accounting Review* 90 (1):351-369.

Public Company Accounting Oversight Board (PCAOB). 2015. *Inspection Observations Related to PCAOB "Risk Assessment" Auditing Standards (No. 8 through No. 15).* PCAOB Release No. 2015-007. Washington, D.C.: PCAOB.

Public Company Accounting Oversight Board (PCAOB). 2018. *Auditing accounting estimates, including fair value measurements, and amendments to PCAOB auditing standards.* PCAOB Release No. 2018-005. Washington, D.C.: PCAOB.

Ricci, M. A. 2019. How higher performance on client service affects auditors' willingness to challenge management's preferred accounting. Working paper, University of Florida.

Roets, A., and A. Van Hiel. 2011. Item selection and validation of a brief, 15-item version of the need for closure scale. *Personality and Individual Differences* 50: 90-94.

Sandra, D. A., and A. R. Otto. 2018. Cognitive capacity limitations and need for cognition differentially predict reward-induced cognitive effort expenditure. *Cognition* 172: 101-106.

Setiawan, W. Y. 2017. Need for cognition: Does it influence professional judgment? *Review of Integrative Business & Economics* 6 (1): 240-248.

Shanteau, J. 1989. Cognitive heuristics and biases in behavioral auditing: Review, comments, and observations. *Accounting, Organizations and Society* 14 (1/2): 165-177.

Sheeran, P., T. L. Webb, and P. M. Gollwitzer. 2005. The interplay between goal intentions and implementation intentions. *Personality and Social Psychology Bulletin* 31 (1): 87-98.

Simon, C. A. 2012. Individual auditors' identification of relevant fraud schemes. *Auditing: A Journal of Practice & Theory* 31 (1): 1-16.

Stajkovic, A. D., G. P. Latham, K. Sergent, and S. J. Peterson. 2018. Prime and performance: Can a CEO motivate employees without their awareness? *Journal of Business and Psychology* 28 (3): 1-12.

Stanovich, K. E., and R. F. West. 2008. On the relative independence of thinking biases and cognitive ability. *Journal of Personality and Social Psychology* 94 (4): 672-695.

Strobel, A., M. Fleischhauer, C. Luong, and A. Strobel. 2018. Predicting everyday life behavior by direct and indirect measures of need for cognition. *Journal of Individual Differences* 39 (2): 107-114.

Tan, H., and R. Libby. 1997. Tacit managerial knowledge versus technical knowledge as determinants of audit expertise in the field. *Journal of Accounting Research* 35 (1): 97-113.

Tidwell, P. S., C. J. Sadowski, and L. M. Pate. 2000. Relationships between need for cognition, knowledge, and verbal ability. *The Journal of Psychology* 134 (6): 634-644.

Vansteenkiste, M., J. Simons, W. Lens, K. M. Sheldon, and E. L. Deci. 2004. Motivating learning, performance, and persistence: The synergistic effects of intrinsic goal contents and autonomy-supportive contexts. *Journal of Personality and Social Psychology* 87 (2): 246-260.

Verplanken, B., P. T. Hazenberg, and G. R. Palenewen 1992. Need for cognition and external information search effort. *Journal of Research in Personality* 26 (2): 128-136.

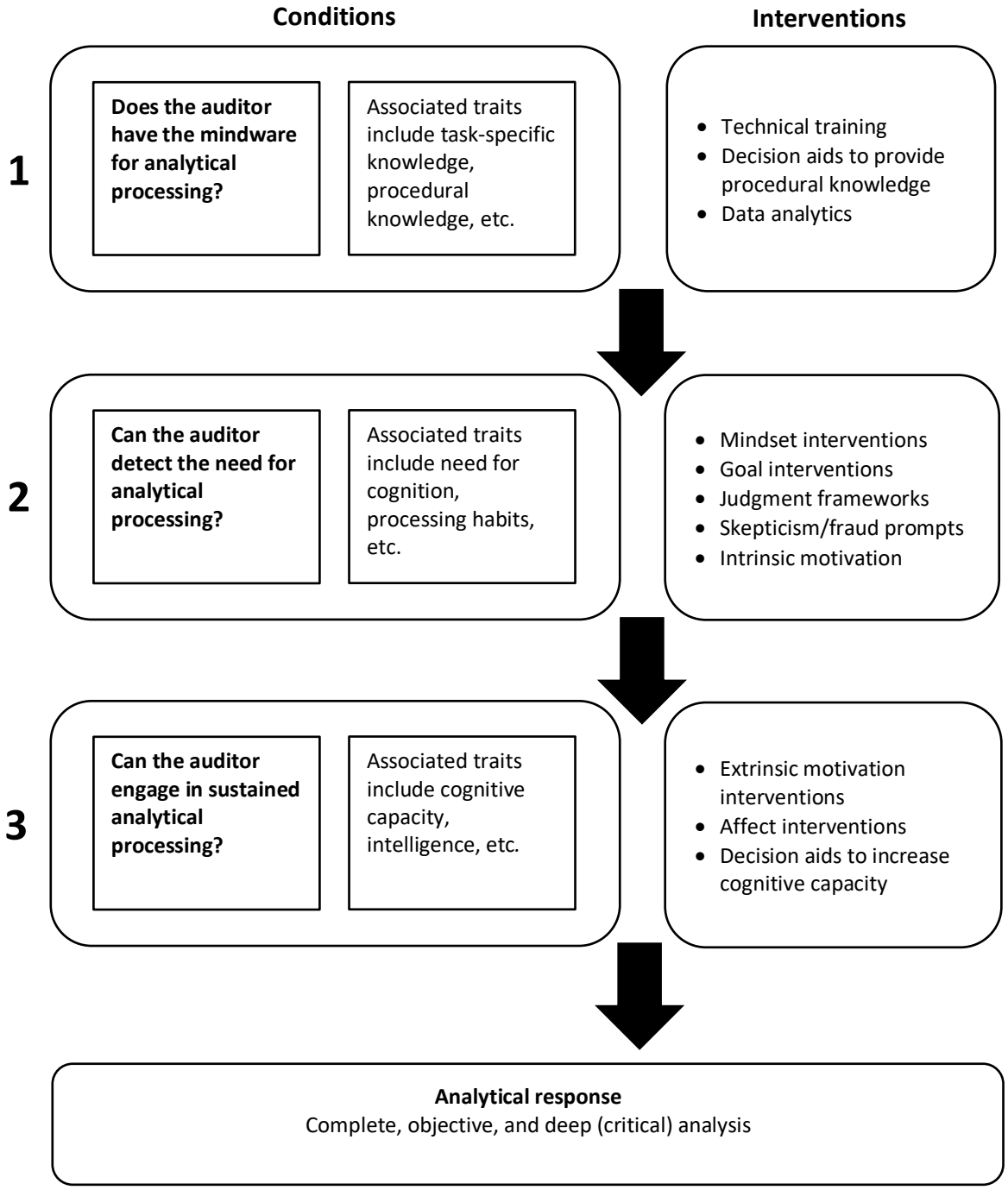
Viator, R. E., P. L. Bagley, B. G. Barnes, and N. L. Harp. 2014. Measuring reflective cognitive capacity: A methodological recommendation for accounting research of feedback effects. *Behavioral Research in Accounting* 26 (2): 131-160.

Weber, E., and E. Johnson. 2009. Mindful judgment and decision making. *Annual Review of Psychology* 60: 53-85.

Wilks, T.J. 2002. Predecisional distortion of evidence as a consequence of real-time audit review. *The Accounting Review* 77 (1): 51-71.

Zhou, Y. (D.). 2019. Counteracting the directional influence of incentives on auditor judgment. Working paper. University of Illinois.

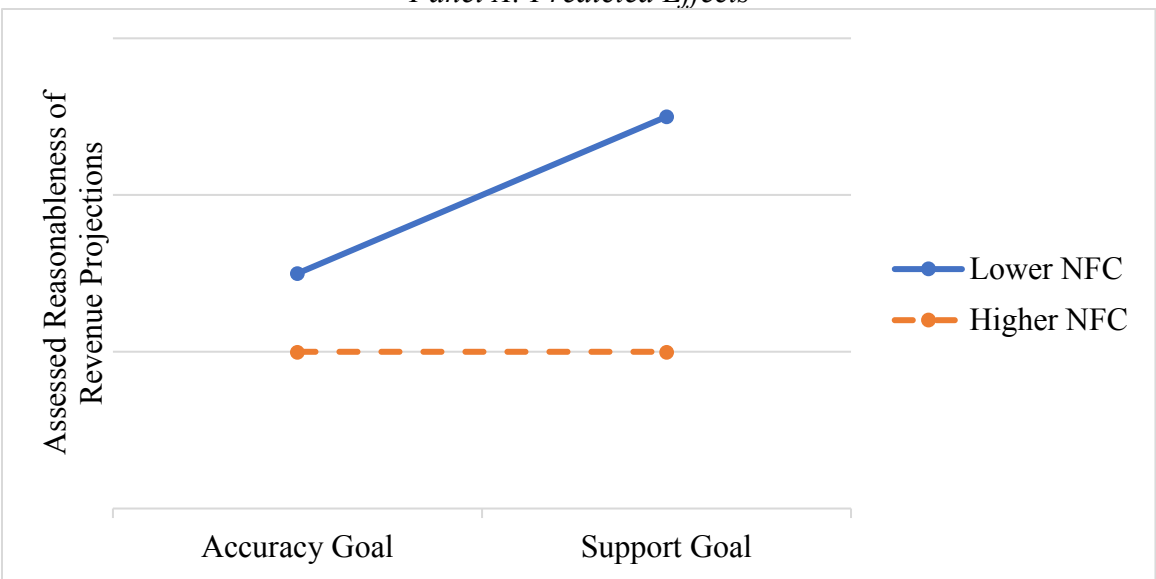
Figure 1 Conditions and interventions for high-quality judgments in complex audit tasks



The traits and interventions included in the figure above are intended as selected examples to clarify the framework. Other traits and interventions can also address the conditions for high-quality complex audit judgments.

Figure 2 The effect of need for cognition and goal on assessed reasonableness of the revenue projections

Panel A: Predicted Effects



Panel B: Observed Effects

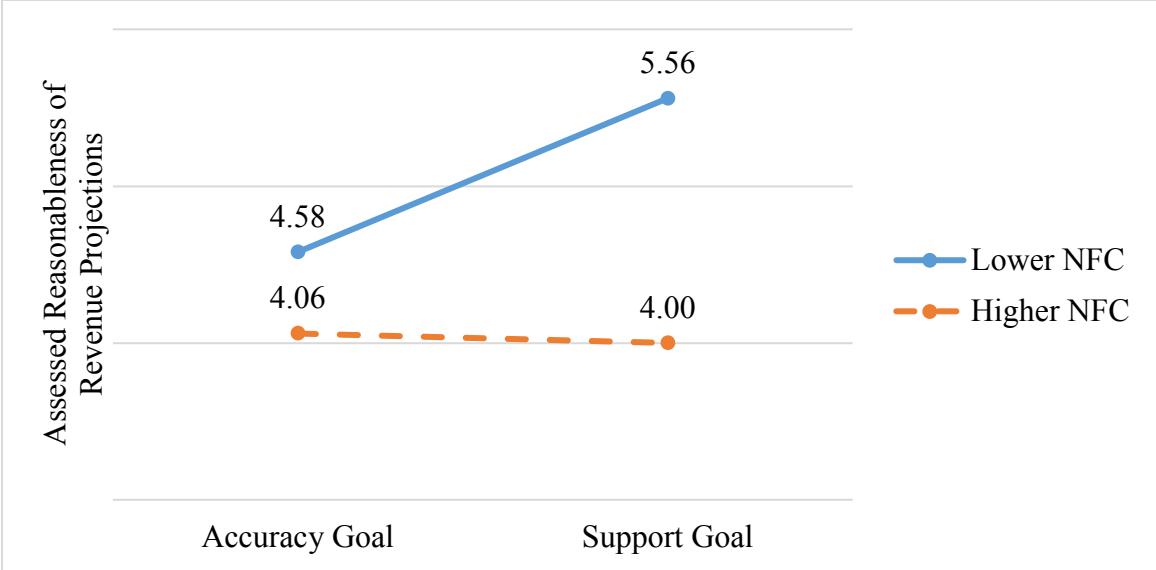
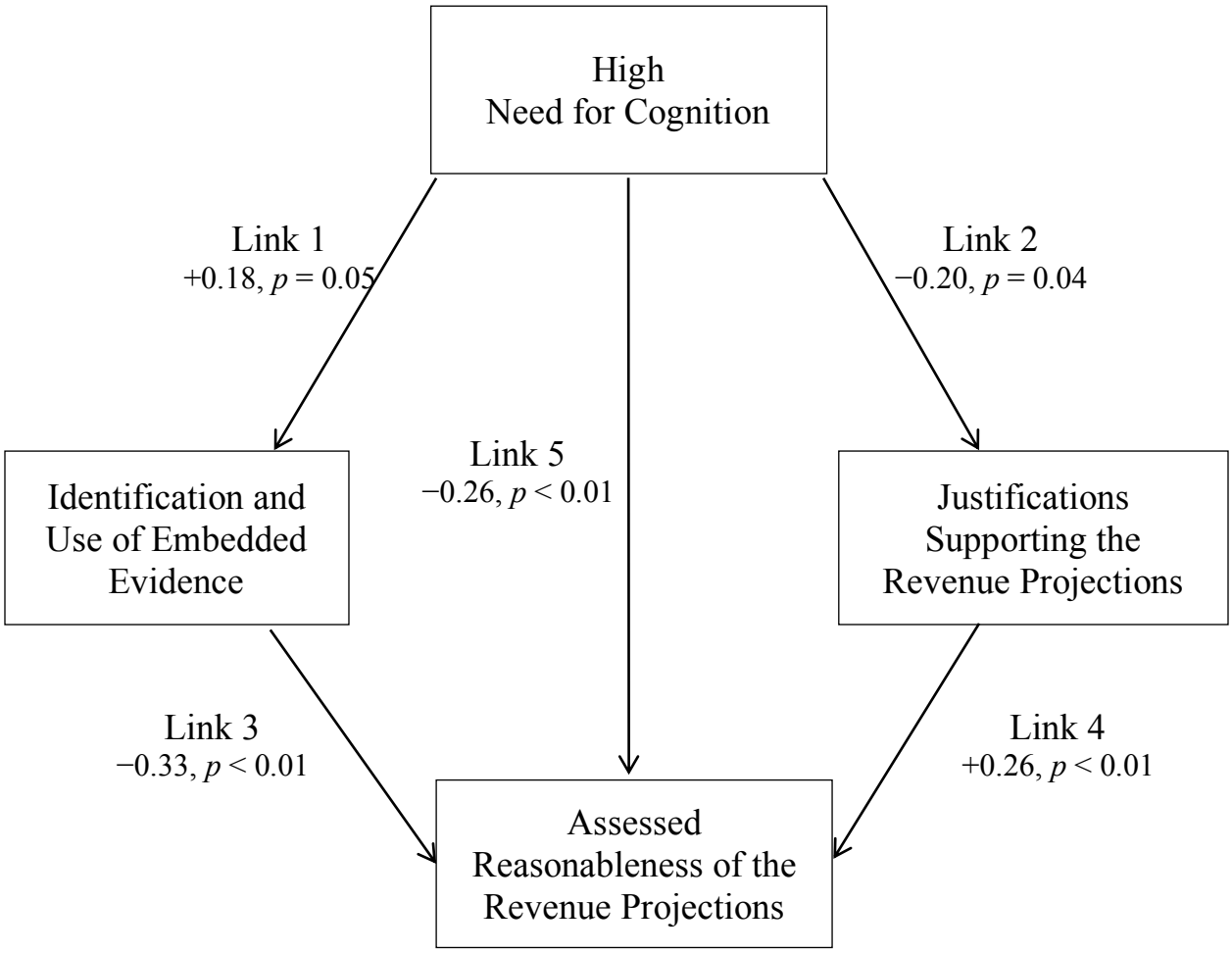
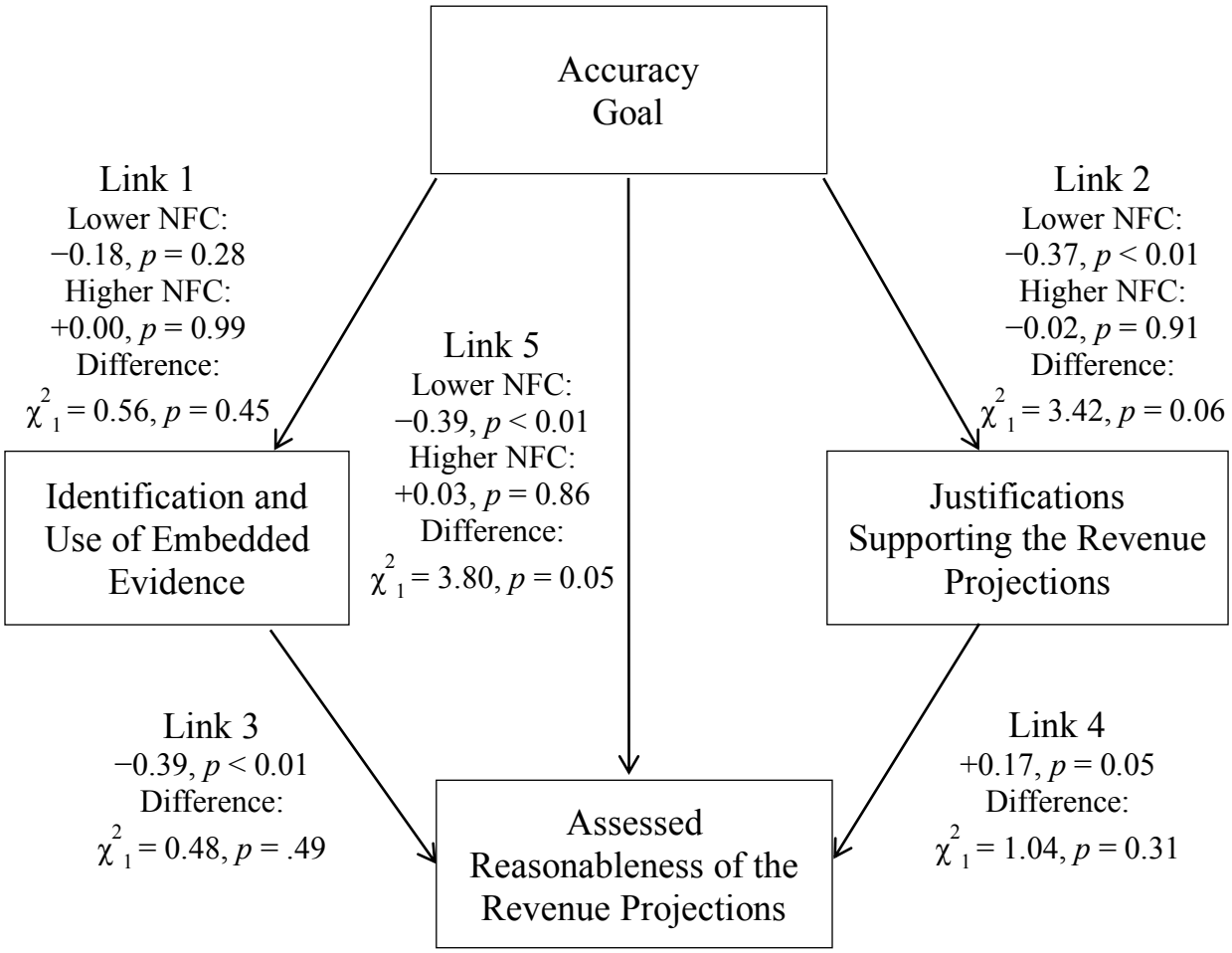


Figure 3 Structural equations model for the effects of need for cognition on reasonableness assessments



This figure summarizes how identification and use of embedded contradictory evidence and reasons supporting the projections jointly mediate the effect of need for cognition and goal on assessed reasonableness of the revenue projections using a structural equation model. The model fits the data well ($\chi^2_1 = 0.11$, p = 0.74, Comparative Fit Index = 1.00, Root Mean Square Error of Approximation = 0.00). All p-values are one-sided, reflecting directional predictions.

Figure 4 Structural equations model for the effects of goal on reasonableness assessments, conditional on level of need for cognition



PROCESS Tests of Indirect Effects of Model Mediators

Meditator: *Accuracy Goal* → *Identification and Use of Embedded Evidence* → *Reasonableness Assessment*
Indirect Effect When Lower NFC: 90% CI of Indirect Effect (-0.500, 0.087)
Indirect Effect When Higher NFC: 90% CI of Indirect Effect (-0.340, 0.287)
Index of Moderated Mediation (Difference Between Conditional Indirect Effects) ***90% CI (-0.233, 0.607)***

Meditator: *Accuracy Goal* → *Justifications Supporting the Revenue Projections* → *Reasonableness Assessment*
Indirect Effect When Lower NFC: 90% CI of Indirect Effect (0.056, 0.653)
Indirect Effect When Higher NFC: 90% CI of Indirect Effect (-0.181, 0.165)
Index of Moderated Mediation (Difference Between Conditional Indirect Effects) ***90% CI (-0.715, -0.013)***

This figure summarizes how identification of reasons supporting the revenue projections solely mediates the effect of goal on assessed reasonableness of the revenue projections, when in the lower NFC condition, using a nested structural equation model. The model fits the data well (χ^2_4 = 2.82, p = 0.59, Comparative Fit Index = 1.00, Root Mean Square Error of Approximation = 0.00).

All p-values for higher NFC links are two-sided and all p-values for joint or lower NFC links are one-sided, reflecting directional predictions. PROCESS testing was conducted using the PROCESS package for SPSS (Hayes 2018), 90% confidence intervals for the indirect effect and index of moderated mediation are estimated using 5,000 bootstrapped re-samples (with replacement) and are biased corrected. Intervals excluding zero indicate significance.

TABLE 1
Effects of need for cognition and goal on assessed reasonableness of the revenue projections

Panel A: Descriptive Statistics – mean (standard deviation) [n]

	Accuracy Goal	Support Goal
Lower NFC	4.58 (1.07) [19]	5.56 (1.31) [16]
Higher NFC	4.06 (1.43) [17]	4.00 (1.15) [19]

Panel B: Analysis of Variance

Source	*df*	*SS*	*MS*	*F*	*p-value*
NFC	1	19.14	19.14	12.39	< 0.001
Goal	1	3.77	3.77	2.44	0.123
NFC × Goal	1	4.80	4.80	3.10	0.083
Error	67	103.51	1.53		

Panel C: Simple Main Effects

	df	*F*	*p-value*
Effect of NFC given Accuracy goal	1, 67	1.57	0.214
Effect of NFC given Support goal	1, 67	13.73	< 0.001
Effect of Goal given Higher NFC	1, 67	0.02	0.888
Effect of Goal given Lower NFC	1, 67	5.44	0.022

Panel D: Planned Contrast: Joint Test of H1 and H2

Source	*df*	*SS*	*MS*	*F*	*p-value*
Contrast [−1,+3,−1,−1]*†	1	22.57	22.57	14.61	< 0.001
Error	67	103.51	1.54		

The dependent variable measures auditors' responses to "…how likely is it that management's revenue projections are reasonable?" on a scale of 0 (not likely at all) to 10 (extremely likely). The revenue projections were unreasonable considering the information provided about economic indicators and past firm performance; smaller numbers represent better decision quality. Goal was manipulated as accuracy (when the word search contained the following words: delve, dig, drive, examine, explore, flower, perch, probe, question, scrutinize, table, vase and walk) or support (when the word search contained the following words: affirm, approve, confirm, drive, endorse, flower, perch, support, table, verify, validate, vase and walk). NFC was measured and scored in accordance with Cacioppo et al. (1984). NFC is coded as higher when the NFC score was greater than or equal to the median NFC score of 65 and lower otherwise

†We use the following contrast weights (Buckless and Ravenscroft 1990) to perform a joint test of H1 and H2: −1 (Lower NFC/Accuracy; Cell A), +3 (Lower NFC/Support; Cell B), −1 (Higher NFC/Accuracy; Cell C), −1 (Higher NFC/Support; Cell D). Thus, our test is: (Cell B) > (Cell A + Cell C + Cell D).