COMPLEX ESTIMATES AND AUDITOR RELIANCE
ON ARTIFICIAL INTELLIGENCE

ABSTRACT

Audit firms are investing millions of dollars to develop artificial intelligence (AI) systems that will help auditors execute challenging tasks (e.g., evaluating complex estimates). Audit firms assume AI will enhance audit quality. However, a growing body of research documents "algorithm aversion" – the tendency for individuals to discount computer-based advice more heavily than human advice, although the advice is identical otherwise. Auditor susceptibility to algorithm aversion could prove costly for the profession and financial statements users. Accordingly, we examine how algorithm aversion manifests in auditor decisions using an experiment that manipulates the source of contradictory audit evidence (human specialist versus AI specialist system) and the degree of structure within the client's estimation process (higher versus lower) for a complex estimate. Consistent with theory, we find evidence that algorithm aversion amplifies the persuasive effect of greater estimation structure, making auditors more likely to discount contradictory audit evidence and accept management's preferred estimates.

1. Introduction

Audit firms are making substantial investments in advanced technologies with the goal of enhancing the effectiveness, efficiency, and decision-usefulness of audits. One of the most promising advanced technologies under consideration is the application of machine learning or artificial intelligence (AI) on audit engagements. Industry experts estimate that each of the 'Big 4' firms currently invest $250 million per year on AI and machine learning technology (Alliott [2017]). AI can synthesize large amounts of diverse and unstructured data, and some firms are harnessing these abilities to help auditors perform tasks that have traditionally been performed by human specialists, such as evaluating complex accounting estimates (e.g., commercial loan grades; KPMG [2016]). The audit profession believes that applying advanced technologies such as AI to the audit setting will enhance audit quality and provide significant benefits for auditors and clients (FEI [2017], EY [2018]). These benefits, however, will only materialize if auditors consider and incorporate the information produced by such systems into their evidence evaluation. Therefore, this study examines when and how receiving contradictory evidence from a firm's AI system (i.e., "specialist system") – rather than a firm's human specialist – influences auditors' evaluations of management's complex estimates.[1]

A growing body of research finds that individuals often exhibit "algorithm aversion" – the human tendency to discount advice from algorithms and rely more readily on human input, as compared to computer-generated input (e.g., Önkal, Goodwin, Thompson, Gonul, and Pollock [2009], Eastwood, Snook, and Luther [2012], Dietvorst, Simmons, and Massey [2015], Yeomans, Shah, Mullainathan, and Kleinberg [2017]). For example, identical stock forecasting

---

[1] Following the Brookings Institution (e.g., West and Allen [2018]), our concept of AI incorporates intentionality, intelligence (i.e., "machine learning"), and adaptability. Consistent with audit firms intending to use their in-house AI systems similar to the way they currently use in-house "human" valuation specialists (e.g., KPMG [2016]), we hereafter refer to these as "specialist systems" and "human specialists", respectively.

advice has a greater influence on individuals' predictions when they believe that advice comes from a human instead of a computer-based model (Önkal et al. [2009]). Additionally, algorithm aversion persists even when individuals receive feedback that the algorithmic predictions are more accurate than their own predictions (Dietvorst et al. [2015]). Research also finds that algorithm aversion occurs in highly subjective settings (Yeomans et al. [2017]).

The unique features of the audit environment make it uncertain whether and how algorithm aversion might manifest in auditor judgments and in the specific context of this study – auditors' evaluations of complex estimates. Auditors lack sufficient expertise to perform some specialized tasks on their own, and in those cases, they are encouraged to rely on advice from experts (Martin, Rich, and Wilks [2006], PCAOB [2017a], PCAOB [2017b]). Additionally, due to the high degree of subjectivity and potential for management bias, auditors have strong incentives to rely on evidence from their firm's specialists, particularly when that evidence contradicts management's estimates. Thus, auditors might be quite willing to rely on audit evidence from AI-based specialist systems. Interestingly, some regulators have expressed concerns that auditors might over-rely on advanced audit technologies (Harris [2017]). Yet, at the same time, regulators have criticized auditors for under-relying on human specialists (e.g., PCAOB [2017a], Griffith [2018]). Therefore, the potential impact of algorithm aversion on auditor judgments remains an open research question. Furthermore, the effects of algorithm aversion might manifest uniquely in the audit setting.

When evaluating complex estimates, auditors evaluate their firm's evidence in conjunction with management's evidence. Accordingly, we also examine how algorithm aversion influences the way auditors respond to persuasive attributes of management's evidence. Auditors often evaluate the reasonableness of management's complex estimates by testing

management's estimation process (Griffith, Hammersley, and Kadous [2015]). We propose that the degree of structure within management's estimation process will influence auditors' evaluations of complex estimates. Following research on task structure and audit structure (e.g., Abdolmohammadi [1999], Hyatt and Prawitt [2001]), we define a more-structured estimation process as one with well-defined steps, fewer alternative inputs, and fewer potential solution paths, which constrains judgment and increases consistency in the estimation process. Alternatively, a less-structured process is one that involves fewer explicit steps and more alternative inputs, requiring more judgment and creating a more varied set of potential solution paths. Research finds that auditors prefer well-defined tasks and information that is more objective, quantified, and verifiable (Bamber, Snowball, and Tubbs [1989], Joe, Vandervelde, and Wu [2017]). Accordingly, we propose that greater estimation structure is a persuasive attribute of management's estimates that can sway auditors toward evaluating management's estimates and related evidence more favorably.

Research documents that individuals use a relative-weighting process to evaluate competing information (e.g., Birnbaum [1976], Birnbaum, Wong, and Wong [1976]). As the persuasiveness of one piece of information increases, that piece of information is more heavily weighted and there is a corresponding (and proportionate) decline in the weighting of opposing information (Birnbaum and Stegner [1979]; Birnbaum and Mellers [1983]). We theorize that algorithm aversion will exacerbate the degree to which persuasive evidence attributes lead to a tradeoff in the weighting of competing information. That is, we propose persuasive information cues will have a greater influence on judgments when weighed against competing information from a computer-based source versus a human source. Accordingly, we predict that the persuasive influence of greater estimation structure will be amplified when auditors receive

3

contradictory evidence from a specialist system instead of a human specialist, leading auditors to judge the client's balance as more reasonable and propose smaller adjustments.

We conduct an experiment with 170 audit senior participants, manipulating the source of firm-provided evidence (human specialist versus specialist system) and the degree of structure in the client's estimation process (higher versus lower). Participants in all conditions receive the same audit evidence from their own firm's specialist regarding a banking client's allowance for loan losses ("ALL"). This evidence suggests that the client's ALL is understated (i.e., net income is overstated). The source of the firm-provided contradictory evidence is either the in-house valuation group (i.e., human specialist) or the proprietary AI system (i.e., specialist system); we use identical language to describe the accuracy, reliability, and calibration of the evidence in both source conditions. In the lower structure condition, management's estimation process relies on the judgement of loan officers and credit analysts, who use a variety of methods to develop estimates for a key input (collateral values) for the ALL estimate. In the higher structure condition, management's estimation process relies on client-selected, detailed market data to update collateral values in a uniform and systematic manner. Auditors' consideration of available audit evidence forms the basis for their beliefs about whether and to what extent management's estimates should be adjusted. Therefore, participants' proposed audit adjustments serve as our dependent measure.

Consistent with our theory-based expectations, we find that persuasive attributes of management's evidence cause auditors to weight their own firm's contradictory evidence differently, depending on whether it comes from a human specialist or specialist system. Specifically, higher (versus lower) estimation structure in management's process leads to lower proposed audit adjustments when the audit firm's contradictory evidence comes from a specialist

system instead of a human specialist. Additional analyses reveal that auditors judge the quality of management's evidence more favorably when it is higher in estimation structure, regardless of the source of the firm's contradictory evidence (a specialist system or human specialist). However, consistent with theory, these more favorable perceptions of management's evidence quality lead to smaller proposed adjustments when the firm's evidence comes from a specialist system instead of a human specialist.

Our study extends two streams of research. First, we contribute to a growing body of research in psychology and management science documenting individuals' reluctance to allow computer-generated input to substitute for human advice (i.e., algorithm aversion). This study is the first to provide evidence of algorithm aversion in auditor judgments and it also demonstrates that in the audit setting, algorithm aversion is more nuanced than was documented in the prior literature. Specifically, algorithm aversion was only triggered when auditors weighed system evidence against client evidence from a more-structured estimation process, but when the subjectivity in client evidence was more obvious (i.e., less-structured estimation process), auditors' judgments were not susceptible to algorithm aversion. Thus, it is perhaps encouraging that auditors only exhibit algorithm aversion when they perceive that the evidence supporting management's estimate appears stronger. Nevertheless, it is still worrisome that algorithm aversion might make it easier to sway auditors into thinking that potentially biased estimates are fairly stated.

Second, we contribute to the growing literature around complex accounting estimates. Measurement uncertainty continues to be a critically important risk for financial reporting stakeholders (Bratten, Gaynor, McDaniel, Montague, and Sierra [2013], Dennis, Griffin, and Johnstone [2018]). Additionally, due to the high degree of measurement uncertainty and

subjectivity, appropriately evaluating complex estimates remains a challenge for auditors (Cannon and Bedard [2017], Joe et al. [2017]). Recent findings indicate that auditors are willing to discount (and perhaps even ignore) contradictory evidence from valuation specialists (PCAOB [2017a], Griffith [2019]). We contribute to this literature by identifying estimation structure as an attribute of management's estimates that can contribute auditors' propensity to discount contradictory evidence, especially when that evidence comes from a specialist system.

Our findings are relevant to audit firms and regulators. The audit environment is quickly evolving due to technological advancement (Raphael [2017], Tysiac [2017]). We provide evidence that the implementation of advanced specialist systems could alter auditor judgments in a way that inadvertently undermines audit quality. Though it is important to evaluate the reliability and appropriateness of new audit tools, it is equally important to consider how auditors will interact with and incorporate evidence from these tools on their engagements.

2. Theory and Hypothesis Development

2.1 BACKGROUND

Audit firms are making significant investments in advanced technologies such as data analytics, drones, and robotic process automation (PwC [2017], Austin, Carpenter, Christ, and Nielson [2018], Christ, Emett, Summers, and Wood [2019]). Some firms have deployed proprietary applications that streamline audit processes and enable the use of mobile devices to collect audit evidence (e.g., for inventory observations; Deloitte [2018]). Other firms are in early stages of implementing robotic process automation for routine and simple audit tasks, such as debt and cash confirmations (Cooper, Holderness, Sorensen, and Wood [2018]). One of the most advanced technologies under consideration is the incorporation of AI on audit engagements.

Audit firms plan to use AI to assist auditors with some of the most challenging audit tasks, such as evaluating management's complex estimates (KPMG [2016], Murphy [2017]). For

example, KPMG is developing an AI system to help auditors evaluate commercial loan grades (KPMG [2016]). With its ability to integrate and process large amounts of diverse and unstructured data through machine learning, AI is well-suited to help auditors evaluate management's assumptions and develop independent estimates. Therefore, as audit firms continue developing AI systems to produce evidence around complex estimates – a task traditionally performed by human specialists – it is critical to understand how auditors will respond to the evidence provided by these AI systems.[2]

2.2 AUDITORS AND ALGORITHM AVERSION

A substantial body of research from psychology, medicine, and management science documents the human tendency to prefer and rely more on information when it comes from a human source rather than when it comes from a computer source (Promberger and Baron [2006], Önkal et al. [2009], Eastwood et al. [2012], Dietvorst et al. [2015], Yeomans et al. [2017]). For example, Önkal et al. [2009] show that when forecasting stock prices, individuals are more likely to discount computer-generated input than human input – although the information provided by the two sources is otherwise identical. Similarly, when predicting student performance, individuals prefer to rely on their own predictions (or predictions from another person) rather than predictions produced by an algorithm, even after receiving feedback that the algorithm's predictions are consistently more accurate than their own (Dietvorst et al. [2015]). The literature refers to this tendency to discount computer-generated "advice" in favor of human advice as algorithm aversion.

---

[2] It is important to note that AI systems differ from decision aids in that they produce audit evidence by simulating human judgment – and auditors are expected to incorporate this evidence into their decisions. Thus, prior accounting research that examined auditors' use of decision aids (e.g., checklists to promote adherence to accounting standards or firm policies) on fraud risk assessments or internal controls testing tasks do not apply to auditor's use of AI (e.g., Kachelmeier and Messier [1990], Messier [1995], Glover, Prawitt, and Spilker [1997], Anderson, Moreno, and Mueller [2003], Asare and Wright [2004]).

Research on algorithm aversion typically examines objective prediction tasks (e.g., forecasting student performance) with observable, correct answers (e.g., realized GPAs). [3] However, recent findings indicate that algorithm aversion also occurs in subjective settings where gaining consensus on the correct answer(s) is more difficult (Yeomans et al. [2017]). Specifically, Yeomans et al. [2017] find that although an algorithmic joke recommendation system outperforms human recommenders, individuals are reluctant to rely on advice from the system when making joke recommendations to others or when receiving joke recommendations for themselves. Yeomans et al.'s [2017] findings are particularly relevant to our experimental context (i.e., auditing complex estimates) because they demonstrate that algorithm aversion can persist in highly subjective domains where accuracy is ill-defined.

Auditing management's complex estimates is challenging because the estimates are highly subjective, lack objectively correct answers, and there is limited opportunity for timely outcome feedback (Martin et al. [2006], Christensen, Glover, and Wood [2012]; Bratten et al. [2013], Griffith [2018]). As a means of reducing the risk and uncertainty associated with complex estimates, it is common for firm-employed valuation specialists to assist auditors with their evaluations of these estimates (PCAOB [2015, 2017a], Cannon and Bedard [2017]). Importantly, due to the high degree of subjectivity associated with complex estimates, auditors often receive evidence from their firm specialists that contradicts management's estimates (Griffith et al. [2015a], Cannon and Bedard [2017], Griffith [2019]). The rapid advancement and planned use of audit technologies point to a future in which auditors are considering contradictory evidence produced by their firm's specialist system instead of a human specialist.

---

[3] Some research suggests that algorithm aversion occurs because individuals expect computer-generated information to be perfect. Thus, when individuals observe any inaccuracies in highly reliable (but not perfect) computer-generated advice, they will discount this advice more than human advice (Dzindolet, Pierce, Beck and Dawe [2002], Madhavan and Wiegmann [2007], Dietvorst et al. [2015], Prahl and Van Swol [2017]).

While firms anticipate that these specialist systems will improve audit outcomes, the literature on algorithm aversion suggests that auditors will weight their own firm's evidence differently (e.g., discounting contradictory evidence) when it comes from a specialist system instead of a human specialist.

However, there are important reasons why algorithm aversion effects might be more nuanced in the audit setting when compared to contexts examined in prior studies. First, regulators have expressed concerns that auditors will over-rely on advanced audit technologies (Harris [2017]). Consistent with this concern, prior research finds that individuals are more willing to rely on advice from others (i.e., humans) as task complexity increases (Schrah, Dalal, and Sniezek [2006], Gino and Moore [2007]). Due to the difficulty associated with auditing complex estimates, for which auditors often lack sufficient expertise to evaluate (Martin et al. [2006], Bratten et al. [2013], PCAOB [2015, 2017a, 2017b], Griffith [2019]), auditors might be more willing to rely on evidence from specialist systems than the participants in prior algorithm aversion studies. Second, complex estimates are particularly vulnerable to management bias (Bratten et al. [2013]) and regulators continue to caution auditors to evaluate the risk of management bias in the subjective aspects of estimation processes (e.g., PCAOB [2007], PCAOB [2010a], Hanson [2012], PCAOB [2018 p. 1]). Thus, auditors have significant legal and regulatory incentives to avoid discounting contradictory evidence from their firms' expert systems (e.g., PCAOB [2010b, 2015]). Therefore, we consider how algorithm aversion may manifest in a more nuanced manner in the context of auditing complex estimates.

2.3 AUDITORS' CONSIDERATION OF COMPETING INFORMATION AND ALGORITHM AVERSION

While theory from the algorithm aversion literature generally suggests that auditors will rely less on evidence that comes from a specialist system (instead of a human specialist), auditors

do not evaluate evidence from firm specialists in isolation. Auditors consider any contradictory evidence obtained from firm specialists in conjunction with other evidence they have gathered, including evidence provided by management, when forming their own opinion about whether management's estimates are fairly stated (PCAOB [2010a, 2014a, 2014b], Hanson [2012], CPAB [2015a, 2015b], IFIAR [2015], PCAOB [2017a]). When evidence from these sources are in conflict, auditors must use professional judgment to determine how to incorporate the competing evidence into their decision making (PCAOB [2015, 2017a, 2017b]).

Research indicates that individuals use a relative-weighting process to reconcile conflicting information (Birnbaum [1976], Birnbaum et al. [1976]). In this relative-weighting process, as the persuasive strength of one piece of evidence increases, that information is more heavily weighted and there is a proportionate decline in the weighting of competing information (Birnbaum and Stegner [1979], Birnbaum and Mellers [1983]). Thus, persuasive information attributes can prompt a tradeoff in the influence of two competing information sources. Based on prior algorithm aversion research, we further theorize that the magnitude of this tradeoff will depend on whether the opposing information comes from a computer system or a human source. Specifically, we expect attributes that can sway an individual that one perspective is correct or reasonable will have a greater persuasive influence on judgments when the opposing perspective (or information) comes from an algorithm instead of a human.

Following this, we expect algorithm aversion to manifest in the audit environment by amplifying the degree to which persuasive attributes of management's evidence influence auditors' related judgments. In the context of auditing complex estimates, we expect the persuasive attributes of management's evidence will influence auditor judgments more strongly when the firm's contradictory evidence comes from a specialist system instead of a human

10

specialist. Prior accounting research has identified many evidence attributes (e.g., amount, consistency, congruency with management incentives, quantification) that are persuasive in accounting and auditing decision contexts (Caster and Pincus [1996], Goodwin [1999], Kadous, Koonce, and Towry [2005], Kaplan, O'Donnell, and Arel [2008], Joe et al. [2017]). For example, Joe et al. [2017] demonstrate that a higher degree of quantification in information can increase auditors' comfort level with the reasonableness of management's subjective estimates, even if that quantification does not provide new diagnostic evidence. Although there are several attributes of audit evidence that can have a persuasive influence on auditor judgments, we focus on an important contextual feature of management's estimation process that has not been examined by prior research – the degree of structure within management's estimation process (i.e., "estimation structure").

2.4 COMPLEX ESTIMATES AND ESTIMATION STRUCTURE

Unlike account balances based on historical cost, where relatively more objective audit evidence is available, complex estimates are inherently ambiguous, uncertain, and lack verifiability (Bratten et al. [2013]). The task of evaluating the reasonableness of management's complex estimates is correspondingly difficult for auditors (Christensen et al. [2012], Glover, Taylor, and Wu [2019]). Research finds that auditors typically approach this audit task by testing management's estimation process (Griffith et al. [2015a]). However, accounting standards allow for a variety of estimation methods, and even highly trained experts can disagree about the best method for developing an estimate for a given asset or liability (Bratten et al. [2013]). Because accounting standards allow managers significant discretion in how they develop estimates, auditors encounter a wide variety of estimation methods in the field, including estimation processes with varying degrees of structure (Bratten et al. [2013]).

Prior research characterizes structured tasks as well-defined tasks with fewer alternative inputs and solution paths, thus requiring less judgment. In contrast, unstructured tasks have many alternative inputs and solution paths, and require more judgment (Payne [1976], Bonner [1994], Abdolmohammadi [1999], Bratten et al. [2013]). Similarly, Hyatt and Prawitt [2001 p. 265] describe a relatively unstructured audit firm as one that offers little "guidance or other mechanisms to encourage control and uniformity…" In contrast, a more structured firm imposes "more specific guidance and control mechanisms to enhance consistency and uniformity." We extend this concept of structure to management's estimation processes. For example, management might use a more-structured estimation process that is marked by a well-defined methodology with detailed steps and well-specified inputs, which give rise to more consistent and uniform estimates (i.e., a limited solution path set). Alternatively, management could use a less-structured process that involves fewer explicit steps and more alternative inputs, resulting in more judgment and a more varied solution path set.

Research finds that accountants and auditors are averse to ambiguity and have a strong preference for well-defined and more verifiable tasks (Bamber et al. [1989], Nelson and Kinney [1997], Kadous et al. [2005], Luippold and Kida [2012], Zimbelman and Waller [1999]). Auditors appear to cope with the ambiguity and uncertainty associated with complex estimates by focusing on the quantifiable and verifiable aspects of management's estimate (Griffith et al. [2015a], Joe et al. [2017]). For example, Griffith et al. [2015a p. 858] observe that auditors are more comfortable using a "highly structured…checklist -like approach [that] easily accommodates verifying discrete components of management's estimates," even though doing so can prevent auditors from making more holistic evaluations of the estimates.[4] Overall, this

---

[4] Griffith et al. [2015a p. 856] note that, "the verification approach makes auditors more likely to overlook or justify conflicting evidence" and more likely to be influenced by management.

research suggests that auditors prefer higher levels of structure and less ambiguity. Accordingly, we expect that a relatively higher degree of estimation structure will make management's evidence around complex estimates relatively more persuasive for auditors. Therefore, as management's estimation process becomes more structured, auditors will place more weight on management's evidence, which will also cause auditors to more heavily discount contradictory evidence from firm specialists (i.e., a relative tradeoff between the two competing sources of audit evidence).

In our earlier discussion about algorithm aversion and auditors' consideration of competing information, we propose that persuasive attributes of management's evidence will have a greater influence on auditor judgments when auditors receive contradictory evidence from a specialist system instead of a human specialist. Thus, when auditors evaluate conflicting information related to management's estimates, we predict that the persuasive effect of greater estimation structure will be amplified when the source of the firm's contradictory evidence is a specialist system instead of a human specialist. Accordingly, we propose the following interaction hypothesis:

> *Hypothesis:* Relatively more structure in management's estimation process will lead to greater auditor discounting of contradictory evidence, especially when that contradictory evidence comes from a specialist system instead of a human specialist.

3. Method

3.1 PARTICIPANTS

Participants are 170 Big 4 audit seniors with a reported mean of 4.02 years of public accounting experience.[5] Participants report spending approximately 50 percent of their time

---

[5] We obtained institutional review approval for this study. We received 199 complete responses from experienced audit senior associates. We also received three additional responses from individuals who indicated they had less than one year of public accounting experience, and we omit the responses from these three participants as they do

13

working on public clients and are relatively likely to provide input into decisions related to proposed audit adjustments in a typical year (mean of 5.10 on a scale from 0 = "not at all likely" to 7 = "highly likely"). Overall, participants have experience consistent with auditors who are typically involved with auditing complex estimates (e.g., Griffith et al. [2015a]). Very few participants (4 percent) report experience working with AI systems on an audit engagement; consistent with our understanding that audit firms have not yet fully implemented these systems.[6]

3.2 EXPERIMENTAL AUDIT CASE

Participants assume the role of an in-charge auditor on the financial statement audit of Heartland National Bank. Participants first receive background information about their hypothetical audit firm (Clark & Miller, LLP) and Heartland's allowance for loan losses ("ALL"). Participants then receive information about the methodology that Clark & Miller uses to evaluate clients' ALL calculations. In all conditions, Clark & Miller's methodology involves using information from a variety of sources to develop independent loan grades, including the use of either an in-house human valuation specialist or the firm's AI-based specialist system.

Following this, participants receive information about the current audit of Heartland's ALL. Differences between Heartland's loan grades and Clark & Miller's independent loan grades indicate a potential audit difference that would overstate earnings by $28 million. Case details indicate that the audit team's investigation finds that the root cause of these differences relates to disagreements about estimated collateral values, which are a key input in the ALL calculation. Due to rapidly-increasing real estate prices, evidence from both management and the audit firm indicates that many of the appraisals used to determine collateral values are "stale"

---

not possess the requisite experience for the experimental task. We also exclude 29 responses from participants that provided incorrect answers to manipulation check questions. See Results section for additional discussion.
[6] When we exclude participants that indicate experience working with AI systems or control for their experience with AI in our models, our results and inferences remain unchanged.

(i.e., outdated) as of the balance sheet date – and therefore need to be adjusted upward. However, Heartland's methodology adjusts these collateral values more aggressively (i.e., higher values) than Clark & Miller's methodology.

Participants then view side-by-side reports from Clark & Miller and Heartland that summarize the respective methodologies that each uses to roll-forward collateral values from the appraisal date to the balance sheet date. These reports indicate that Clark & Miller examined three commercial real estate price indices, while Heartland's management examined only the two most aggressive of these three indices. Consequently, Heartland's methodology results in significantly higher collateral values (resulting in lower ALL balance and higher earnings) than Clark & Miller's methodology. These case details suggest that management may have estimated collateral values, and thus the ALL balance, in a biased manner.

Following these reports, we provide participants with a comprehensive summary of the issue. The last part of this summary informs participants that Heartland management is confident in its methodology and prefers not to make an adjustment to the ALL. Immediately following this summary, participants recommend a proposed adjustment.[7] When evaluating management's complex estimates, the auditor's consideration of the available evidence (and relative weighting of that evidence) forms the basis for whether there is a proposed adjustment to management's estimate, and the magnitude of such adjustment. For example, full reliance on firm-provided contradictory evidence would result in larger proposed adjustments. In contrast, discounting evidence that conflicts with management's estimate would result in greater agreement with management's preferred balance and, on average, result in smaller proposed adjustments.

---

[7] Senior leaders at a participating firm indicated that experienced senior associates are qualified to recommend proposed adjustments and that partners would likely consider these recommendations in their own decisions. For brevity, we refer to this variable as "proposed adjustments" throughout the remainder of the manuscript.

Therefore, the effects predicted by our hypothesis should be evident in auditors' proposed adjustments. Finally, participants complete a post-experiment questionnaire. Figure 1 presents the flow of the experimental procedures.

[Insert Figure 1 about here.]

3.3 INDEPENDENT VARIABLES

We use a 2 x 2 between-subjects experimental design that manipulates the source of an audit firm's evidence around a complex estimate (i.e., "*Source*") and the degree of structure in the client's estimation process (i.e., "*Structure*"). We manipulate *Source* at two levels: human specialist and specialist system. In the human specialist condition, Clark & Miller employs an internal group of specialists that calculates independent loan grades for individual loans. In the specialist system condition, we inform participants that Clark & Miller utilizes a proprietary AI system (i.e., a specialist system called the "Amadeus System") that develops these grades. The firm methodology for calculating loan grades and the resulting reports (i.e., audit evidence) from these two sources are identical in both conditions.

When describing these firm sources of evidence, we include statements in both treatments to equalize the acceptability of the two sources and legitimize the specialist system in the same way that audit firms legitimize human specialists in practice (see Appendix A). For example, participants in both conditions are told that their firm considers the resulting loan grades (from either source) to be an approved source of audit evidence. Overall, these design choices help familiarize the participants with the source of audit evidence and ensure that observed effects are due to the human/non-human nature of the information source rather than reluctance to deviate from more traditional firm methodologies.

16

We also manipulate *Structure* at two levels: higher structure and lower structure. In the higher structure condition, Heartland's management uses a more limited and consistent solution path to develop estimates, relying on detailed, verifiable market data to roll forward collateral value estimates. In the lower structure condition, management relies heavily on the judgement of its loan officers and credit analysts to update underlying collateral values using a variety of methods. For example, these loan officers and credit analysts can use information from comparable sales, local market trends, and/or discussions with real estate brokers, yielding a more varied solution path for estimates (see Appendix B for more detail). This design holds management's estimated collateral values and the potential for management bias in these values constant and manipulates only the process by which management generates those values.[8]

4. Results

4.1 MANIPULATION CHECKS

We evaluate our *Source* manipulation by asking participants whether the audit team received input from the firm's internal valuation group or the Amadeus system. Only 11 participants incorrectly identified the *Source* of the audit firm's evidence. With regard to our *Structure* manipulation, we varied whether management's estimation process relies on the systematic application of detailed market data (higher structure) or on the judgment and expertise of loan officers and/or credit analysts (lower structure). Accordingly, we asked participants to correctly identify management's estimation methodology and 18 additional participants were

---

[8] It is important to note that management has significant discretion to alter inputs and manipulate recorded amounts in both conditions of estimation structure. For example, swapping out a single price index in the higher structure condition influences the collateral value estimates for every loan and can materially alter the ALL calculation. Similarly, systematically biased judgments in the lower structure condition generate inappropriate collateral values at the loan level that can accumulate to a material misstatement across the portfolio.

unable to do so. Since these 29 participants did not properly attend to our manipulations, we eliminate their responses from our analyses, resulting in a final sample of 170 participants.[9]

To further gauge the effectiveness of our *Structure* manipulation, we ask participants, "To what extent would you characterize management's method for estimating updated collateral values as structured?" (1 = "Not at All Structured" and 7 = "Very Structured"). Participants in the higher structured condition reported that management's estimation process was more structured than participants in the lower structured condition (4.8 versus 3.8, respectively; $p <$ 0.01), indicating that *Structure* is successfully manipulated in our final sample.

As previously discussed, we include language in the case materials to equalize the legitimacy of the specialist-provided evidence across both levels of our *Source* manipulation – including the firm's endorsement of the respective *Source* (see Appendix B). Accordingly, we ask participants to indicate whether they agree with the statement that their audit firm views the respective *Source* as a "credible source of audit evidence" (1 = "Strongly Disagree" and 7 = "Strongly Agree"). Participants in the human specialist and specialist system conditions report similarly high levels of agreement with that statement (5.92 versus 5.61, respectively; $p > 0.10$, two-tailed, untabulated). Further, mean responses in both conditions are significantly higher than the scale midpoint of 3.5 ($p < 0.01$, two-tailed, untabulated). Thus, participants in both *Source* conditions were similarly assured that their respective source was deemed reliable and approved by the firm.

---

[9] While all of our results remain significant at the reported levels when we include responses from 10 participants who only missed the *Source* question, the interaction effect becomes insignificant when we either include responses from 18 participants who only missed the *Structure* question ($p = 0.14$) or all 199 observations ($p = 0.16$) (one participant missed both manipulation check questions). Overall, the participants who missed these questions did not sufficiently attend to the case details and the inclusion of these responses adds noise to our statistical analyses, weakening the power of our tests. We discuss how manipulation check failures affect each of our experimental conditions in the notes to Table 1. Consistent with directional expectations, *p*-values reported throughout the paper are equivalent to one-tailed tests, unless otherwise noted.

4.2 HYPOTHESIS TESTS

We test our hypothesis using a 2 x 2 ANOVA model with auditors' proposed adjustments ("*Proposed*") as the dependent variable and *Source* and *Structure* as independent variables. Table 1, Panel A, presents descriptive statistics for *Proposed* and Table 1, Panel B, reports the model results. Figure 2 presents a graphical illustration of these results.

[Insert Table 1 and Figure 2 about here.]

Our hypothesis predicts that higher (versus lower) structure in management's estimation process will cause auditors to more heavily discount contradictory evidence when the source of the firm's contradictory evidence is a system specialist versus a human specialist. This suggests that the effect of *Structure* on *Proposed* will be more negative in the specialist system condition than in the human specialist condition. Consistent with our hypothesis, there is a significant interaction ($F_{1,166} = 4.00$, p = 0.02). Additionally, the graphical pattern in Figure 2 and the simple effect tests reported in Table 1, Panel C support our hypothesis. Specifically, we find that when the source of the firm's contradictory evidence was a human specialist, there was no statistical difference in the mean proposed adjustments across the higher and lower structure conditions (19.81 versus 22.13, $F_{1,166} = 1.11$; p = 0.15). However, when the source was the firm's specialist system, auditors' average proposed adjustments were smaller in the higher structure condition than in the lower structure condition (11.20 versus 19.94; $F_{1,166} = 13.89$, p < 0.01). Overall, the observed pattern of the results suggest that algorithm aversion amplifies the persuasive influence of *Structure* in management's estimation process, consistent with our hypothesis.

Given the significant *Source* X *Structure* interaction, we also conduct simple effect tests to further explore the effects of algorithm aversion in the audit setting. Consistent with algorithm aversion, when estimation structure is higher, auditors discount the firm's contradictory evidence

19

more heavily in the specialist system condition than in the human specialist condition. As shown in Table 1, Panel D, participants in the higher structure condition propose smaller adjustments when the firm's contradictory evidence comes from a specialist system instead of a human specialist (11.20 versus 19.81; $F_{1,166} = 13.21$, $p < 0.01$). However, in the lower structure condition, proposed adjustments do not vary depending on the source of the firm's contradictory evidence (22.13 vs. 19.94; $F_{1,166} = 1.01$, $p = 0.16$). These results suggest that the effects of algorithm aversion are likely contextually dependent in the audit setting.

4.3 MODERATED MEDIATION ANALYSES

In developing our expectations, we theorize that auditors will prefer more-structured estimation processes (relative to less-structured processes), leading to relatively smaller proposed adjustments. We further theorize that this effect will be stronger when the audit firm's contradictory evidence comes from a system specialist instead of a human specialist. To provide additional support for our theory, we use conditional process analyses to examine whether the perceived quality of management's evidence ("*Quality*") mediates our predicted effects of *Structure* and *Source* on *Proposed* (i.e., moderated mediation).[10]

There are two ways *Source* might moderate the indirect effect of *Structure* on *Proposed* through *Quality*. First, consistent with Model 8 in the PROCESS Macro (Hayes [2018]), *Source* might moderate the effect of *Structure* on *Quality*. Specifically, higher estimation structure might more readily translate to favorable perceptions of management's evidence in the specialist system condition than in the human specialist condition, thereby leading to relatively smaller proposed adjustments. Alternatively, higher structure might lead to higher perceived evidence quality, regardless of *Source*. Then, due to algorithm aversion, *Source* could amplify the

---

[10] We ask participants to rate the quality of the evidence provided by management on a seven-point scale (1 = "Very Low Quality"; 7 = "Very High Quality").

influence that higher perceived evidence quality has on auditors' adjustment decisions. That is, consistent with Model 15 of the PROCESS Macro, *Source* might moderate the effect of *Quality* on *Proposed*, such that the effect of *Quality* on *Proposed* is more negative in the specialist system condition than in the human specialist condition. The following analyses help disentangle whether algorithm aversion affects the way auditors initially *perceive* management's evidence and/or how auditors ultimately *use* management's evidence to make decisions.

We use Model 8 to test the first of these two mediation possibilities (see Figure 3, Panel A). Results reveal that the indirect effect of *Structure* on *Proposed* through *Quality* (i.e., $\beta_a * \beta_b$) is significant in both the specialist system condition (90% confidence interval of -2.91 to -0.45, equivalent to $p < 0.05$) and the human specialist condition (90% confidence interval of -3.89 to -0.54). However, the index of moderated mediation is not significant (i.e., 90% confidence interval of -0.74 to 1.97), indicating that these indirect effects do not differ across our *Source* conditions. These results suggest that auditors evaluate the quality of management's evidence more favorably when estimation structure is higher, regardless of whether the firm's contradictory evidence comes from a specialist system or human specialist.

[Insert Figure 3 about here.]

We next use Model 15 to examine whether *Source* instead moderates the degree to which *Quality* affects *Proposed* (see Figure 3, Panel B). The results show that the indirect effect is significant in the specialist system condition (90% confidence interval of -5.53 to -1.43), but not in the human specialist condition (90% confidence interval of -2.37 to 0.49). Additionally, the index of moderated mediation is significant (90% confidence interval of -4.96 to -0.34). This indicates that the indirect effect of *Structure* on *Proposed,* through *Quality*, is significantly more negative in the specialist system condition than in the human specialist condition, consistent with

21

our hypothesis development. Collectively, we interpret these findings as additional support for our underlying theory. Namely, algorithm aversion affects the way auditors *use* management's evidence, rather than the way they *perceive* that evidence.

4.4 SOURCE CREDIBILITY

Consistent with algorithm aversion, we are primarily interested in how the nature (i.e., human vs. non-human) of a firm source of evidence affects auditors' reliance on that evidence. Although we explicitly inform participants in both *Source* conditions that the firm considers the specialist to be an approved source of audit evidence (see Appendix A), it is possible that participants nonetheless could have perceived differences in credibility across the two sources of evidence (specialist system vs. human specialist). We therefore test whether our results are robust to these potential differences.

Source credibility theory identifies objectivity and expertise as factors that affect the credibility of a source (Birnbaum and Stegner [1979], Pornpitakpan [2004]). Accordingly, participants rate their perceptions of the objectivity (*Objectivity*) and expertise (*Source_Expertise*) of the specialist system or human specialist.[11] Perceptions of objectivity do not differ across our *Source* conditions ($p = 0.72$, two-tailed, untabulated). However, participants in the specialist system condition report higher levels of concern about expertise than those in the human specialist condition ($p < .01$, two-tailed, untabulated). To rule out the possibility that these participants' perceptions of source credibility drive our results, we run a 2 x 2 ANCOVA that is similar to our main hypothesis-testing model but includes *Objectivity* and *Source_Expertise* as covariates. The ANCOVA results reported in Table 2 are similar to those in

---

[11] We ask participants to rate the objectivity of the firm source (i.e., the specialist system or human specialists) (1 = "Not at All Objective"; 7 = "Very Objective") and to assess the extent to which they were concerned about the knowledge and expertise of the firm source (1 = "Not at All Concerned"; 7 = "Very Concerned").

our main analyses and our inferences are unchanged. Specifically, the *Source* X *Structure* interaction ($F_{(1, 164)} = 3.22$, p = 0.04) result remains significant.

[Insert Table 2 about here.]

Similarly, when we include *Objectivity* and *Source_Expertise* as covariates in our PROCESS Macro models, the significance levels we report in our Moderated Mediation Analyses are the same (untabulated) and our inferences are unchanged. Therefore, although participants report relatively higher levels of concern about *Source_Expertise* in the specialist system condition, our results are fully robust to controlling for these differences. Thus, consistent with algorithm aversion, the human/non-human nature of the firm source appears to be the primary driver of our findings, not source credibility or related expertise concerns.[12]

5. Conclusion

To date the auditing profession has invested hundreds of millions of dollars with plans for further investments to develop and implement AI systems and the leaders of multi-national accounting firms and the audit profession assert that these investments will enhance audit quality (FEI 2017; EY 2018). One area that audit firms have targeted for AI innovation is the use of specialist systems to provide "expert" evaluations and recommendations to assist auditors in the performance of complex tasks. The profession is likely keen on implementing expert AI systems because human specialists are both a costly and scarce resource, and research has noted that audit teams are sometimes reluctant to use specialists because of the associated costs (Bratten et al.

---

[12] It is also possible that auditors are reluctant to adjust management's estimates in the specialist system condition because they believe it will be difficult to persuade management to adjust the estimate based on AI-produced audit evidence. Accordingly, we ask participants to rate the likelihood that management could be convinced to adjust an estimate primarily based on the evidence provided by the specialist system or human specialist. Participants in the human specialist condition rated this likelihood as higher than those in the specialist system condition (p = 0.01, untabulated). However, similar to our analyses related to *Source_Expertise*, when we control for this *Convincing* measure in our ANCOVA and PROCESS models, the significance of our results and our inferences are unchanged.

[2013], Griffith et al. [2015a], Griffith [2014]). While the audit profession is optimistic that the implementation of AI will enhance audit quality, research has not examined how auditors will interact with these new specialist systems or how AI might influence the way auditors evaluate evidence. Prior research in psychology and management science, however, documents that individuals are susceptible to algorithm aversion – the tendency to discount computer-generated advice more severely than human advice. Thus, if algorithm aversion occurs in auditing, there could be significant consequences for the performance of audits, and particularly for the high-risk audit areas that require specialist expertise. Motivated by these concerns, this study examines how algorithm aversion manifests in the context of auditing complex estimates.

Consistent with theory, we find that algorithm aversion impacts how auditors consider contradictory evidence provided by their firm's specialist. Specifically, we find that, when management employs a relatively more-structured estimation process, auditors more heavily discount their firm's contradictory evidence when that evidence comes from a specialist system instead of a human specialist. This finding demonstrates that algorithm aversion amplifies the persuasive effect of greater estimation structure. It is important to note that managers have discretion over their estimation methods and inputs, as well as other persuasive attributes of evidence (e.g., quantification of evidence; Joe et al. [2017]). Therefore, our findings suggest that algorithm aversion among auditors can make management's strategic attempts to influence auditor judgments through these evidence attributes more effective, thereby increasing the likelihood of auditors accepting their client's potentially biased estimates. More broadly, our findings also raise the concerning possibility that algorithm aversion might increase the overall effectiveness other management persuasion tactics (e.g., explicitly stated preferences, concessions, ingratiation) (Jenkins and Haynes [2003], Wolfe, Mauldin, and Diaz [2009],

Robertson [2010]). Thus, we highlight a possible unintended consequence of employing cognitive technologies in the audit setting.

Our study makes significant contributions to two streams of research. First, our findings have implications for research related to auditing complex estimates. PCAOB inspection findings indicate that auditors frequently fail to consider contradictory evidence identified by valuation specialists (PCAOB [2017a]). Furthermore, qualitative evidence suggests that jurisdictional motivations can lead auditors to discount (or even alter) the evidence provided by specialists if that evidence does not conform to their own perspective (Griffith [2019]). Research also finds that implemental mindsets and lower risk perceptions can increase auditors' propensity to incorrectly conclude that management's biased estimates are reasonable (Griffith, Hammersley, Kadous, and Young [2015], Griffith [2018]). We extend this literature by identifying estimation structure as a contextual feature of management's estimation process that causes auditors to have more favorable perceptions of management's estimates and increases auditors' willingness to discount contradictory evidence, particularly if that evidence is produced by an AI system.

Second, our study contributes to emerging research in psychology and management science around algorithm aversion. Prior studies that document algorithm aversion effects do so in more objective, less complex, and low-stakes task settings. Our findings demonstrate that, in the audit setting, the effects of algorithm aversion are nuanced and contextually-dependent. We document that algorithm aversion manifests in the behavior of experienced professionals with strong incentives to rely on the related computer-generated evidence – but only in the context where they judged the opposing evidence to be particularly strong. Our findings also demonstrate that in settings with competing information, algorithm aversion serves to amplify the persuasive effect of the evidence attributes under consideration.

Our study is subject to some limitations that provide several interesting opportunities for future research. First, we examine the effects of algorithm aversion only in the setting of auditing complex estimates, which can be a highly challenging and highly subjective task for auditors. We use this specific setting because audit firms are currently directing their AI investments toward similar efforts (e.g., KPMG [2016]). Future research might examine auditor reliance on AI in more objective audit tasks. Second, this study measures auditors' reactions to a new and novel source of audit evidence. It is possible that the effects we detect will change over repeated exposures that reduce this novelty. Specifically, auditors might become more willing to rely on these systems as they become more familiar with them. That said, previous research finds that individuals are more willing to rely on algorithms when they have *no* experience using the algorithm because they have not seen the algorithm err (Dietvorst et al. [2015]). With repeated exposure and usage, auditors are also likely to encounter both positive and negative experiences as they use AI systems. Future research could examine how the valence of these experiences shapes the way auditors rely on these systems. Finally, future research could explore theory-grounded interventions that mitigate the effects algorithm aversion and, ultimately, help the auditing profession recognize the full benefits of its investments in cognitive technologies. Our study takes the important first step of identifying when and how receiving audit evidence from an AI system alters auditor decisions around complex accounting estimates.

REFERENCES

ABDOLMOHAMMADI, M. J. "A Comprehensive Taxonomy of Audit Task Structure, Professional Rank and Decision Aids for Behavioral Research." *Behavioral Research in Accounting* 11 (1999): 51-92.

ALLIOTT GROUP. "Alliott Group Worldwide Conference Highlights How Mid-Market Firms Can Win in Rapidly Changing Professional Services Market." 2017. Available at https://www.alliottgroup.net/international-alliance-law-accounting-firms-news/2017-worldwide-conference-report/.

ANDERSON, J. C.; K. K. MORENO; AND J. M. MUELLER. "The Effect of Client vs. Decision Aid as a Source of Explanations Upon Auditors' Sufficiency Judgments: A Research Note." *Behavioral Research in Accounting* 15 (2003): 1-11.

ASARE, S. K., AND A. M. WRIGHT. "The Effectiveness of Alternative Risk Assessment and Program Planning Tools in a Fraud Setting." *Contemporary Accounting Research* 21 (2004): 325-352.

AUSTIN, A. A.; T. CARPENTER; M. H. CHRIST; AND C. NIELSON. "The Data Analytics Transformation: Evidence from Auditors, CFOs, and Standard-Setters." Working paper, University of Richmond and University of Georgia, 2018. Available at https://pdfs.semanticscholar.org/e308/2c715f168c2c2569ebe93ad449117858234e.pdf.

BAMBER, E. M.; D. SNOWBALL; AND R. M. TUBBS. "Audit Structure and Its Relation to Role Conflict and Role Ambiguity: An Empirical Investigation." *The Accounting Review* 64 (1989): 285-299.

BONNER, S. E. "A Model of the Effects of Audit Task Complexity." *Accounting, Organizations and Society* 19 (1994): 213-234.

BIRNBAUM, M. H. Intuitive Numerical Prediction. *The American Journal of Psychology* 89 (1976): 417-429.

BIRNBAUM, M. H., AND B. A. MELLERS. "Bayseian Inference: Combining Base Rates with Opinions of Sources Who Vary in Credibility." *Journal Personality and Social Psychology* 45 (1983): 792-804.

BIRNBAUM, M. H., AND S. E. STEGNER. "Source Credibility in Social Judgment: Bias, Expertise, and the Judge's Point of View." *Journal of Personality and Social Psychology* 37 (1979): 48-74.

BIRNBAUM, M. H.; R. WONG; AND L. K. WONG. "Combining Information from Sources That Vary in Credibility." *Memory & Cognition* 4 (1976): 330-336.

BRATTEN, B.; L. M. GAYNOR; L. MCDANIEL; N. R. MONTAGUE; AND G. E. SIERRA. "The Audit of Fair Values and Other Estimates: The Effects of Underlying Environmental, Task, and Auditor-Specific Factors." *Auditing: A Journal of Practice & Theory* 32 (2013): 7-44.

CANADIAN PUBLIC ACCOUNTABILITY BOARD (CPAB). "CPAB Report on 2015 Inspections of Canada's Big Four Accounting Firms." 2015a. Available at http://www.cpab-ccrc.ca/Documents/Annual%20Reports/CPAB_2015_Big_Four_Report_Highlights_EN_Final.pdf.

CANADIAN PUBLIC ACCOUNTABILITY BOARD (CPAB). "CPAB Report on 2014 All Firm Inspections of the Quality of Audits Across Canada's Big Four Accounting Firms: Highlights." 2015b. Available at http://www.cpab-ccrc.ca/Documents/Topics/Public%20Reports/CPAB_2014_PR_Highlights_EN_FNL.pdf

CANNON, N. H., AND J. C. BEDARD. "Auditing Challenging Fair Value measurements: Evidence from the Field." *The Accounting Review* 92 (2017): 81-114.

CASTER, P., AND K. V. PINCUS. "An Empirical Test of Bentham's Theory of the Persuasiveness of Evidence." *Auditing: A Journal of Practice & Theory* 15 (1996): 1-22.

CHRIST, M. H.; S. A. EMETT; S. L. SUMMERS; AND D. A. WOOD. "Prepare for Takeoff: Improving Audit Efficiency and Effectiveness with Drone-enabled Inventory Audit Procedures. Working paper, University of Georgia, Arizona State University, and Brigham Young University, 2019. Available at https://business.okstate.edu/site-files/docs/cepd/christ-emett-summers-and-wood-for-osu.pdf.

CHRISTENSEN, B. E.; S. M. GLOVER; AND D. A. WOOD. "Extreme Estimation Uncertainty in Fair Value Estimates: Implications for Audit Assurance. *Auditing: A Journal of Practice & Theory* 31 (2012): 127-146.

COOPER, L. A.; D. K. HOLDERNESS, JR.; T. L. SORENSEN; D. A. WOOD. "Robotic Process Automation in Public Accounting." *Accounting Horizons* (2019): Forthcoming.

DELOITTE LLC. "Solve for Now. Build for Next. The Deloitte Audit." 2018. Available at https://www2.deloitte.com/content/dam/Deloitte/global/Documents/Audit/gx-audit-solve-for-now-build-for-next.pdf.

DENNIS, S. A.; J. B. GRIFFIN; AND K. M. JOHNSTONE. "The Value Relevance of Managers' and Auditors' Disclosures About Material Measurement Uncertainty." *The Accounting Review* (2019): Forthcoming.

DIETVORST, B. J.; J. P. SIMMONS; AND C. MASSEY. "Algorithm Aversion: People Erroneously Avoid Algorithms After Seeing Them Err." *Journal of Experimental Psychology: General* 144 (2015): 114-126.

DZINDOLET, M. T.; L. G. PIERCE; H. P. BECK; AND L. A. DAWE. "The Perceived Utility of Human and Automated Aids in a Visual Detection Task." *Human Factors* 44 (2002): 79-94.

EASTWOOD, J.; B. SNOOK; AND K. LUTHER. "What People Want from Their Professionals: Attitudes Toward Decision-making Strategies." *Journal of Behavioral Decision Making* 25 (2012): 458-468.

ERNST & YOUNG (EY). "How Artificial Intelligence Will Transform the Audit." 2018. Available at https://www.ey.com/en_us/assurance/how-artificial-intelligence-will-transform-the-audit.

FINANCIAL EXECUTIVE INTERNATIONAL (FEI). "What Impact Will AI Have on the Audit?" 2017. Available at https://daily.financialexecutives.org/impact-will-ai-audit/.

GINO, F., AND D. A. MOORE. "Effects of Task Difficulty on Use of Advice." *Journal of Behavioral Decision Making* 20 (2007): 21-35.

GLOVER, S. M.; D. F. PRAWITT; AND B. C. SPILKER. "The Influence of Decision Aids on User Behavior: Implications for Knowledge Acquisition and Inappropriate Reliance." *Organizational Behavior and Human Decision Processes,* 72 (1997): 232-255.

GLOVER, S. M.; M. H. TAYLOR; AND Y. WU. "Mind the Gap: Why Do Experts Have Differences of Opinion Regarding the Sufficiency of Audit Evidence Supporting Complex Fair Value Measurements?" *Contemporary Accounting Research* (2019): Forthcoming.

GOODWIN, J. "The Effects of Source Integrity and Consistency of Evidence on Auditors' Judgments." *Auditing: A Journal of Practice & Theory* 18 (1999): 1-16.

GRIFFITH, E. E. "The Role of Valuation Specialists in Auditing Complex Estimates." 2014. PhD diss., University of Georgia.

GRIFFITH, E. E.; J. S. HAMMERSLEY; AND K. KADOUS. "Audits of Complex Estimates as Verification of Management Numbers: How Institutional Pressures Shape Practice." *Contemporary Accounting Research* 32 (2015a): 833-863.

GRIFFITH, E. E; J. S. HAMMERSLEY; D. YOUNG; AND K. KADOUS. "Auditor Mindsets and Audits of Complex Estimates." *Journal of Accounting Research* 53 (2015b): 49-77.

GRIFFITH, E. E. "When Do Auditors Use Specialists' Work to Improve Problem Representations of and Judgments About Complex Estimates?" *The Accounting Review* 93 (2018): 177-202.

GRIFFITH, E. E. "Auditors, Specialists, and Professional Jurisdiction in Audits of Fair Values." *Contemporary Accounting Research* (2019): Forthcoming.

HARRIS, S. B. "Technology and the Audit of Today and Tomorrow." 2017. Speech delivered at PCAOB/AAA Annual Meeting: Washington DC, April 20, 2017. Available at https://pcaobus.org/News/Speech/Pages/Harris-statement-PCAOB-AAA-4-20-17.aspx.

HANSON, J. D. "Auditing the Future." 2012. Speech delivered at PCAOB/Fair Value Measurements and Reporting Conference National Harbor, MD, June 7, 2012. Available at https://pcaobus.org/News/Speech/Pages/06072012_HansonAICPA.aspx.

HAYES, A. F. Introduction to Mediation, Moderation, and Conditional Process Analysis: A Regression-based Approach. Guilford Publications, 2018.

HYATT, T. A., AND D. F. PRAWITT. "Does Congruence Between Audit Structure and Auditors' Locus of Control Affect Job Performance?" *The Accounting Review* 76 (2001): 263-274.

INTERNATIONAL FORUM OF INDEPENDENT AUDIT REGULATORS (IFIAR). "Report on 2014 Survey of Inspection Findings." 2015. Available at http://www.h3c.org/textes/Etude_IFIAR_2014-EN.pdf.

JENKINS, J. G., AND C. M. HAYNES. "The Persuasiveness of Client Preferences: An Investigation of the Impact of Preference Timing and Client Credibility." *Auditing: A Journal of Practice & Theory* 22 (2003): 143-154.

JOE, J. R.; S. D. VANDERVELDE; AND Y.-J. WU. "Use of High Quantification Evidence in Fair Value Audits: Do Auditors Stay in Their Comfort Zone?" *The Accounting Review* 92 (2017): 89-116.

KACHELMEIER, S. J., AND W. F. MESSIER JR. "An Investigation of the Influence of a Nonstatistical Decision Aid on Auditor Sample Size Decisions." The *Accounting Review* 65 (1990): 209-226.

KADOUS, K. L.; L. KOONCE; AND K. L. TOWRY. "Quantification and Persuasion in Managerial Judgment." *Contemporary Accounting Research* 22 (2005): 643-686.

KAPLAN, S. E,; E. F. O'DONNELL; AND B. M. AREL. "The Influence of Auditor Experience on the Persuasiveness of Information Provided by Management." *Auditing: A Journal of Practice & Theory* 27 (2008): 67-83.

KPMG, LLP. "Harnessing the Power of Cognitive Technology to Transform the Audit." 2016. Available at https://assets.kpmg/content/dam/kpmg/us/pdf/2016/11/us-audit-CognitiveFactSheet.pdf.

LEVENE, H. "Robust Tests for Equality of Variances." *Contributions to Probability and Statistics*: *Stanford Studies in Mathematics and Statistics* (1960): 278-92.

LUIPPOLD, B. L., AND T. E. KIDA. "The Impact of Initial Information Ambiguity on the Accuracy of Analytical Review Judgments." *Auditing: A Journal of Practice & Theory* 31 (2012): 113-129.

MADHAVAN, P., AND D. A. WIEGMANN. "Effects of Information Source, Pedigree, and Reliability on Operator Interaction with Decision Support Systems." *Human Factors* 49 (2007): 773-785.

MARTIN, R. D.; J. S. RICH; AND T. J. WILKS. "Auditing Fair Value Measurements: A Synthesis of Relevant Research." *Accounting Horizons* 20 (2006): 287-303.

MESSIER, W. F. "Research in and Development of Audit Decision Aids." *Judgment and Decision Making Research in Accounting and Auditing*, edited by R. Ashton and A. Ashton. Cambridge: Cambridge University Press, 1995.

MURPHY, H. "Auditing to be Less of a Burden as Accountants Embrace AI." 2017. *Financial Times.* Available at https://www.ft.com/content/0898ce46-8d6a-11e7-a352-e46f43c5825d.

NELSON, M. W., AND W. R. KINNEY JR. "The Effect of Ambiguity on Loss Contingency Reporting Judgments." *Accounting Review* (1997): 257-274.

ÖNKAL, D.; P. GOODWIN; M. THOMSON; S. GÖNÜL; AND A. POLLOCK. "The Relative Influence of Advice from Human Experts and Statistical Methods on Forecast Adjustments." *Journal of Behavioral Decision Making* 22 (2009): 390-409.

PAYNE, J. W. "Task Complexity and Contingent Processing in Decision Making: An Information Search and Protocol Analysis." *Organizational Behavior and Human Performance* 16 (1976): 366-387.

PORNPITAKPAN, C. "The Persuasiveness of Source Credibility: A Critical Review of Five Decades' Evidence." *Journal of Applied Social Psychology* 34 (2004): 243-281.

PRAHL, A., AND L. V. SWOL. "Understanding Algorithm Aversion: When is Advice From Automation Discounted?" *Journal of Forecasting* 36 (2017): 691-702.

PRICEWATERHOUSECOOPERS (PWC). "Confidence in the Future. Human and Machine Collaboration in the Audit." 2017. Available at https://www.pwc.co.uk/services/audit-assurance/statutory-audit/insights/confidence-in-the-future-audit-innovation/human-and-machine-collaboration-in-the-audit.html.

PROMBERGER, M., AND J. BARON. "Do Patients Trust Computers?" *Journal of Behavioral Decision Making* 19 (2006): 455-468.

PUBLIC COMPANY ACCOUNTING OVERSIGHT BOARD (PCAOB). *Staff Audit Practice Alert No. 2: Matters Related to Auditing Fair Value Measurements of Financial Instruments and The Use of Specialists.* WASHINGTON, DC: PCAOB, 2007.

PUBLIC COMPANY ACCOUNTING OVERSIGHT BOARD (PCAOB). *The Auditor's Responses to the Risks of Material Misstatement.* Auditing Standard 2301. Washington, DC: PCAOB, 2010a.

PUBLIC COMPANY ACCOUNTING OVERSIGHT BOARD (PCAOB). *Report on Observations of PCAOB Inspectors Related to the Audit Risk Areas Affected by the Economic Crisis.* Release No. 2010-006. Washington, DC: PCAOB, 2010b.

PUBLIC COMPANY ACCOUNTING OVERSIGHT BOARD (PCAOB). *PCAOB Release No. 104-2014.166. KPMG LLP's Quality Control Remediation Submissions.* Washington, DC: PCAOB, 2014a.

PUBLIC COMPANY ACCOUNTING OVERSIGHT BOARD (PCAOB). *Staff Consultant Paper: Auditing Accounting Estimates and Fair Value Measurements.* Washington, DC: PCAOB, 2014b.

PUBLIC COMPANY ACCOUNTING OVERSIGHT BOARD (PCAOB). *Staff Consultation Paper No. 2015-01, the Auditors' Use of the Work of Specialists.* Washington, DC: PCAOB, 2015.

PUBLIC COMPANY ACCOUNTING OVERSIGHT BOARD (PCAOB). *Proposed Amendments to Auditing Standards for Auditor's Use of the Work of Specialists*. PCAOB Release No. 2017-003, 2017a.

PUBLIC COMPANY ACCOUNTING OVERSIGHT BOARD (PCAOB). *Proposed Auditing Standard – Auditing Accounting Estimates, Including Fair Value Measurements And Proposed Amendments To PCAOB Auditing Standards.* PCAOB Release No. 2017-002, 2017b.

PUBLIC COMPANY ACCOUNTING OVERSIGHT BOARD (PCAOB). *Auditing Accounting Estimates, Including Fair Value Measurement.* PCAOB Release No. 2018-005, 2018.

RAPHAEL, J. "Rethinking the Audit." 2017. *Journal of Accountancy*. Available at https://www.journalofaccountancy.com/issues/2017/apr/rethinking-the-audit.html.

ROBERTSON, J. C. "The Effects of Ingratiation and Client Incentive on Auditor Judgment." *Behavioral Research in Accounting* 22 (2010): 69-86.

SCHRAH, G. E.; R. S. DALAL; AND J. A. SNIEZEK. "No Decision-maker is an Island: Integrating Expert Advice with Information Acquisition." *Journal of Behavioral Decision Making* 19 (2006): 43-60.

TYSIAC, K. "How to Enable Audit Innovation." 2017. *Journal of Accountancy*. Available at https://www.journalofaccountancy.com/issues/2017/apr/audit-innovation.html

WEST, D. M., AND J. R. ALLEN. "How Artificial Intelligence is Transforming the World." 2018. *Brookings Institution.* Available at https://www.brookings.edu/research/how-artificial-intelligence-is-transforming-the-world/.

WOLFE, C. J; E. G. MAULDIN; AND M. C. DIAZ. "Concede or Deny: Do Management Persuasion Tactics Affect Auditor Evaluation of Internal Control Deviations?" *The Accounting Review* 84 (2009): 2013-2037.

YEOMANS, M.; A. K. SHAH; S. MULLAINATHAN; AND J. KLEINBERG. "Making Sense of Recommendations." Working paper. Harvard University, 2017. Available at https://scholar.harvard.edu/files/sendhil/files/recommenders55.pdf.

ZIMBELMAN, M. F., AND W. S. WALLER. "An Experimental Investigation of Auditor-Auditee Interaction Under Ambiguity." *Journal of Accounting Research* 37 (1999): 135–155.

	Human specialist condition	Specialist system condition
Source of audit evidence:	Your firm's department of specialized professionals	Your firm's proprietary AI system
Qualifications:	"These internal valuation specialists have advanced degrees and professional certifications. They also have significant experience with audits of commercial loans (on large numbers of clients), and they continue to receive extensive and rigorous training in their areas of expertise." "The firm has invested significant resources developing and supporting the valuation group"	"To develop the Amadeus system, your firm partnered with a large international technology company with leading experts in artificial intelligence. Additionally, the firm gathered input from valuation specialists with expertise in commercial loan grading (e.g., advanced degrees, professional certifications, significant experience, and extensive and rigorous training)." "The firm has invested significant resources developing and supporting the Amadeus system"
Method:	"applies firm-approved methodologies to evaluate information from a variety of sources…uses information from clients as well as external information to develop independent loan grades"	
Firm endorsement:	"Your firm has indicated that the [Amadeus system's/internal valuation group's] overall predictions are reasonably accurate and are considered an approved source of audit evidence"	

Notes: The purpose of Appendix A is to demonstrate how the language in our instrument equalizes legitimacy and credibility across the human specialist and specialist system conditions.

APPENDIX B
Structure Manipulation

Lower Degree of Structure

Our commercial loan portfolio is unique and highly diversified. Therefore, we relied heavily on the expertise of our loan officers and credit analysts when updating collateral values for our 12/31/2017 allowance calculation. Specifically, our loan officers and credit analysts determine whether the most recent appraisal still reliably reflects a property's current market value.

When these parties determine that an appraised value is "stale", they must either order a new appraisal, obtain a Broker Price Opinion (BPO), or document an alternative valuation based on research. BPOs are significantly less comprehensive than appraisals; however, BPOs can be performed more quickly because they account for less data. When neither a current appraisal nor a current BPO is available, loan officers and credit analysts collaborate to update the most recent appraised value based on their own research (e.g., comparable sales, local market trends, discussions with brokers). For the majority of the loans in our portfolio, our loan officers and credit analysts relied on either BPOs or independent research to update appraised values.

Higher Degree of Structure

Our commercial loan portfolio is unique and highly diversified. Therefore, we relied heavily on detailed market data when updating collateral values for our 12/31/2017 allowance calculation. Specifically, we obtained detailed monthly data from the RCA and NCREIF price indices at the industry level for each of the metropolitan markets in which the collateral underlying our commercial loan portfolio is located. Depending on the property type/metropolitan market combination, annualized increases in collateral values varied from 0.2% to 19.1%.

We then used the detailed monthly information from these indices to update collateral values from the appraisal date to 12/31/2017. For example, office properties in Indianapolis increased by 0.8%, 0.7%, and 0.8% in October, November, and December, respectively. We would therefore apply those respective changes to update a 9/30/2017 appraisal for an office building in Indianapolis over the period from 10/1/2017 through 12/31/2017.

Notes: Appendix B presents how structure in management's estimation process is manipulated across the higher and lower structure conditions.

FIGURE 1
Experimental Design Flow

| Part I: Instructions/Background	Part II Summary of Methodology	PART IV Specialist/Client Memos	PART V Summary	PART VI Variables/PEQ
Participants receive background information about the client, engagement status, and allowance calculation	Participants are informed of the audit firm's approach and methodology of auditing the allowance*, the client's methodology of estimating the allowance**, and the audit findings around the allowance (and related potential misstatement)	Participants receive side-by-side memos from the audit firm and the client summarizing respective methodologies. Audit firm's specialist memo references three price indices that suggest smaller increase in collateral values, while the client memo references two price indices that suggest larger increase in collateral values. Client memo also contained description of management's estimation process.**	Participants receive summary of potential audit difference, summary of audit firm's assessment*, and summary of management's position and estimation process.**	Participants proposed an audit adjustment, completed post-experimental questionnaire, manipulation check questions, and demographic questions.

Notes: Figure 1 presents the flow of the experimental design.

* *Source* is manipulated as human specialist or specialist system.
** *Structure* is manipulated as higher estimation structure or lower estimation structure within the client's estimation process.

FIGURE 2
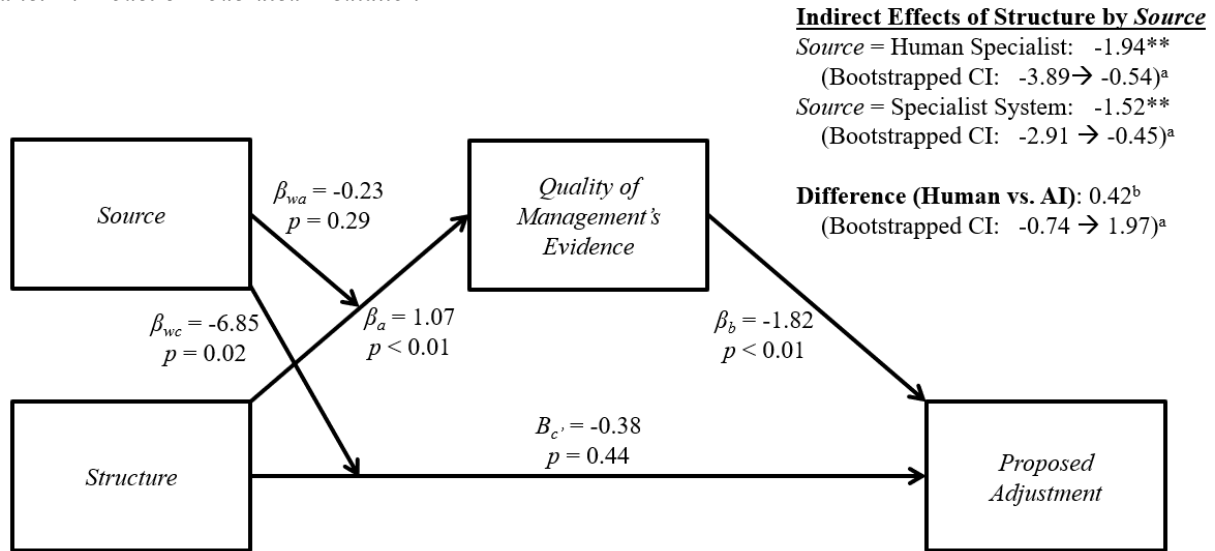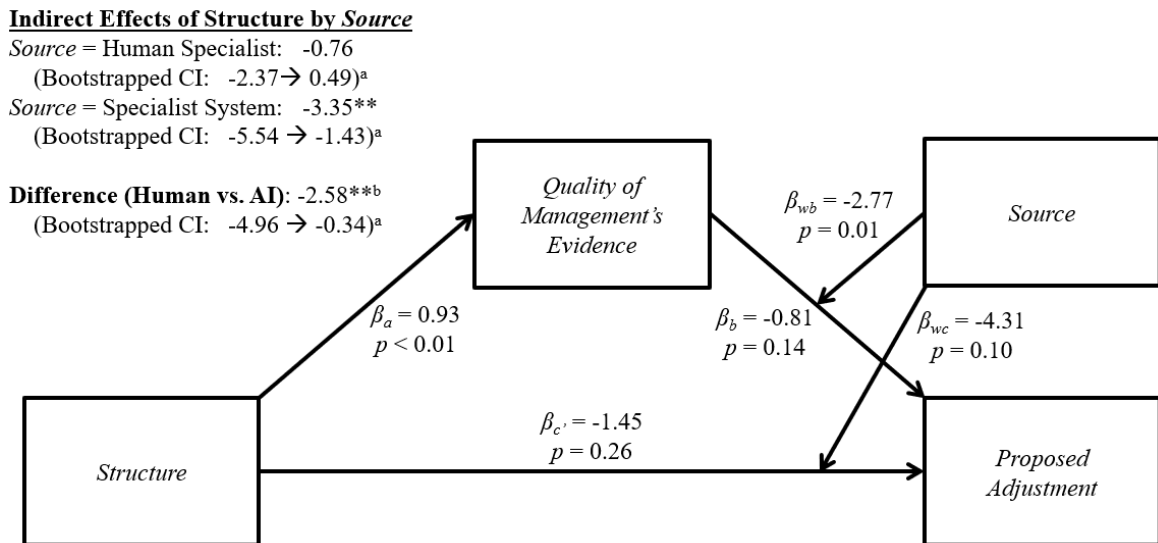**Graphical Representation of Results**
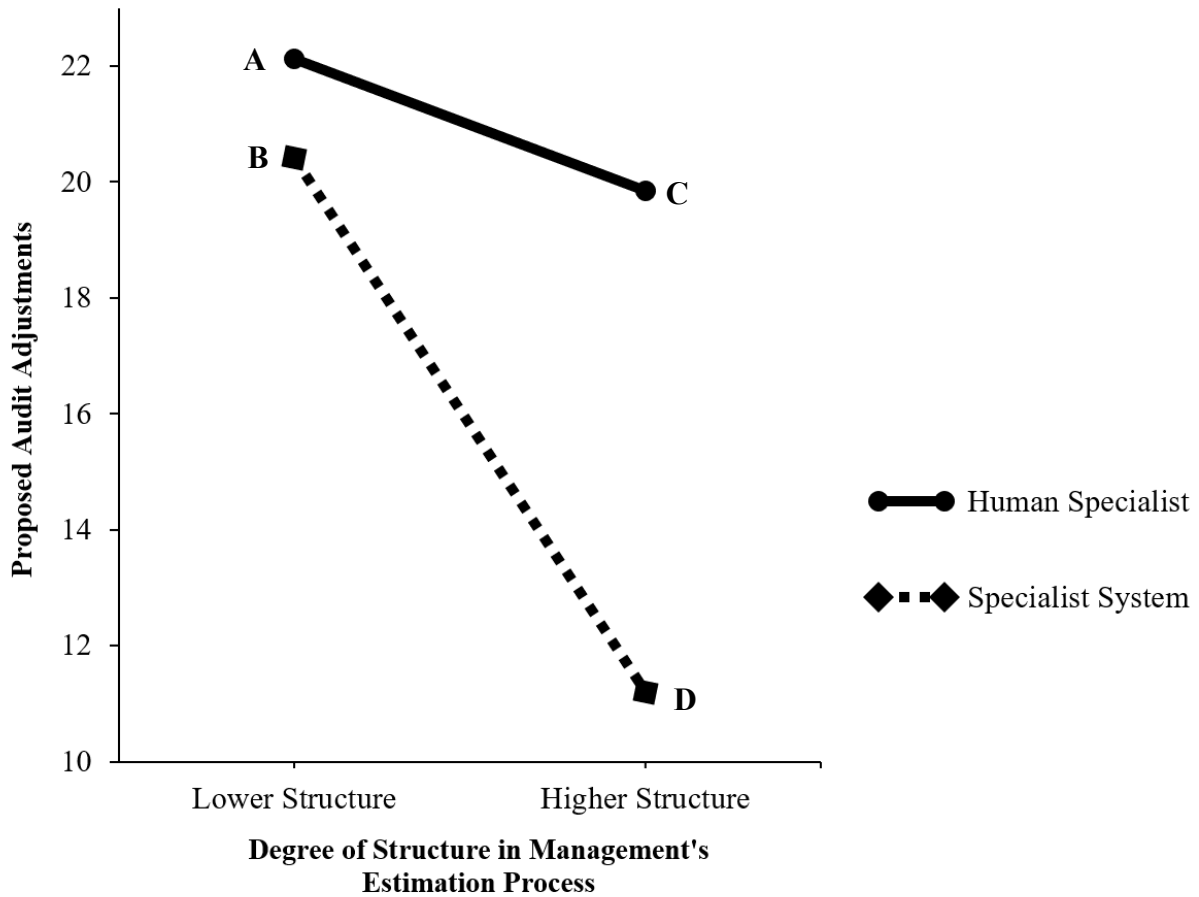
Notes: The dependent variable is participants' proposed audit adjustment (in millions). We manipulate the source of the audit firm's evidence at two levels (human specialist versus specialist system), between participants. We also manipulate the degree of structure in management's estimation process at two levels (lower versus higher), between participants.

FIGURE 3
Moderated Mediation Analysis

Panel A: Model 8 Moderated Mediation

<u>**Indirect Effects of Structure by *Source***</u>
Source = Human Specialist: -1.94**
  (Bootstrapped CI: -3.89→ -0.54)[a]
Source = Specialist System: -1.52**
  (Bootstrapped CI: -2.91 → -0.45)[a]

Difference (Human vs. AI): 0.42[b]
  (Bootstrapped CI: -0.74 → 1.97)[a]

Source

$\beta_{wa} = -0.23$
$p = 0.29$

Quality of Management's Evidence

$\beta_{wc} = -6.85$
$p = 0.02$

$\beta_a = 1.07$
$p < 0.01$

$\beta_b = -1.82$
$p < 0.01$

Structure

$B_{c'} = -0.38$
$p = 0.44$

Proposed Adjustment

Panel B: Model 15 Moderated Mediation

<u>**Indirect Effects of Structure by *Source***</u>
Source = Human Specialist: -0.76
  (Bootstrapped CI: -2.37→ 0.49)[a]
Source = Specialist System: -3.35**
  (Bootstrapped CI: -5.54 → -1.43)[a]

Difference (Human vs. AI): -2.58**[b]
  (Bootstrapped CI: -4.96 → -0.34)[a]

Quality of Management's Evidence

Source

$\beta_{wb} = -2.77$
$p = 0.01$

$\beta_a = 0.93$
$p < 0.01$

$\beta_b = -0.81$
$p = 0.14$

$\beta_{wc} = -4.31$
$p = 0.10$

Structure

$\beta_{c'} = -1.45$
$p = 0.26$

Proposed Adjustment

Notes: ** denote statistical significance equivalent to $p < 0.05$, one-tailed, respectively

[a] We use confidence intervals from bootstrapped sampling distributions (based on 10,000 bootstrap samples) to test the significance of indirect effects (Hayes [2018]). Since we have directional predictions for all indirect effects, we use 90% confidence intervals (i.e., bounded at 0.05 and 0.95) to test whether one-tailed p-values are less than 0.05.

[b] We test the significance of this difference (calculated as $\beta_{wa} * \beta_b$ in Panel A and $\beta_a * \beta_{wb}$ in Panel B) to determine whether the effect of *Structure* is mediated by *Quality of Management's Evidence*, contingent on *Source*.

TABLE 1
TABLE 1
Hypothesis-Testing Model

Panel A: Proposed Adjustments by Condition: Least squares mean (delta-method standard error) [n] Cell

	Lower Structure	Higher Structure	Overall
Human Specialist	22.13	19.81	20.97
	(1.52)	(1.58)	(1.10)
	[47]	[43]	[90]
	A	C	
Specialist System	19.94	11.20	16.13
	(1.55)	(1.76)	(1.17)
	[45]	[35]	[80]
	B	D	
Overall	21.04	15.51	
	(1.09)	(1.19)	
	[92]	[78]	

Panel B: ANOVA Results

Source	Sum of Squares	df	F	p
Structure	1,283.77	1	11.84	< 0.01
Source	1,223.91	1	11.29	< 0.01
Source × Structure	433.25	1	4.00	0.02
Error	17,997.05	166		

Panel C: Simple Effects of Structure

	df	F	p
Across Human Specialist: Cell A > Cell C	1,166	1.11	0.15
Across Specialist System: Cell B > Cell D	1,166	13.89	< 0.01

Panel D: Simple Effects of Source

	df	F	p
Across Lower Structure: Cell A > Cell B	1,166	1.01	0.16
Across Higher Structure: Cell C > Cell D	1,166	13.21	< 0.01

Notes: The dependent variable is participants' proposed audit adjustments. We manipulate the source of the audit firm's evidence at two levels (Human Specialist versus Specialist System), between participants. We also manipulate the degree of structure in management's estimation process at two levels (lower versus higher), between participants. Consistent with our directional hypothesis, all reported p-values are equivalent to a one-tailed test. Our unbalanced cell sizes reflect differences in both manipulation check failure rates and otherwise unusable responses. All three of the participants who indicated they had inadequate public accounting experience (i.e., less than one year) were randomly assigned to Cell D and were excluded from our final sample. Additionally, we excluded three, nine, five, and twelve participants from Cells A, B, C, and D, respectively, due to manipulation check failures. Although the cell sizes are unbalanced, Levene's [1960] test for equality of variances is not significant ($F_{1,166} = 1.93$; $p = 0.13$, two-tailed), indicating that the assumption of homogeneity of variances has not been violated in our ANOVA model.

TABLE 2
Analysis of Covariance Results

Source	Sum of Squares	df	F	p
Structure	1,168.27	1	11.31	< 0.01
Source	425.96	1	4.06	0.02
Source × Structure	338.14	1	3.22	0.04
Objectivity (covariate) (two-tailed p-value)	182.77	1	1.74	0.19
Source_Expertise (covariate) (two-tailed p-value)	458.97	1	4.37	0.04
Error	17,213.47	164		

Notes: The dependent variable is participants' proposed audit adjustments. We manipulate the source of the audit firm's evidence at two levels (Human Specialist versus Specialist System), between participants. We also manipulate the degree of structure in management's estimation process at two levels (lower versus higher), between participants. All tests include the effects of the covariates *Objectivity* and *Source_Expertise*. Consistent with our directional hypothesis, all reported p-values are equivalent to a one-tailed test, unless otherwise noted.