CHICAGO BOOTH

Man Versus Machine: Complex Estimates and Auditor Reliance on Artificial Intelligence

BENJAMIN P. COMMERFORD ⓘD,* SEAN A. DENNIS ⓘD,†
JENNIFER R. JOE ⓘD,‡ AND JENNY W. ULLA ⓘD§

ABSTRACT

Audit firms are investing billions of dollars to develop artificial intelligence (AI) systems that will help auditors execute challenging tasks (e.g., evaluating complex estimates). Although firms assume AI will enhance audit quality, a growing body of research documents that individuals often exhibit "algorithm aversion"—the tendency to discount computer-based advice more heavily than human advice, although the advice is identical otherwise. Therefore, we conduct an experiment to examine how algorithm aversion manifests in auditor judgments. Consistent with theory, we find that auditors receiving contradictory evidence from their firm's AI system (instead of a human specialist)

*University of Kentucky; †University of Central Florida; ‡University of Delaware; §University of Nevada, Las Vegas

propose smaller adjustments to management's complex estimates, particularly when management develops their estimates using relatively objective (vs. subjective) inputs. Our findings suggest auditor susceptibility to algorithm aversion could prove costly for the profession and financial statements users.

1. Introduction

Some of the largest audit firms have touted plans to invest billions of dollars in audit technologies in the coming years, with the goal of enhancing the effectiveness, efficiency, and decision-usefulness of audits (Bloomberg Tax [2020]). One of the most promising advanced technologies under consideration is the application of machine learning or artificial intelligence (AI) on audit engagements. 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]). In doing so, audit firms are carefully developing "narrow AI" systems to perform specific tasks within certain programmable parameters.[1] Currently, the goal for these AI systems is to use machine learning to replicate specific (yet complex) judgments that auditors and human specialists make (e.g., Deloitte [2014, 2016], PwC [2017]). Firms believe that applying these advanced technologies 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–rather than a firm's human specialist–influences auditor judgments related to complex estimates.

Auditors lack the requisite expertise to evaluate management's complex estimates on their own, and are therefore expected to rely on advice from

---

[1] According to the Brookings Institution (e.g., West and Allen [2018]), AI incorporates intentionality, intelligence (i.e., "machine learning"), and adaptability. "General AI" refers to a fully autonomous system that can learn about any problem and then solve it (e.g., Deloitte [2018a]). With narrow AI systems (e.g., KPMG [2016]), firms can maintain more direct control over their audit methodologies and how they comply with current auditing standards. This focus on narrow AI also is consistent with documented concerns around potential regulatory scrutiny and legal liability that can arise from the use of autonomous, unsupervised AI (e.g., Christ, Emett, Summers, and Wood [2021], Emett, Kaplan, Mauldin, and Pickerd [2021]). Importantly, while firms appear focused on narrow AI currently, they intend to use this technology to assist with some of the most challenging tasks that humans perform during audits (e.g., KPMG [2016]).

experts when doing so (Martin, Rich, and Wilks [2006], Griffith, Hammersley, and Kadous [2015], PCAOB [2017a, b]). Due to the subjectivity and related potential for management bias inherent in complex estimates, auditors have strong incentives to rely on evidence from their own firm's expert specialists. However, as audit firms begin using AI systems rather than human specialists to produce evidence, auditors will transition from weighing the advice of humans to weighing the advice of machines. Although some regulators have expressed concerns that auditors might over-rely on advanced audit technologies (e.g., Harris [2017]), theory suggests the risk of auditor *under*-reliance is also a concern.

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., Eastwood, Snook, and Luther [2012], Dietvorst, Simmons, and Massey [2015]). For example, identical stock forecasting advice has a greater influence on individuals' own predictions when they believe that advice comes from a human instead of a computer-based model (Önkal et al. [2009]). This aversion persists even when individuals receive feedback suggesting the algorithmic predictions are more accurate than their own (Dietvorst, Simmons, and Massey [2015]). Algorithm aversion is particularly evident in tasks that are highly subjective (e.g., movie and joke recommendations), possibly due to beliefs that algorithms lack the necessary capabilities to perform subjective tasks effectively (Castelo, Bos, and Lehmann [2019], Yeomans et al. [2019]). Furthermore, as task uncertainty increases, individuals exhibit greater algorithm aversion (Dietvorst and Bharti [2020]). Therefore, because evaluating management's complex estimates is inherently subjective and involves a high degree of uncertainty, auditors' judgments will likely exhibit some degree of algorithm aversion in this context. Specifically, this literature suggests auditors will more heavily discount firm-provided evidence around a complex estimate when it comes from an AI system than when that same evidence comes from a human specialist.

Algorithm aversion could be especially problematic in competitive information environments, such as the audit setting, because discounting one source's information causes a corresponding increase in the weight placed on *competing* information (Birnbaum [1976], Birnbaum and Stegner [1979]). For example, Birnbaum and Stegner [1979] develop and test a scale-adjustment model that shows lower perceived expertise in one source creates a trade-off, whereby advice from that source is discounted and competing advice is more heavily weighted. Therefore, if auditors discount firm-provided evidence (vis-à-vis algorithm aversion), this will likely cause them to more heavily weight management-provided evidence, making them more willing to accept management's preferences. For instance, auditors consider evidence from both management and firm sources when deciding whether and to what extent management's estimates should be adjusted. Drawing on the algorithm aversion and advice-taking literatures,

we predict that receiving contradictory evidence from an AI system instead of a human specialist will cause auditors to propose lower adjustments to management's complex estimates.

Because auditors operate in a competitive information environment, we draw further on the advice-taking literature to investigate whether the nature of management's evidence moderates the degree to which algorithm aversion influences auditor judgments. The Birnbaum and Stegner [1979] scale-adjustment model demonstrates that the expertise of one source and the perceived objectivity of a *competing* information source interactively affect judgments. Specifically, the same degree of concern about one source's expertise creates a trade-off that is *even greater* when the competing source is perceived to be more objective. Thus, if auditors are concerned about the adequacy of their firm's AI system (i.e., aversion), then the Birnbaum and Stegner [1979] model suggests that algorithm aversion effects will be amplified when management's competing evidence appears relatively more objective.

Although all complex accounting estimates are inherently subjective, the underlying information management uses to develop these estimates can contain varying degrees of objectivity/subjectivity (e.g., Kadous, Koonce, and Towry [2005], Bratten et al. [2013], Joe, Vandervelde, and Wu [2017]). For example, management can rely on inputs that are relatively more objective (e.g., verifiable external data) and/or rely on inputs that are relatively more subjective (e.g., management assumptions; Kadous, Koonce, and Towry [2005], FASB [2010], Griffin [2014], PCAOB [2017b]). Further, research suggests auditors will find objective inputs to be more persuasive than subjective inputs (e.g., Kadous, Koonce, and Towry [2005], Joe, Vandervelde, and Wu [2017]). Accordingly, we predict that algorithm aversion will have a greater impact on auditors' judgments (i.e., proposed adjustments will be even smaller) when management uses objective (vs. subjective) inputs to develop their estimates.

We conduct an experiment with 170 audit senior participants, manipulating the source of audit firm-provided evidence (human specialist vs. AI system) and the nature of the inputs used by management (subjective vs. objective) to develop a complex estimate. Participants in all conditions receive the same audit evidence from their own firm's specialist (based on the same underlying inputs) regarding a banking client's allowance for loan losses ("ALL"). This evidence contradicts management's evidence (i.e., "contradictory evidence") and suggests that the client's ALL is materially understated (i.e., net income is materially overstated). We manipulate the source of the audit firm-provided contradictory evidence as either the in-house valuation group (i.e., human specialist) or the firm's proprietary AI system. Importantly, we use identical language to describe the accuracy and reliability of the audit evidence derived from each firm source. We also manipulate the nature of the inputs used by management. In the subjective inputs condition, management relies heavily on the judgment of loan officers and credit analysts, who use a variety of methods and information

(e.g., discussions with real estate brokers) to develop estimates for a key parameter in the ALL estimate (i.e., collateral values). In the objective inputs condition, management relies on client-selected, detailed market data (i.e., real estate price indices) to update collateral values in a standardized manner. Participants' proposed audit adjustments serve as our dependent measure.

Consistent with our theory-based expectations, auditors recommend smaller proposed adjustments, on average, when firm-provided evidence comes from an AI system versus a human specialist (i.e., exhibiting algorithm aversion). We also find that the effect of algorithm aversion is stronger when management uses inputs that are more objective (vs. subjective) to develop their estimates. Importantly, our results suggest that algorithm aversion could have economically significant implications. On average, mean proposed adjustments are 23% lower in the AI system condition than in the human specialist condition. Furthermore, when management uses inputs that are relatively more objective this difference increases, such that mean proposed adjustments are 43% lower in AI system condition than in the human specialist condition.

We also conduct additional analyses using a moderated mediation model to examine the psychological process through which algorithm aversion affects auditors' adjustment decisions. Mediation results reveal that when auditors receive evidence from an AI system instead of a human specialist, they are more concerned that the firm source (i.e., the AI system) might lack the necessary knowledge and expertise. Consequently, auditors also believe the system's evidence will be less convincing to management. However, these concerns about expertise and the ability to convince management only lead to smaller proposed adjustments when management relies on inputs that are relatively more objective. Overall, these results suggest the effects of algorithm aversion on auditors' adjustment decisions are amplified when management's evidence is more persuasive to auditors (i.e., when evidence is more objective).

Our study makes several important contributions and extends two streams of research. First, we contribute to the growing literature around auditing complex accounting estimates. Measurement uncertainty continues to be a critically important risk for financial reporting stakeholders (Bratten et al. [2013], Dennis, Griffin, and Zehms [2019]). 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, Vandervelde, and Wu [2017]). Recent findings indicate that auditors are willing to discount (and perhaps even ignore) contradictory evidence from valuation specialists (PCAOB [2017a], Griffith [2020]). Our findings suggest that receiving evidence from an AI system likely exacerbates this auditor tendency to discount contradictory evidence around management's complex estimates. Thus, the adoption of these AI systems could increase the likelihood of auditors accepting their clients' potentially biased estimates, particularly when management

develops their estimates using inputs that appear relatively more objective in nature.

Second, we contribute to a growing body of research in psychology and management science that documents individuals' reluctance to rely on computer-generated advice. This study is the first to provide evidence of algorithm aversion in auditors' judgments, demonstrating that this phenomenon manifests even for experienced professionals who have strong incentives to rely on the related computer-generated evidence. We also show that, at least in the audit setting, the effects of algorithm aversion are partly attributable to source credibility (i.e., expertise) concerns. We also extend the algorithm aversion literature by demonstrating that the characteristics of competing information sources can influence the extent to which algorithm aversion affects individuals' judgments. Specifically, when a competing information source appears more objective (i.e., less biased) the effects of algorithm aversion are amplified. Our findings have interesting implications for other, non-audit settings in which AI models are beginning to emerge, such as forecasting, lending, and financial reporting (e.g., Ding et al. [2020], Chen et al. [2021], Liu [2021]). These studies find that AI models can outperform human-based models in these settings. Our results highlight that, beyond performance concerns, care and attention in implementing AI is warranted because getting professionals to rely on these AI-based models can be a significant barrier to realizing the benefits of this technology.

Our findings are especially relevant to audit firms and regulators. Although the audit environment is quickly evolving due to technological advancement (Raphael [2017], Tysiac [2017]), auditing standards do not yet contemplate the implications of autonomous and unsupervised AI systems (e.g., PCAOB [2020]). Therefore, our investigation of narrow AI in this setting provides insights that are relevant to the current state of audit firm methodologies and auditor regulation. We provide evidence that the implementation of AI 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 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 and have already begun implementing these technologies to assist auditors with relatively basic and routine audit tasks, such as confirmation testing and inventory observations (PwC [2017], Deloitte [2018b], Cooper et al. [2019], Austin et al. [2020], Christ et al. [2021]). The most advanced technology

under consideration is AI, which firms believe can assist auditors on some of the most challenging audit tasks (Murphy [2017]). Currently, the goal is that these AI systems will make the same judgments that auditors and human specialists would, but through the use of machine learning to synthesize and process large amounts of diverse and unstructured data. However, consistent with concerns about the potential for increased regulatory scrutiny and legal liability associated with the use of more advanced audit technologies (e.g., Christ et al. [2021], Emett et al. [2021]), firms are not yet considering the use of fully autonomous, unsupervised AI systems to produce audit evidence. Rather, firms describe their current ambitions as "narrow AI," which performs specific tasks within certain programmable parameters (PwC [2017]). For example, KPMG is developing an AI system to help auditors evaluate commercial loan grades (KPMG [2016]). 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 respond to evidence provided by these systems.[2]

2.2 CONTRADICTORY AUDIT EVIDENCE AND ALGORITHM AVERSION

When evaluating complex estimates, auditors must consider the evidence from their firm's specialists in conjunction with the evidence provided by management to form an opinion about whether the client's estimates are fairly stated (PCAOB [2010, 2014a, b], Hanson [2012], CPAB [2015a, b], IFIAR [2015], PCAOB [2017a]). Due to the high degree of uncertainty associated with complex estimates, auditors often receive evidence from their firm specialists that contradicts management's estimates (Griffith, Hammersley, and Kadous [2015], Cannon and Bedard [2017], Griffith [2020]). When firm specialists provide evidence that contradicts management's estimates, auditors must use their professional judgment to incorporate the competing evidence into their decisions (PCAOB [2015, 2017a, b]). The rapid advancement and planned use of audit technologies point to a future in which AI systems are a key source of contradictory audit evidence. Although firms anticipate that these AI systems will improve audit outcomes, theory suggests that auditors will respond to their own firm's evidence differently when it comes from a system instead of a human specialist.

Specifically, a growing body of research across multiple fields (psychology, medicine, and management science) indicates that individuals are

---

[2] It is important to note that AI systems differ from decision aids of the past in the degree to which these systems can incorporate diverse, unstructured data sets, and the degree to which they can adapt over time based on those data sets (i.e., machine learning). AI systems are increasingly able to simulate human judgment–and auditors are expected to incorporate the evidence produced into their decisions. Thus, prior research examining 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 does not offer clear implications for 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]).

susceptible to "algorithm aversion"–a tendency to rely more heavily on information produced by a human source than information produced by a computer source (e.g., Promberger and Baron [2006], Önkal et al. [2009], Eastwood, Snook, and Luther [2012], Dietvorst, Simmons, and Massey [2015], Yeomans et al. [2019]). 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 from 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 demonstrates the algorithm's predictions are consistently more accurate than their own (Dietvorst, Simmons, and Massey [2015]).

Recent research also documents algorithm aversion effects in highly subjective settings (Castelo, Bos and Lehmann [2019], Yeomans et al. [2019]). For example, Yeomans et al. [2019] show that although an algorithmic joke recommendation system outperforms human recommenders, individuals are reluctant to rely on advice from the system when making recommendations to others and when receiving recommendations for themselves. Importantly, Castelo, Bos, and Lehmann [2019] posit that algorithm aversion effects occur in subjective task settings because individuals believe algorithms lack the necessary abilities (e.g., human intuition) to perform subjective tasks effectively. Although existing research has yet to fully explain why individuals are susceptible to algorithm aversion, the Castelo, Bos and Lehmann [2019] findings suggest that algorithm aversion effects might arise from beliefs that algorithmic information sources are inherently less capable than human sources. Informed by the algorithm aversion literature, we expect auditors will more heavily discount audit evidence produced by AI systems than evidence produced by human specialists. Moreover, consistent with Castelo, Bos, and Lehmann [2019], we theorize that this discounting effect is due, in part, to auditors' beliefs that AI systems lack the necessary capability (i.e., knowledge and expertise) to evaluate complex estimates. These capability concerns might lead to the additional concern that management will be unwilling to adjust their estimates based on the contradictory evidence provided by an AI system. Consequently, auditors might be more hesitant to propose and negotiate adjustments related to management's complex estimates.

Individuals tend to place less weight on information sources that are perceived to be less credible (e.g., lower expertise; Birnbaum [1976], Birnbaum, Wong, and Wong [1976], Birnbaum and Mellers [1983], Hirst [1994], Griffith [2018]).[3] Furthermore, the Birnbaum and Stegner [1979] scale-adjustment model posits that, much like a traditional balance scale,

---

[3] Source credibility theory identifies objectivity and expertise as key factors that increase the perceived credibility of a source (Hovland, Janis, and Kelley [1966], O'Reilly and Roberts [1976], Birnbaum and Stegner [1979], Pornpitakpan [2004]). Audit research documents that

discounting one source's information causes a proportional *increase* in the weight placed on *competing* information. For example, Birnbaum and Stegner [1979] ask individuals to determine the value of a used car based on estimates provided by two sources, each of whom had previously examined the car. Results indicate that a lower level of expertise (i.e., a novice mechanic vs. an expert mechanic) causes individuals to more heavily discount the estimated value from that source *and* more heavily weight the estimated value provided by the second source. Applying the Birnbaum and Stegner [1979] scale-adjustment model, and theory on algorithm aversion, to the context of auditing complex estimates implies auditors will discount firm-provided contradictory evidence more heavily (and thus more heavily weight management-provided evidence) when that evidence comes from an AI system instead of a human specialist. Thus, the adoption of AI systems could have an unfavorable effect on auditor judgments. For example, auditors' consideration of the available evidence (from both management and firm sources) determines their beliefs about whether and to what extent management's estimates should be adjusted. Placing more weight on management-provided evidence and less weight on firm-provided evidence results in auditors being more accepting of management's preferred estimates, leading to smaller proposed audit adjustments. Formally, we predict the following:

> *H1*: Auditors will propose smaller adjustments to management's estimates when evidence contradicting management's preferences is received from a firm AI system instead of a human specialist.

2.3 OBJECTIVITY AND THE AMPLIFICATION OF ALGORITHM AVERSION

Complex accounting estimates are inherently subjective because they are forward-looking, and managers can incorporate a variety of inputs and assumptions that are difficult to verify (Bratten et al. [2013]). However, the nature of estimation inputs can vary, with some inputs appearing more objective than others (Kadous, Koonce, and Towry [2005], FASB [2010], Griffin [2014], PCAOB [2017b]). For example, inputs based on external data sources appear more objective than inputs based on management's projections and assumptions. Research suggests that accounting professionals have a strong preference for well-defined, more objective evidence/tasks over subjective and ambiguous evidence/tasks (Bamber, Snowball, and Tubbs [1989], Zimbelman and Waller [1999], Kadous, Koonce, and Towry

---

auditors consider expertise and competence to be key source credibility factors when deciding whether to rely on others' work (e.g., Bamber [1983], Griffith [2018]). Building on this, our study examines whether auditors weight their own firm's independent evidence differently depending on whether the evidence source is a firm-employed human specialist or a firm-developed AI system. Given that these sources are equivalently independent from the client, we expect that differences in auditors' weighting of evidence from human specialists versus AI systems will more strongly reflect concerns around expertise, relative to concerns around objectivity.

[2005], Joe, Vandervelde, and Wu [2017]). Furthermore, audit standards identify a higher degree of subjectivity in inputs to management's estimates as an indicator of potential bias (PCAOB [2017b]). Consistent with this regulatory caution, research finds that auditors are less likely to adjust management estimates that are more precise and objective in nature (Nelson, Elliott, and Tarpley [2002], Nelson, Smith, and Palmrose [2005], Griffin [2014]). Relatedly, Kadous, Koonce, and Towry [2005] find that quantification positively influences the persuasiveness of accounting managers' proposals through greater perceived competence of the preparer, but that the influence of quantification on persuasiveness only occurs when proposals are based on objective (rather than subjective) inputs. Overall, these findings suggest auditors prefer objective information and will be more likely to agree with management's estimates when those estimates are based on inputs that are more objective in nature.

Importantly, Kadous, Koonce, and Towry [2005] also demonstrate that objectivity can interact with other credibility factors (i.e., perceived competence). Relatedly, the Birnbaum and Stegner [1979] scale-adjustment model demonstrates that, when there are two competing information sources, the expertise of one source and the perceived objectivity of the competing information source interactively affect judgments. Our first hypothesis follows the findings of Birnbaum and Stegner [1979] that a decrease in one source's expertise (i.e., expert vs. novice mechanic) can cause a trade-off between the two information sources. However, their results also indicate that the trade-off created by a decrease in the first source's expertise is *even greater* when the second source appears more objective. For example, a car buyer will discount a novice mechanic's car value estimate even more heavily (relative to an expert mechanic) when the competing second source is believed to be more objective (e.g., blue book value vs. friend of the seller).

Given that algorithm aversion is likely due, in part, to concerns about capability or expertise (e.g., Castelo, Bos, and Lehmann [2019]), the findings of Birnbaum and Stegner [1979] suggest that algorithm aversion effects will be amplified (i.e., even greater) when there is a competing source that appears relatively more objective. Thus, although H1 suggests that receiving contradictory evidence from an AI system (instead of a human specialist) will cause auditors to propose smaller adjustments, theory further suggests this effect will be *stronger* when management uses inputs that are relatively more objective (vs. subjective) in nature, causing proposed adjustments to be even smaller.[4] This interaction prediction is formalized as follows:

H2: Receiving contradictory evidence from an AI system instead of a human specialist will reduce proposed adjustments to a greater extent when management's estimates are based on inputs that are relatively more objective than subjective.

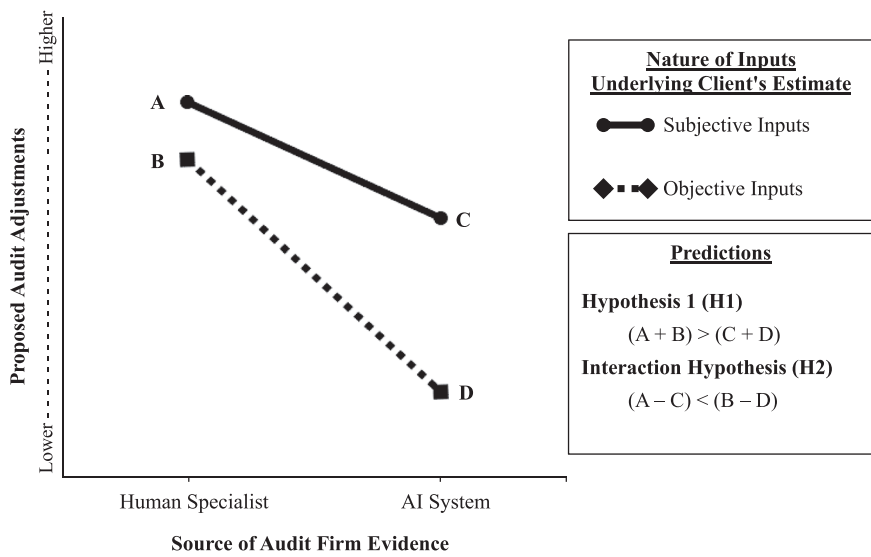[4] Figure 1 depicts the interactive effect suggested by H1 and H2.

Fig. 1.—Illustration of predictions. H1 predicts that auditors will propose smaller adjustments to management's estimates when evidence contradicting management's preferences is provided by an AI system compared to a human specialist (i.e., algorithm aversion). H2 further proposes that the effect predicted in H1 will be stronger when management uses inputs that are relatively more objective (vs. subjective) in nature, causing auditors to propose smaller proposed adjustments. Although not formally predicted, regardless of firm source, we expect when management develops their estimate based on objective inputs, this will increase the perceived objectivity of management's estimate, causing auditors to propose smaller adjustments. In conjunction, H1 and H2 suggest an ordinal interaction will occur.

TABLE 1
Demographic Profile of Participants

Mean public accounting experience (in years)	4.02
Percentage of participants who have experience with banking clients	30%
Average percentage of time working on banking clients	44%
Percent of participants' time working on public clients	50%
Likelihood of providing input for proposed audit adjustments	5.10
(1 = "Not at All Likely" and 7 = "Highly Likely")	
Percentage of participants who have experience using an AI system	4%

3. Methods

3.1 PARTICIPANTS

One hundred and seventy audit seniors from two Big 4 firms participated in this study. Table 1 presents participants' demographic data. Participants have an average of 4.02 years of public accounting experience (SD = 1.19),

with a range of 1–10 years.[5] Importantly, 96% of participants report three years or more of public accounting experience.[6] Additionally, 30% of participants in our sample report experience working on audits of banking clients, indicating that a fair portion of our sample has relevant industry-specific experience. Participants report spending approximately 50% of their time working on public clients and are relatively likely to provide input into decisions regarding proposed audit adjustments in a typical year (mean of 5.10 on a scale from 1 = "Not at All Likely" and 7 = "Highly Likely"). Overall, these participants have experience consistent with auditors who are typically involved with auditing complex estimates (e.g., Griffith, Hammersley, and Kadous [2015], Griffith et al. [2015]). Very few participants (4%) report experience working with AI systems on an audit engagement, consistent with our understanding that audit firms have not yet fully implemented these systems.

3.2 EXPERIMENTAL AUDIT CASE

Prior to conducting our study, we gathered extensive information to fully understand audit firms' plans and intentions around the use of AI. Specifically, we consulted with senior audit partners at different international accounting firms, including a partner that specializes in auditing banks. Each of these partners indicated that their respective firm has a particularly strong interest in developing AI to replicate the judgments that auditors and specialists' make on financial services engagements. Notably, we consulted with the National Innovation Leader of a Big 4 audit firm. This Innovation Leader described their firm's development of a system that can be used to produce loan grades for portfolios of commercial real estate loans. Additionally, after designing our experiment, senior partners at multiple Big 4 firms reviewed and validated our experimental context and instrument.

In our experiment, 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 developing independent loan grades.

---

[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, which we omit from the study because these individuals do not possess the requisite experience for the experimental task. We also exclude 29 responses from participants who provided incorrect answers to the manipulation check questions. See subsection 4.4 for additional discussion regarding these manipulation check failures.

[6] Reported inferences about hypotheses are unchanged if data from participants with less than three years of experience are excluded.

We manipulate the source of these independent loan grades as either the firm's in-house human valuation specialist or the firm's AI system (i.e., the *Source* manipulation described later in subsection 3.3).

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 reduce earnings by $28 million. For simplicity, we do not provide participants with the ALL balance. Instead, in all conditions, the experimental materials state that the potential audit difference is greater than the materiality threshold, indicating this issue warrants significant attention. Case details indicate that these differences relate to disagreements about estimated collateral values, which are a key parameter 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" (i.e., outdated) as of the balance sheet date—and therefore need to be adjusted upward. We manipulate the nature of the inputs that Heartland's management uses for these adjustments as relatively more objective or relatively more subjective (i.e., the *Nature of Inputs* manipulation described later in subsection 3.3) to roll forward collateral value estimates. However, in all conditions, Heartland adjusts these collateral values more aggressively than Clark & Miller, resulting in higher estimated collateral values.

Participants then view side-by-side reports from Clark & Miller and Heartland summarizing the respective methodologies that each uses to roll-forward collateral values from the appraisal date to the balance sheet date. Ultimately, Heartland's methodology results in significantly higher collateral values than Clark & Miller's methodology. In all conditions, management engages in a benchmarking exercise to evaluate how their collateral value updates compare to national price changes. Although details seeded throughout the case indicate that there are three highly reputable price indices that can be used, in all conditions, management takes the same aggressive approach in its benchmarking and relies only on the two most favorable price indices. In contrast, Clark & Miller's specialist report primarily focuses on the two most conservative price indices. Therefore, although Heartland's management asserts that their estimated collateral value increases are reasonable based on their analysis, there are subtle indications that management could have conducted this benchmarking exercise in an optimistically biased manner, and relatedly could have estimated collateral values in a biased manner (i.e., resulting in a lower ALL balance and higher earnings).

Following these reports, participants view 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 (income-decreasing) adjustment to the ALL. Immediately following this

summary, participants recommend a proposed adjustment.[7] Proposed adjustments serves as our dependent measure because this judgment reflects the auditor's consideration and relative weighting of the available evidence (from both management and firm sources). Finally, participants complete a postexperiment questionnaire.

3.3 INDEPENDENT VARIABLES

We use a 2×2 between-subjects experimental design, manipulating the source of an audit firm's evidence around a complex estimate (i.e., "*Source*") and the nature of the inputs used by management to generate the estimate (i.e., "*Nature of Inputs*"). We manipulate *Source* at two levels: human specialist and AI system. In the human specialist condition, the audit firm employs an internal group of human specialists who develop independent loan grades for individual loans. In the AI system condition, we inform participants that Clark & Miller utilizes a proprietary AI system called the "Amadeus System" that develops these grades. The description of the firm's loan grade calculations and the resulting valuation reports (i.e., audit evidence) from these two sources are identical in both conditions. For example, in both *Source* conditions the firm specialist (i.e., the human specialist or the AI system) considers the specific property-type and location when relying on detailed market data (i.e., price indices). Additionally, the firm specialist uses other market indicators (e.g., vacancy rates, rent levels, new construction, and comparable sales) to further adjust collateral values.

Importantly, while AI systems can use different information and methodologies than human specialists, we made the conscious decision to hold the experimental language around this methodology constant across conditions in order to maintain experimental control. Specifically, we designed identical wording to describe the method of updating collateral values in both *Source* conditions, such that the language could be interpreted differently when attributed to a human specialist versus an AI system. Our design choices were informed and validated through the consultations with senior audit partners, as described in subsection 3.2. As a result, our experiment achieves experimental realism while maintaining internal validity (Swieringa and Weick [1982], Peecher and Solomon [2001]).

When describing the sources of firm evidence, we include statements in both treatments to equalize the acceptability of the two sources and legitimize the AI system in the same way that audit firms legitimize human specialists in practice (see the appendix). For example, we inform participants in both *Source* conditions that the audit firm considers the loan grades developed by the respective source to be an approved source of audit evidence. This information helps familiarize participants with the respective

---

[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 hereafter refer to this measure as "proposed adjustments."

source and ensures that observed effects are due to the human/nonhuman nature of the information source rather than reluctance to deviate from more traditional firm methodologies.

Next, we manipulate the *Nature of Inputs* (used by management) at two levels: objective inputs and subjective inputs (e.g., Kadous, Koonce, and Towry [2005]).[8] In the objective inputs condition, Heartland's management relies on detailed, verifiable market data (i.e., real estate price indices) to roll forward collateral value estimates. In the subjective inputs condition, management relies on the judgment of its loan officers and credit analysts to update collateral values using a variety of methods. For example, these loan officers and credit analysts can update the most recent appraised value based on their own research (e.g., comparable sales, local market trends, discussions with real estate brokers).

It is important to note that while subjective inputs can provide management with more degrees of freedom to introduce bias than objective inputs (e.g., Choudhary [2011], Ramanna and Watts [2012], Bratten et al. [2013]), management can effectively bias its collateral value estimates in both *Nature of Inputs* conditions. Specifically, in the objective inputs condition, adding, replacing, or removing a single price index can materially alter the ALL calculation at the portfolio level. In the subjective inputs condition, systematically biased judgments at the individual loan level could accumulate to a material misstatement of the ALL.[9]

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 *Nature of Inputs* manipulation, we varied whether management relies on the systematic application of detailed market data (i.e., objective inputs) or on the judgment and expertise of loan officers and/or credit analysts (i.e., subjective inputs). Accordingly, we asked participants to identify management's methodology and 18 additional participants were unable to do so correctly. As 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.[10]

---

[8] The *Nature of Inputs* manipulation relates only to inputs used by Heartland management. The audit firm (Clark & Miller) uses the same inputs in all conditions.

[9] Research consistently finds that management uses the unverifiable discretion afforded to them in the accounting standards to estimate fair values in a way that opportunistically inflates earnings and stock prices (e.g., Benston [2006], Hilton and O'Brien [2009], Dechow, Myers, and Shakespeare [2010], Choudhary [2011], Ramanna and Watts [2012], Bratten et al. [2013], Li and Sloan [2017]).

[10] We eliminate these responses because we view these multiple-choice questions as straightforward measures of participants' attentiveness to the experimental materials (both

To further examine the effectiveness of our *Nature of Inputs* manipulation, we ask participants, "to what extent do you believe management's process for estimating updated collateral values was objective versus subjective?" (reverse-coded to 1 = "Very Subjective" and 7 = "Very Objective"). Participants in the objective inputs condition reported that management's estimation process was more objective (i.e., less subjective), on average, than participants in the subjective inputs condition (3.89 vs. 3.02, respectively; $p < 0.01$, one-tailed, untabulated), further indicating that *Nature of Inputs* is successfully manipulated in our final sample.[11]

We also ask participants whether 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 AI system conditions report similarly high levels of agreement with that statement (5.92 vs. 5.61, respectively; $p > 0.10$, two-tailed, untabulated). Further, mean responses in both conditions are significantly higher than the scale midpoint ($p < 0.01$, two-tailed, untabulated). Thus, participants in both *Source* conditions were similarly assured that their respective source was deemed reliable and approved by the firm.

4.2 TESTS OF HYPOTHESES

We test our hypotheses using a 2×2 analysis of variance (ANOVA) model with auditors' proposed adjustments ("*Proposed*") as the dependent variable and *Source* and *Nature of Inputs* as independent variables. Table 2, panel A, presents descriptive statistics for *Proposed*, Table 2, panel B, reports ANOVA results, and Figure 2 provides a graphical illustration of these results.[12]

H1 predicts that auditors will propose smaller adjustments to management's estimates when their own firm's contradictory evidence comes from an AI system versus a human specialist. Furthermore, H2 predicts that this effect will be larger when management uses inputs that are relatively more objective (vs. subjective) to develop their estimates. Consistent with H1, we find a significant main effect of *Source* on *Proposed* ($F_{1,166} = 11.29$, $p < 0.01$, one-tailed), indicating that proposed adjustments are, on average, higher in the human specialist condition than in the AI system condition. As indicated in Table 2, panel A, the mean for *Proposed* is only $16.12 million in the AI system condition versus a mean of $21.02 million in the human

---

manipulations are highly salient and appear in multiple places throughout the instrument). We discuss the implications of these manipulation check failures in subsection 4.4.

[11] For directional predictions, we report *p*-values that are one-tailed equivalents in the text.

[12] We perform the Levene [1960] test for equality of variances and the results are 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 data. However, consistent with the nature of proposed audit adjustment decisions, the Shapiro–Wilk test indicates that our *Proposed* variable is not normally distributed ($W = 0.94$, $p < 0.01$, two-tailed; untabulated). We conduct a nonparametric bootstrapping procedure using the Process macro (model 1, untabulated) to examine whether our results are robust to the potential implications of this nonnormality, and our inferences are unchanged.

TABLE 2
Proposed Audit Adjustments

Panel A: Descriptive statistics: mean (standard error) [n] cell

	Human Specialist	AI System	Overall
Subjective inputs	22.13	19.94	21.06
	(1.52)	(1.55)	(1.02)
	[47]	[45]	[92]
	A	C	
Objective inputs	19.81	11.20	15.95
	(1.58)	(1.76)	(1.35)
	[43]	[35]	[78]
	B	D	
Overall	21.02	16.12	18.72
	(1.07)	(1.29)	(0.85)
	[90]	[80]	[170]

Panel B: Conventional ANOVA

Source of Variance	Sum of Squares	df	F	p[a]
Source	1,223.91	1	11.29	<0.01
Nature of Inputs	1,283.77	1	11.84	<0.01
Source × Nature of Inputs	433.25	1	4.00	0.05
Error	17,997.05	166		

Panel C: Simple effects of source

	df	t	p[b]
Across subjective inputs: Cell A > Cell C	1, 166	1.01	0.16
Across objective inputs: Cell B > Cell D	1, 166	3.63	<0.01

The dependent variable ("*Proposed*") is the amount of participants' proposed audit adjustments to the ALL on a scale from $0 million (no adjustment) to $28 million (full adjustment of the difference between the firm's estimate and the client's estimate). We manipulate, between participants, the source of the audit firm's evidence ("*Source*") at two levels (Human Specialist vs. AI System). We also manipulate, between participants, the nature of inputs that management uses to develop their estimate ("*Nature of Inputs*") at two levels (subjective inputs vs. objective inputs).

[a] As ANOVA F-tests are nondirectional, we report two-tailed p-values in panel B. Consistent with our directional predictions, we report one-tailed equivalent p-values in the text. We derive these one-tailed equivalent p-values from the ANOVA contrast t-statistics, which are equivalent to the square roots of the respective F-statistics (see, e.g., Kachelmeier and Williamson [2010]; Piercey [2011]; Saiewitz and Kida [2018]).

[b] The p-values in panel C are one-tailed, consistent with our directional predictions.

specialist condition, yielding a $4.90 million difference across conditions.[13] Thus, we find that, on average, proposed adjustments are 23% lower in the AI system condition versus the human specialist condition and 14% lower than the overall sample mean of $18.72 million.

Consistent with H2, we find a significant interaction between *Source* and *Nature of Inputs* ($F_{1,166} = 4.00$, $p = 0.02$, one-tailed). To further evaluate this interaction, we also examine the simple effects of *Source*. As reported in

---

[13] Relative to the potential audit difference of $28 million, the mean proposed adjustment in the AI system condition ($16.12 million) is about 58% of the maximum potential adjustment amount. In contrast, the mean proposed adjustment in the human specialist condition ($21.02 million) is about 75% of the maximum potential adjustment.
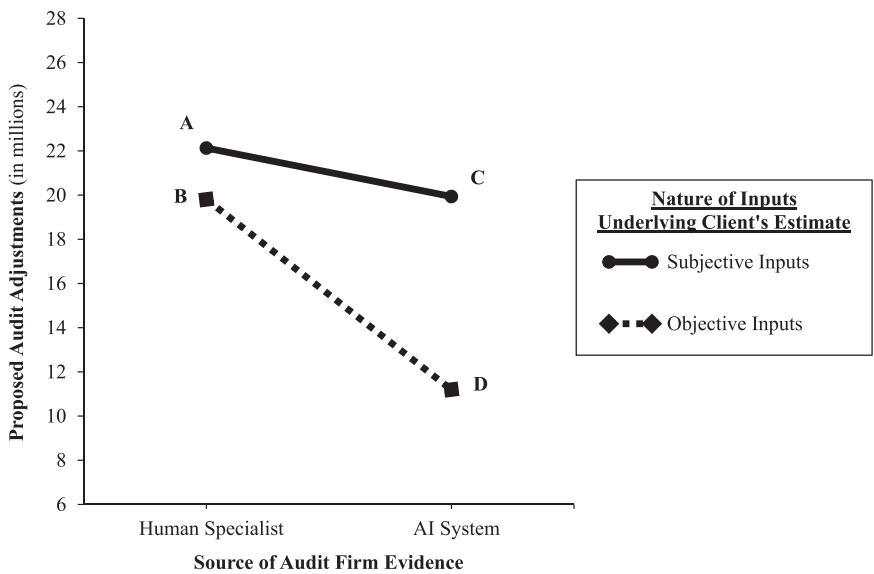
FIG 2.—Observed effects of source and nature of inputs on proposed audit adjustments. See notes to Table 2 for descriptions of dependent variable and independent factors.

Table 2, panel C, the simple effect of *Source* is significant in the objective inputs condition ($t_{166} = 3.63$, $p < 0.01$, one-tailed). Although the simple effect of *Source* in the subjective inputs condition is not statistically significant ($t_{166} = 1.01$, $p = 0.16$, one-tailed), the overall mean pattern depicted in Figure 2 is visually consistent with our expectations (see Figure 1). As noted earlier, the mean proposed adjustment is \$4.90 million lower in the AI system condition than in the human specialist condition. However, this mean difference increases to \$8.61 million when management's inputs are more objective (\$19.81 million vs. \$11.20 million) and decreases to \$2.19 million difference when management's inputs are more subjective (\$22.13 million vs. \$19.94 million), consistent with the predicted interaction effect. Notably, the mean proposed adjustment in the AI system/objective inputs condition (\$11.20 million) is 43% lower than the mean in human specialist/objective inputs condition (\$19.81 million) and 40% lower than the overall sample mean (\$18.72 million). Given that participants were instructed that the \$28 million audit difference was material, these differences are quite significant and suggest that algorithm aversion can have a meaningful influence on amounts reported within the financial statements. Collectively, our results provide support for H1 and H2.[14]

---

[14] We gather data on participants' public accounting experience and position, as well as their specific experience with public clients, banking clients, artificial intelligence, and proposed adjustment decisions. We run several models to examine whether any of our experience

4.3 MODERATED MEDIATION ANALYSES

In developing our hypotheses, we draw on theory that suggests algorithm aversion might be partially attributable to concerns about the capabilities (e.g., expertise) of AI systems (Birnbaum and Stegner [1979], Castelo, Bos, and Lehmann [2019]). This is particularly plausible in the audit setting because source expertise/competence is a key consideration for auditors when evaluating audit evidence (Bamber [1983], Hirst [1994], Griffith [2018]). Furthermore, in the audit setting, auditors must use the available audit evidence to "justify proposed adjustments and persuade management that the current estimate is unreasonable" (Griffith [2018, p. 178]). Thus, expertise concerns about the firm's source of contradictory evidence could lead auditors to worry that they might have difficulty convincing management to adjust their estimate based on the contradictory evidence. In this section, we provide additional analyses to demonstrate that these mechanisms drive the effects we observe in our tests of H1 and H2. Accordingly, we examine whether the effect of *Source* on *Proposed* operates sequentially through participants' concerns about the firm specialist's expertise (i.e., *Source Expertise Concern)* and their perceptions of the likelihood that management can be convinced to book an adjustment based on the specialist's evidence (i.e., *Convincing to Management*).[15] Additionally, because our theory predicts the effect of algorithm aversion will be amplified (stronger) when management uses inputs that are relatively more objective (vs. sub-

———————

measures influence the effects we observe and find that this is not the case. First, we run a series of six 2×2 analysis of covariance (ANCOVA) models that control for each of these demographic measures. The main effects and interactions in all models (untabulated) have the same significance levels as the main effects and interaction in our main analysis. We then adapt the 2×2 ANOVA model from our main analyses (i.e., with proposed adjustments as the dependent variable) and run six new models, each with a three-way interaction between our two manipulated variables and one of the respective demographic measures (where demographic measures are mean-centered continuous variables). In all of these models, the three-way interaction is insignificant (all two-tailed p-values are greater than 0.35). Furthermore, the main effects of our two manipulated variables, as well as the predicted two-way interaction between these variables, remain significant at the same levels as in our main analysis. Collectively, these sensitivity tests indicate that none of our experience measures influence the effects we observe. In addition, we run a series of models on the full sample ($n = 199$) to test whether there are systematic demographic differences across conditions, despite our use of random assignment. We find no significant main effects or interactions in any of these models, indicating that none of our demographic measures differ systematically across conditions (all p-values > 0.19, two-tailed).

[15] Both process measures (i.e., *Source Expertise Concern* and *Convincing to Management*) are collected as part of the post experiment questionnaire. To capture participants' perceptions of *Source Expertise Concern*, we asked participants to rate the extent to which they were concerned about the firm specialist's expertise (1 = "Not at All Concerned" and 7 = "Very Concerned"). To measure *Convincing to Management*, we asked participants to assess the likelihood that management could be convinced to adjust an estimate based on evidence provided by the firm specialist (1 = "Highly Unlikely" and 7 = "Highly Likely"). One participant did not respond to the *Convincing to Management* measure, and thus, reported mediation analyses include 169 observations.

TABLE 3
Descriptive Statistics for Mediator Variables

Panel A: *Source Expertise Concern*: mean (standard error) [*n*] cell

	Human Specialist	AI System	Overall
Subjective inputs	2.89	4.16	3.51
	(0.20)	(0.21)	(0.15)
	[47]	[45]	[92]
	A	C	
Objective inputs	2.90	4.63	3.69
	(0.21)	(0.24)	(0.19)
	[42]	[35]	[77]
	B	D	
Overall	2.90	4.36	3.59
	(0.15)	(0.16)	(0.12)
	[89]	[80]	[169]

Panel B: *Convincing to Management*: mean (standard error) [*n*] cell

	Human Specialist	AI System	Overall
Subjective inputs	3.96	3.38	3.67
	(0.22)	(0.23)	(0.15)
	[47]	[45]	[92]
	A	C	
Objective inputs	3.38	2.66	3.05
	(0.23)	(0.26)	(0.19)
	[42]	[35]	[77]
	B	D	
Overall	3.69	3.06	3.39
	(0.16)	(0.17)	(0.12)
	[89]	[80]	[169]

This table provides descriptive statistics for the two mediators presented in Figure 3. *Source Expertise Concern* measures participants' assessment of how concerned they were about the firm specialist's expertise (1 = "Not at All Concerned" and 7 = "Very Concerned"). *Convincing to Management* measures participants' assessment of the likelihood that management can be convinced to book an adjustment based on the specialist's evidence (1 = "Highly Unlikely" and 7 = "Highly Likely").

jective), we also examine whether and how this sequential indirect effect on auditors' proposed adjustments differs depending on the nature of management's inputs (i.e., moderated mediation). We present descriptive statistics for *Source Expertise Concern* and *Convincing to Management* in Table 3, panels A and B, respectively.

Following the procedures described by Hayes [2018], we conduct a moderated mediation analysis using the SPSS PROCESS macro (Model 89) with participants' evaluations of *Source Expertise Concern* and *Convincing to Management* as serial mediators. Figure 3 presents our results and describes our process measures in more detail. Results indicate that in the objective inputs condition, the effect of *Source* on *Proposed* operates indirectly through *Source Expertise Concern* and *Convincing to Management* (90% confidence interval of −1.48 to −0.11, indicating a one-tailed *p*-value less than 0.05), though this sequential indirect effect is not significant in the subjective

Indirect Effects by *Nature of Inputs*
Nature of Inputs = Subjective: **-0.16**
  (Bootstrapped CI: **-0.67 → 0.24**)[a]
Nature of Inputs = Objective: **-0.70****
  (Bootstrapped CI: **-1.48 → -0.11**)[a]

Index of Moderated Mediation (Subjective vs. Objective): -0.54*
  (Bootstrapped CI: **-1.11 → -0.03**)[a]

$\beta_b = -0.21$
$p < 0.01$

Source Expertise Concern → *Convincing to Management*

$\beta_a = 1.46$
$p < 0.01$

$\beta_d = -0.31$
$p = 0.12$

$\beta_f = -1.21$
$p = 0.02$

$\beta_e = 1.32$
$p < 0.01$

$\beta_{wf} = 2.13$
$p = 0.03$

$\beta_{c'} = -2.58$
$p = 0.07$

Source (1=AI System; 0=Human Specialist)

$\beta_{wc'} = -7.56$
$p = 0.02$

$\beta_{we} = 1.74$
$p = 0.05$

Proposed Adjustments

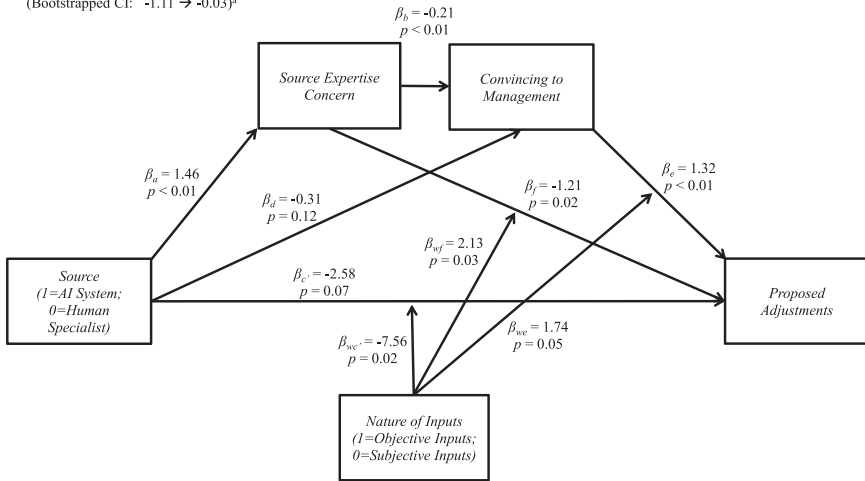*Nature of Inputs (1=Objective Inputs; 0=Subjective Inputs)*

FIG 3.—Moderated mediation analysis. The diagram represents a moderated sequential mediation model (Hayes [2018]). We use a Model 89 with two mediators. Specifically, the effect of *Source* on *Proposed* is expected to operate sequentially through concerns about the expertise of firm specialist (i.e., *Source Expertise Concern*) and the perceived likelihood that management could be convinced to adjust their estimate (i.e., *Convincing to Management*). Consistent with H2, *Nature of Inputs* is depicted as having a moderating effect on the indirect effect on *Source* to *Proposed* through *Source Expertise Concern* and *Convincing to Management*. All continuous variables are mean-centered to facilitate interpretation of the coefficients. [a]We use confidence intervals from bootstrapped sampling distributions (based on 10,000 bootstrap samples) to test the significance of indirect effects (Hayes [2018]). Reflecting our directional predictions, 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 and 80% confidence intervals (i.e., bounded at 0.10 and 0.90) to test whether one-tailed p-values are less than 0.10. * denotes statistical significance equivalent to $p < 0.10$, one-tailed. ** denotes statistical significance equivalent to $p < 0.05$, one-tailed.

inputs condition (90% confidence interval of -0.67 to 0.24). Consistent with H2, which predicts that the effect of *Source* on *Proposed* (and thus the associated indirect effects) is moderated by *Nature of Inputs*, we find that the index of moderated mediation is marginally significant (i.e., 80% confidence interval of -1.11 to -0.03, indicating $p < 0.10$, one-tailed). Collectively, the individual links in this model indicate that receiving contradictory evidence from an AI system instead of a human specialist results in greater expertise concerns, which leads to lower perceived likelihoods that management can be convinced to adjust their estimate. However, these concerns lead to smaller proposed audit adjustments only when management utilizes objective inputs to develop their estimates. Overall, the pattern of these conditional indirect effects is consistent with our theory

that algorithm aversion effects are stronger when management uses objective (vs. subjective) inputs.[16]

4.4 MANIPULATION CHECK FAILURES

In our main analyses, we exclude responses from 29 participants who failed the manipulation checks. We exclude these observations because our manipulations are overt and repeated in multiple places throughout the experimental materials, and we view the manipulation check questions as straightforward tests of whether participants attended to the study.[17] We perform additional analyses to validate this rationale and provide further insight regarding these manipulation check failures. As discussed below, we find troubling and paradoxical patterns in these 29 responses that suggest a critical lack of attention to the experimental materials.

We first examine the internal consistency of participants' answers to two separate sets of related questions in our postexperimental questionnaire (all results untabulated). We ask participants to rate the degree of professional judgment exhibited by management (*Judgment*; 1 = "None at All" and 7 = "A Great Deal") and the perceived bias within management's estimate (*Bias*; 1 = "Not at All Biased" and 7 = "Very Biased). If participants attended to the experimental materials, then we should observe negative correlations between participants' *Subjectivity/Objectivity* ratings (as described in subsection 4.1, 1 = "Very Subjective" and 7 = "Very Objective") and both their *Judgment* and *Bias* ratings. As expected, these correlations are negative and significant in our main sample of 170. However, both correlations are *insignificant* among the 29 participants who failed a manipulation check. Moreover, the correlation between *Subjective/Objective* and *Bias* is inexplicably directionally positive in these 29 responses. We also ask participants to rate their levels of concern about the specialist's expertise (1 = "Not at All Concerned" and 7 = "Very Concerned") and the likelihood that management would find the firm's evidence convincing (1 = "Highly Unlikely" and 7 = "Highly Likely"). Although we find a negative and significant correlation between these measures in our main sample, this correlation is (paradoxically) directionally positive, though insignificant, among the 29 participants who failed a manipulation check. Collectively, we interpret these results as evidence that these 29 participants were not attentive

---

[16] Although our moderated mediation analyses provide additional support for our theory-based expectations, we acknowledge that inferences from these analyses rely on process variables that are measured obtrusively and after participants respond to the main dependent variable. Capturing process measures in this manner creates greater potential for carryover effects and reverse causality threats, which limits the validity of our process inferences (Asay, Guggenmos, Kadous, Koonce, and Libby [2021]).

[17] Theory on reading comprehension, memory, and recall holds that the probability of attentive individuals failing to recall the same information encountered over successive cognitive processing cycles (as in the different sections of our instrument) decreases *exponentially* as the information is repeated (Dijk and Kintsch [1978, p. 369]). Moreover, including observations from inattentive participants tends to obscure results (Maniaci and Rogge [2014]).

to our experimental materials—and therefore should not be included in our analyses.[18]

We further examine whether including these 29 observations in our analyses changes our results and inferences. An ANOVA model (untabulated) with the full sample of 199 (i.e., including all 29 manipulation check failures) continues to yield two significant main effects (both $p < 0.01$, one-tailed), but the interaction becomes insignificant ($p = 0.18$, one-tailed). The ANOVA interaction effect likely weakens because participants who are not attentive to the experimental materials (e.g., those that fail manipulation checks) create noise in the data and reduce statistical power (Maniaci and Rogge [2014]). Additionally, it is important to note that traditional ANOVA models test for disordinal interactions and, in doing so, will attribute most of the explained variance in any ordinal interaction to the main effects (Buckless and Ravenscroft [1990], Guggenmos, Piercey, and Agoglia [2018]), making it difficult for our ANOVA model to detect the type of ordinal interaction suggested by our hypotheses (see Figure 1).

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. Leaders of multinational accounting firms and the audit profession assert that these investments will enhance audit quality (FEI [2017], EY [2018], Bloomberg Tax [2020]). One area that audit firms have targeted for technological innovation is the use of AI systems to provide "expert" evaluations and recommendations to assist auditors in the performance of complex tasks. The profession is likely keen on implementing these 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. [2013], Griffith, Hammersley, and Kadous [2015]). Although 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 AI systems or how AI might influence the way auditors evaluate evidence. Prior research in psychology

---

[18] The number of participants failing a manipulation check in the *AI System/ Objective Inputs* condition (12) is larger than in the other conditions (between 3 and 9). We therefore collect additional data to examine whether this condition (i.e., Cell D) stands out in terms of time taken, difficulty, or understandability. Nineteen additional auditors read the same case materials we used in the main experiment but did so via Qualtrics, which allows us to measure time taken to process the case materials. Additionally, participants assess case understandability and case difficulty instead of dependent variable questions and post-experiment questions used in the main experiment. Although the small sample size precludes meaningful statistical tests, descriptive results reveal that participants in Cell D spend essentially the same amount of time as those in the other conditions and that the case is similarly understandable and similarly difficult in all conditions (data untabulated).

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 influences the way auditors weigh contradictory evidence from their firm's specialists. Specifically, our findings suggest that auditors discount their own firm's contradictory evidence more heavily when it comes from an AI system rather than a human specialist. We further find that this effect is amplified when management's competing evidence appears more objective (i.e., relies on inputs that are more objective vs. subjective). This finding is consistent with the Birnbaum and Stegner [1979] scale-adjustment model and suggests that algorithm aversion has a greater influence on auditors' judgments when management's evidence appears more credible.

Our study makes several important contributions to research and practice. 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) contradictory evidence provided by specialists if that evidence does not conform to their own perspective, which threatens audit and financial reporting quality (Barr-Pulliam et al. [2020], Griffith [2020]). We provide evidence that the adoption of these AI systems can lead to further discounting of firm-provided contradictory evidence, particularly when management develops their estimates using inputs that appear relatively more objective in nature. It is important to note that managers have significant discretion over their estimation methods and inputs, as well as other attributes of evidence (e.g., quantification of evidence; Joe, Vandervelde, and Wu [2017]). Therefore, our findings also raise the troubling possibility that algorithm aversion among auditors can increase the likelihood of auditors accepting their client's potentially biased estimates. Thus, we highlight a possible unintended consequence of employing cognitive technologies in the audit setting.

Our study also contributes to emerging research in psychology and management science around algorithm aversion. Prior studies documenting algorithm aversion effects do so primarily in more objective, less complex, and low-stakes task settings. We document that algorithm aversion manifests in the behavior of experienced professionals with strong incentives to rely on the related computer-generated evidence. Emerging research shows that AI models can outperform human-based models in forecasting, lending, financial reporting, and auditing settings (e.g., Ding et al. [2020], Liu [2021], Chen et al. [2021], Christ et al. [2021]). Our findings add a

cautionary note that the benefits of AI models will only be realized if professionals are willing to rely on the advice that comes from these models. Additionally, unlike other algorithm aversion studies, we examine this phenomenon in a setting where individuals must weigh competing evidence from multiple sources. Our findings demonstrate that in such settings, algorithm aversion effects are amplified when the competing information sources appear more credible. Furthermore, although prior research presents evidence of algorithm aversion, this literature provides little explanation for why and how this phenomenon affects judgments. Our study provides evidence that algorithm aversion partly operates through source credibility concerns.

Our study is subject to some limitations that provide several interesting opportunities for future research. First, to achieve experimental control, the human specialist and the AI system in our study use the same information and provide participants with identical audit evidence (i.e., the information content is held constant in all conditions). Although this design choice is aligned with firms' current focus on developing "narrow" AI systems to replicate human judgments, our findings may not generalize to more ambitious AI systems that firms might use in the future. For instance, firms might begin to employ systems that incorporate a greater degree of autonomy and adaptability, though they might be hesitant to do so without more authoritative guidance from standard setters and regulators (Christ et al. [2021]). Nevertheless, future research could examine how auditor reliance is affected by various system characteristics and capabilities. As these systems become more capable, auditors might become more willing to rely on them. For example, making individuals explicitly aware that an algorithm performs better than a human can increase reliance on algorithms (Castelo, Bos, and Lehmann [2019]). However, despite the algorithm's superior performance, findings suggest this awareness does not create a significant preference for algorithms over humans. Thus, even as these systems advance, some degree of algorithm aversion might persist if system performance is not vastly superior to its human counterparts or if the system capabilities are not well-understood by the auditors using these systems.

Second, this study measures auditors' reactions to a novel source of audit evidence. It is possible that this novelty will decrease after repeated exposure to these AI systems, and as a result the effects we observe might subside over time. Specifically, auditors might become more willing to rely on AI systems as they gain experience and familiarity with them. Previous research, however, 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, Simmons, and Massey [2015]). With repeated exposure and usage, auditors are likely to encounter both positive and negative experiences (e.g., errors) as they use AI systems. Future research could examine how the valence of these experiences shapes the way auditors rely on their firm's AI systems. 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. Future research could explore theory-grounded interventions that mitigate the effects of algorithm aversion and, ultimately, help the auditing profession to realize the full benefits of its investments in cognitive technologies.

APPENDIX

Comparison of Source Treatments

	Human Specialist Condition	AI 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."	

The purpose of this appendix is to illustrate how the language in the experimental instrument equalizes legitimacy across the human specialist and AI system conditions.

REFERENCES

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): 325–52.

ASAY, H. S., R. GUGGENMOS, K. KADOUS, L. KOONCE, and R. LIBBY "Theory Testing and Process Evidence in Accounting Experiments." Working paper, Cornell University, Emory

University, University of Iowa, and University of Texas at Austin, 2021. Available at https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3485844

Austin, A. A.; T. Carpenter; M. H. Christ; and C. Nielson. "The Data Analytics Journey: Interactions Among Auditors, Managers, Regulation, and Technology." Working paper, University of Richmond and University of Georgia, 2020. Available at https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3214140.

Bamber, E. M. "Expert Judgment in the Audit Team: A Source Reliability Approach." *Journal of Accounting Research* 21 (1983): 396–412.

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): 285–99.

Barr-Pulliam, D.; J. R. Joe; S. Mason; and K. A. Sanderson. "The Auditor-Valuation Specialist Coopetitive Alliance in the Fair Value Audit of Complex Financial Instruments." Working paper, University of Louisville, University of Delaware, DePaul University, and Bentley University, 2020. Available at https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3620440

Benston, G. J. "Fair-Value Accounting: A Cautionary Tale from Enron." *Journal of Accounting and Public Policy* 25 (2006): 465–84.

Birnbaum, M. H. "Intuitive Numerical Prediction." *The American Journal of Psychology* 89 (1976): 417–29.

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): 330–36.

Bloomberg Tax. "Big Four Invest Billions in Tech, Reshaping Their Identities." 2020. Available at https://news.bloombergtax.com/financial-accounting/big-four-invest-billions-in-tech-reshaping-their-identities.

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.

Buckless, F. A., and S. P. Ravenscroft. "Contrast coding: A Refinement of ANOVA in Behavioral Analysis." *The Accounting Review* (1990): 933–45.

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*, 2015a. Available at http://www.cpab-ccrc.ca/Documents/Topics/Public%20Reports/CPAB_2014_PR_Highlights_EN_FNL.pdf

Canadian Public Accountability Board (CPAB). *CPAB Report on 2015 Inspections of Canada's Big Four Accounting Firms*, 2015b. Available at http://www.cpab-ccrc.ca/Documents/Annual%20Reports/CPAB_2015_Big_Four_Report_Highlights_EN_Final.pdf.

Cannon, N. H., and J. C. Bedard. "Auditing Challenging Fair Value measurements: Evidence from the Field." *The Accounting Review* 92 (2017): 81–114.

Castelo, N.; M. W. Bos; and D. R. Lehmann. "Task-Dependent Algorithm Aversion." *Journal of Marketing Research* 56 (2019): 809–25.

Chen, X.; Y. H. Cho; Y. Dou; and B. Lev. "Fundamental Analysis of Detailed Financial Data: A Machine Learning Approach." Working paper, New York University, 2021. Available at https://research.chicagobooth.edu/arc/journal-of-accounting-research/jar-annual-conference/papers

Choudhary, P. "Evidence on Differences Between Recognition and Disclosure: A Comparison of Inputs to Estimate Fair Values of Employee Stock Options." *Journal of Accounting and Economics* 51 (2011): 77–94.

Christ, M. H.; S. A. Emett; S. L. Summers; and D. A. Wood. "Prepare for Takeoff: Improving Asset Measurement and Audit Quality with Drone-Enabled Inventory Audit Procedures." *Review of Accounting Studies, forthcoming* (2021): 1–21.

COOPER, L. A.; D. K. HOLDERNESS, JR.; T. L. SORENSEN; and D. A. WOOD. "Robotic Process Automation in Public Accounting." *Accounting Horizons* 33 (2019): 15–35.

DECHOW, P. M.; L. A. MYERS; and C. SHAKESPEARE. "Fair Value Accounting and Gains from Asset Securitizations: A Convenient Earnings Management Tool with Compensation Side-Benefits." *Journal of Accounting and Economics* 49 (2010): 2–25.

DELOITTE, L. L. C. "Demystifying Artificial Intelligence." 2014. Available at https://www2.deloitte.com/us/en/insights/focus/cognitive-technologies/what-is-cognitive-technology.html

DELOITTE, L. L. C. "Minds and Machines: The Art of Forecasting in the Age of Artificial Intelligence." 2016. Available at https://www2.deloitte.com/us/en/insights/deloitte-review/issue-19/art-of-forecasting-human-in-the-loop-machine-learning.html

DELOITTE, L. L. C. "Artificial Intelligence." 2018a. Available at https://www2.deloitte.com/content/dam/Deloitte/nl/Documents/deloitte-analytics/deloitte-nl-data-analytics-artificial-intelligence-whitepaper-eng.pdf

DELOITTE, L. L. C. "Solve for Now. Build for Next. The Deloitte Audit." 2018b. 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. ZEHMS. "The Value Relevance of Managers' and Auditors' Disclosures About Material Measurement Uncertainty." *The Accounting Review* 94 (2019): 215–43.

DIETVORST, B. J., and S. BHARTI. "People Reject Algorithms in Uncertain Decision Domains Because They Have Diminishing Sensitivity to Forecasting Error." *Psychological Science* 31 (2020): 1302–14.

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): 114–26.

DING, K.; B. LEV; X. PENG; T. SUN; and M. A. VARSARHELYI. "Machine Learning Improves Accounting Estimates: Evidence from Insurance Payments." *Review of Accounting Studies* 25 (2020): 1098–134.

DIJK, T. A. V., and W. KINTSCH. "Cognitive Psychology and Discourse: Recalling and Summarizing Stories." *Current Trends in Text Linguistics* (1978): 61–80.

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): 458–68.

EMETT, S. A.; S. E. KAPLAN; E. MAULDIN; and J. S. PICKERD. "Auditing with Data and Analytics: External Reviewers' Judgments of Audit Quality and Effort." Working paper, Arizona State University, University of Mississippi, and University of Missouri, 2021. Available at https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3544973

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 ACCOUNTING STANDARDS BOARD (FASB). *Fair Value Measurements.* Statement of Accounting Standards Codification (ASC) Section 820. Norwalk, CT: FASB, 2010.

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

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): 232–55.

GRIFFIN, J. B. "The Effects of Uncertainty and Disclosure on Auditors' Fair Value Materiality Decisions." *Journal of Accounting Research* 52 (2014): 1165–93.

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* 37 (2020): 245–76

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 (2015): 833–63.

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

GUGGENMOS, R. D.; M. D. PIERCEY; and C. P. AGOGLIA. "Custom Contrast Testing: Current Trends and a New Approach." *The Accounting Review* 93 (2018): 223–44.

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

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

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

HILTON, A. S., and P. C. O'BRIEN. "Inco Ltd.: Market Value, Fair Value, and Management Discretion." *Journal of Accounting Research* 47 (2009): 179–211.

HIRST, D. E. "Auditors' Sensitivity to Source Reliability." *Journal of Accounting Research* 32 (1994): 113–26.

HOVLAND, C. I., I. L. JANIS, and H. H. KELLEY. "A Summary of Experimental Studies of Opinion Change," in Marie Jahoda & Neil Waren (eds.), *Attitudes.* Baltimore, Maryland: Penguin Books, (1966): 139–51.

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.

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): 209–26.

KACHELMEIER, S. J., and M. G. WILLIAMSON. "Attracting Creativity: The Initial and Aggregate Effects of Contract Selection on Creativity-Weighted Productivity." *The Accounting Review* 85 (2010): 1669–91.

KADOUS, K. L.; L. KOONCE; and K. L. TOWRY. "Quantification and Persuasion in Managerial Judgment." *Contemporary Accounting Research* 22 (2005): 643–86.

KPMG, L. L. P. "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): 47, 278–92.

LI, K. K., and R. G. SLOAN. "Has Goodwill Accounting Gone Bad?" *Review of Accounting Studies* 22 (2017): 964–1003.

LIU, M. "Assessing Human Information Processing in Lending Decisions: A Machine Learning Approach." Working paper, Boston College, 2021. Available at https://research.chicagobooth.edu/arc/journal-of-accounting-research/jar-annual-conference/papers

MANIACI, M. R., and R. D. ROGGE. "Caring About Carelessness: Participant Inattention and Its Effects on Research." *Journal of Research in Personality* 48 (2014): 61–83.

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. (207–28). Cambridge: Cambridge University Press, 1995.

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

NELSON, M. W.; J. A. ELLIOTT; and R. L. TARPLEY. "Evidence from Auditors about Managers' and Auditors' Earnings Management Decisions." *The Accounting Review* 77 (2002): 175–202.

NELSON, M. W.; S. D. SMITH; A. Z. V. PALMROSE. "The Effect of Quantitative Materiality Approach on Auditors' Adjustment Decisions." *The Accounting Review* 80 (2005): 897–920.

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.

O'REILLY, C. A., and K. H. ROBERTS. "Relationships Among Components of Credibility and Communication Behaviors in Work Units." *Journal of Applied Psychology* 61 (1976): 99–102.

PEECHER, M. E., and I. SOLOMON. "Theory and Experimentation in Studies of Audit Judgments and Decisions: Avoiding Common Research Traps." *International Journal of Auditing* 5 (2001): 193–203.

PIERCEY, M. D. "Documentation Requirements and Quantified Versus Qualitative Audit Risk Assessments." *Auditing: A Journal of Practice & Theory* 30 (2011): 223–48.

PORNPITAKPAN, C. "The Persuasiveness of Source Credibility: A Critical Review of Five Decades' Evidence." *Journal of Applied Social Psychology* 34 (2004): 243–81.

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): 455–68.

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

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. Washington, DC: PCAOB, 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. Washington, DC: PCAOB, 2017b.

PUBLIC COMPANY ACCOUNTING OVERSIGHT BOARD (PCAOB). "Changes in Use of Data and Technology in the Conduct of Audits." PCAOB Research and Standard Setting Projects. Washington, DC: PCAOB, 2020. Available at https://pcaobus.org/oversight/standards/research-standard-setting-projects/changes-use-data-technology-conduct-audits

RAMANNA, K., and R. L. WATTS. "Evidence on the Use of Unverifiable Estimates in Required Goodwill Impairment." *Review of Accounting Studies* 17 (2012): 749–80.

RAPHAEL, J. "Rethinking the Audit." *Journal of Accountancy* 223 (2017): 29–32. Available at https://www.journalofaccountancy.com/issues/2017/apr/rethinking-the-audit.html.

SAIEWITZ, A., and T. KIDA. "The Effects of an Auditor's Communication Mode and Professional Tone on Client Responses to Audit Inquiries." *Accounting, Organizations and Society* 65 (2018): 33–43.

SWIERINGA, R. J., and K. E. WEICK. "An Assessment of Laboratory Experiments in Accounting." *Journal of Accounting Research* 20 (1982):56–101.

TYSIAC, K. "How to Enable Audit Innovation." Journal of Accountancy 233 (2017): 33–35. 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*. Washington, DC: Brookings Institution Press, 2018. Available at https://www.brookings.edu/research/how-artificial-intelligence-is-transforming-the-world/.

YEOMANS, M.; A. K. SHAH; S. MULLAINATHAN; and J. KLEINBERG. "Making Sense of Recommendations." *Journal of Behavioral Decision Making* 32 (2019): 403–14.

ZIMBELMAN, M. F., and W. S. WALLER. "An Experimental Investigation of Auditor-Auditee Interaction Under Ambiguity." *Journal of Accounting Research* 37 (1999): 135–55.