EVIDENCE SPECIFICITY AND AUDITORS' HEURISTIC RESPONSES TO HUMAN AND NON-HUMAN SPECIALISTS

Benjamin P. Commerford
University of Kentucky
ben.commerford@uky.edu

Blake A. Holman
University of Kentucky
blakeaholman@uky.edu

April 2022

We would like to thank Nargess Golshan, Jiv Johnson, Jenny Ulla, and workshop participants from the University of Kentucky and the University of Denver for their guidance and valuable suggestions. We also thank the Center for Audit Quality for its financial assistance, helpful feedback, and for providing participants for this study through their Research Advisory Board grant. While funding for the research described in this article was provided by the Center for Audit Quality, the views expressed in this article and its content are those of the authors alone and not those of the Center for Audit Quality. Lastly, we gratefully acknowledge funding from the Von Allmen School of Accountancy at the University of Kentucky.

EVIDENCE SPECIFICITY AND AUDITORS' HEURISTIC RESPONSES TO HUMAN AND NON-HUMAN SPECIALISTS

ABSTRACT

PCAOB inspection reports document persistent deficiencies in auditors' use of specialists when evaluating complex estimates. In this study, we examine whether the degree of specificity within specialist explanations influences auditors' elaboration, and ultimately, reliance on specialist-provided evidence. In addition, we examine whether the type of engaged specialist (i.e., human versus system) impacts the extent of auditors' reliance on that specialist. We find that a less specific explanation constrains auditors' elaboration, resulting in judgments indicative of algorithm aversion. Interestingly, when given a greater opportunity to elaborate (via a more specific explanation), auditors increase their reliance on system-provided evidence and *reduce* their reliance on evidence from human specialists. Our findings demonstrate that when auditors' elaboration is relatively low, their decisions are heavily influenced by heuristic cues (e.g., human versus non-human source). Yet, when elaboration is relatively high, auditors' decisions are based on the underlying evidence rather than superficial factors.

I. INTRODUCTION

The inclusion of complex estimates within the financial statements has become increasingly prevalent (Boritz, Kochetova-Kozloski, Robinson, and Wong [2020]) and due to the high levels of subjectivity, inherent ambiguity, and uncertainty contained in these estimates, auditors often have trouble providing assurance on such estimates (e.g., Martin, Rich, and Wilks [2006]; Griffith, Hammersley and Kadous [2015]; Doty [2017]). To address these difficulties, auditors frequently engage valuation specialists to help with the valuation assertion of complex estimates (Cannon and Bedard [2017]; PCAOB [2018a]; Griffith [2018], [2020]). Despite auditors' use of valuation specialists, recent research and PCAOB inspection reports document persistent challenges and deficiencies in auditors' testing of complex estimates (e.g., PCAOB [2016], [2017]; Joe, Vandervelde, and Wu [2017]) and their use of specialists (Griffith [2018], [2020]). For example, at times, auditors' often fail to adequately test the data and assumptions underlying valuation specialists' estimates, leading to over-reliance on advice from specialists (Glover, Taylor, and Wu [2017]). At the same time, some research contends that auditors often ignore specialists' work (Griffith [2020]). Collectively, these contrary findings suggest that auditors do not always appropriately elaborate and rely on specialist-provided evidence.

In this study, we examine whether the degree of specificity within specialist explanations influences auditors' elaboration, and thus, reliance on specialist-provided evidence. The Elaboration Likelihood Model (ELM) posits that variations in judgments and decisions depend on a person's level of elaboration (i.e., effortful cognitive processing), which falls along a continuum ranging from low to high (Petty and Briñol [2014]). Under low elaboration, individuals use simple decision rules or cognitive heuristics to formulate their judgments (Dijkstra [2001]). Therefore, the ELM suggests that when auditors are provided with a specialist report that contains only a brief outline of the work performed (i.e., a less-specific explanation), their capacity to elaborate is constrained, leading to greater use of simple cues when determining how to rely on the available evidence. In contrast, under high elaboration, judgments and decisions are more likely to result from a person's careful examination of the information contained in the message. Accordingly, when auditors are provided with more specific explanations, the opportunity to elaborate is no longer inhibited, resulting in auditor's basing their judgments on the true merits of the evidence being presented. Overall, the ELM suggests that auditors will be more (less) susceptible to heuristics and biases when elaboration is low (high).

We also examine this research question in light of the rapid advancement of audit technologies. Audit firms are making substantial investments in advanced technologies to help auditors perform challenging tasks, such as evaluating management's complex accounting estimates (Austin, Carpenter, Christ, and Nielson [2020]; Bloomberg Tax [2020]). In some cases, the goal is for these systems to perform tasks that have traditionally been performed by human specialists (KPMG [2016]; Murphy [2017]). For example, KPMG is developing a system that employs artificial intelligence (AI) to evaluate commercial loan grades (KPMG [2016]). However, early evidence suggests auditors more heavily discount the same contradictory audit evidence when they believe it is provided by an AI-based system instead of a human specialist (Commerford, Dennis, Joe, and Ulla [2022]). Thus, the human or non-human nature of an evidence source is likely a heuristic cue that auditors will rely on, particularly when their elaboration is low. Accordingly, we predict that low elaboration, arising from less-specific specialist explanations, will lead to greater reliance on evidence from human specialists than specialist systems, consistent with algorithm aversion. In contrast, if elaboration leads auditors to base their judgments on the true merits of the evidence being presented, then this should reduce the influence of peripheral cues, mitigating the effects of algorithm aversion.

Furthermore, if higher elaboration leads to more thoughtful consideration of the available evidence, then such consideration could increase *or decrease* auditors' reliance on the work of specialists. On one hand, greater elaboration could increase auditors' understanding of and comfort with that work, which may promote increased reliance. On the other hand, as auditors receive more specifics around a specialist's methodology and assumptions, they may become more critical of the specialist's work and consequently more hesitant to rely on it. We propose that the directional effect of increased elaboration on auditors' reliance on specialist-provided evidence will depend on the human nature of the specialist (i.e., human versus non-human).

Research suggests individuals heuristically discount recommendations from algorithms and systems, but finds individuals are less averse to system-based recommendations when they have a greater understanding of how a system formulates its recommendations (Yeomans, Shah, Mullainathan, and Kleinberg [2019]). As such, we expect when auditors receive evidence from a novel and unfamiliar evidence source – like an AI-based specialist system – greater specificity, and thus greater elaboration, will increase auditors' reliance on the system-provided evidence. In contrast, research finds individuals tend to heuristically assume advice from human experts is more valuable than advice from non-experts, even before evaluating the evidence (Meshi, Biele, Korn, and Heekeren [2012]). Greater elaboration should mitigate this heuristic reliance on expert advice and encourage a more thoughtful consideration of the evidence itself. Therefore, we expect greater specificity around the specialists' assumptions will *decrease* auditor reliance on evidence from human specialists.

To test our predictions, we conduct a 2 x 2 between-subjects experiment with 98 audit managers, manipulating the source of audit evidence (human specialist versus specialist system) and the degree of specificity contained within the specialist's explanation around their methodology and assumptions (more versus less specific). To manipulate the source of audit evidence, participants receive audit evidence from either a firm-employed human valuation specialist or from a firm's proprietary AI system (i.e., specialist system). In the less-specific condition, the valuation specialists' memo contains a brief outline of assumptions and methodology used to evaluate management's estimate. In the more-specific condition, the valuation specialists' memo provides a more thorough and precise discussion around the testing methodology and assumptions incorporated by the specialist.

We use a client's Insurance Claim Reserve estimate as our experimental setting. Participants in all conditions receive the same audit evidence regarding a client's Insurance Claims Reserve, which suggests that the client's current Insurance Claims Reserve balance is understated by $28 million, an amount greater than the materiality threshold on the engagement. Case materials indicate that the potential audit difference arises from a disagreement between management and the audit firm's specialist system (or human valuation specialist) over estimated loss projections. Participants evaluate the available evidence and then make a recommendation for a proposed audit adjustment. The magnitude of the participants' proposed audit adjustments serves as our main dependent variable, with larger proposed adjustments indicating higher reliance on evidence provided by the specialist (Commerford et al. [2022]).

Consistent with theory, we find that auditors differentially rely on their own firm's contradictory evidence, depending on whether it comes from a human valuation specialist or specialist system. Specifically, when the opportunity for elaboration is constrained (i.e., when provided with less specific explanations), auditors propose lower audit adjustments when the audit firm's contradictory evidence comes from a system specialist compared to that of a human valuation specialist, indicative of algorithm aversion. However, when given a greater opportunity to elaborate (via the provision of a more specific explanation), auditors increase their reliance on contradictory evidence from a specialist system, as evidenced by higher proposed audit adjustments. Interestingly, when provided with a more (versus less) specific explanation, auditors reduce their reliance on evidence from a human valuation specialist to a level consistent with that of a specialist system.

The findings of this study are important for several reasons. First, this study has implications for auditors' evaluations of complex estimates and reliance on the work of specialists, which continues to be an area of interest for both firms and regulators. Audit standards require auditors to carefully review the assumptions and findings of the specialist's work and consider the implications that work may have regarding the reasonableness of management's estimates (PCAOB 2018b). However, at times, auditors appear to disregard contradictory evidence identified by specialists (PCAOB [2017]), and at other times, they seem to simply scan specialists' memos for conclusions (Griffith [2018]). This lack of elaboration is concerning, as our findings demonstrate that when elaboration is low, this can make auditors' judgments more susceptible to heuristics and biases. Specifically, we find that when elaboration is low, auditors rely on the heuristic cue of human versus non-human specialist to determine their extent of reliance on the specialist-provided evidence. Thus, low elaboration can increase the potential for auditors' judgments to be influenced by superficial factors that are not necessarily indicative of audit risk.

We also find that the degree of specificity contained within a specialists' report can influence an auditors' ability to elaborate on the specialists' work. This finding is important as it demonstrates how not only auditor characteristics (e.g. motivation), but also characteristics of specialist-provided evidence, impact auditor's cognitive processing and ultimately their judgments. As such, it is possible that some of the deficiencies identified by the PCAOB and other internal regulators regarding auditors' evaluations of complex accounting estimates could be driven by the nature of the evidence specialists provide to auditors.

In addition, our paper contributes to the growing algorithm aversion literature. Research has theorized that algorithm aversion can manifest from individuals' lack of understanding around how a system formulates its recommendations (Yeomans et al. [2019]). Our study provides evidence consistent with this notion, as participants increase their reliance on system-provided evidence when provided with more specific and detailed explanations. Furthermore, we answer the call by Commerford et al. [2022] to identify factors that can promote greater reliance on evidence from firms' AI-based specialist systems. This study demonstrates that the provision of greater specificity around a specialist system's testing procedures and methodologies helps to mitigate an individual's natural aversion to reliance on advice from non-human sources. Our findings, therefore, have implications for how firms design and implement their systems.

II. THEORY AND HYPOTHESIS DEVELOPMENT

Auditor Elaboration and Evidence Specificity

The inclusion of complex estimates within the financial statements has become increasingly prevalent (Bratten, Gaynor, McDaniel, Montague, and Sierra [2013]; Brown, Grenier, Pyzoha and Reffett [2019]; Boritz et al. [2020]), but due to their high level of subjectivity and inclusion of forward-looking assumptions, they are also very difficult to audit (PCAOB [2009]; Griffith et al. [2015]; Glover et al. [2017]). To assist in this process, auditors frequently engage valuation specialists (Cannon and Bedard [2017]; PCAOB [2018a]; Griffith [2018], [2020]).

Although auditors are expected to thoughtfully consider the evidence produced by these valuation specialists, deficiencies in auditors' use of specialists are commonly documented. For example, prior research finds that auditors often fail to adequately test the underlying data and assumptions used by specialists (Glover et al. [2017]), tend to ignore the details in specialists' work (Griffith [2018]), are prone to accept specialists' conclusions at face value (Griffith et al. [2015]; Glover, Taylor, Wu, and Trotman [2019]), and treat contrary evidence in specialists' work as irrelevant (Griffith et al. [2015]; Griffith [2018]). Collectively, these findings suggest that, at times, auditors fail to sufficiently elaborate on the evidence provided to them from specialists.

The Elaboration Likelihood Model (ELM) posits that individual's judgments and decisions are significantly influenced by their level of elaboration (i.e., effortful cognitive processing). An individual's level of elaboration is determined by both their motivation and capacity to process information (Petty and Briñol [2014]). Motivation to elaborate refers to a person's willingness or desire to engage in effortful thinking in order to assess the validity of a message (Griffith, Nolder, and Petty [2018]), while capacity to elaborate refers to a person's ability and/or opportunity to think (Petty, Wheeler, and Tormala [2003]). When individuals have both high motivation and high capacity they will engage in a "central route" of processing, where attitude changes are based on a thoughtful consideration of issue-relevant information and an integration of that information into an overall position (Petty and Cacioppo [1986]). However, when motivation and/or capacity are relatively low, individuals will follow a "peripheral route" (Petty, Cacioppo, and Schumann [1983]). Under the peripheral route, individuals rely on simple decision rules or cognitive heuristics to formulate their judgments (Petty and Cacioppo [1984], [1986]; Petty, Wegener, Fabriger [1997]).

Elaboration is critical to the quality of auditor's judgments and decisions. For example, Kadous and Zhou [2019] find that making auditors' intrinsic motivation more salient causes auditors to consider a broader set of information and process that information more deeply (i.e., higher elaboration), leading to higher quality assessments of complex estimates. Conversely, audit research also demonstrates that low elaboration can lead to low-quality audit judgments (e.g., Griffith et al. [2018]). For example, Bhattacharjee and Brown [2018] provide evidence that alumni affiliations between auditors and clients can lead to lower elaboration by auditors, and thus, more lenient internal control evaluations. Likewise, other studies find that lower elaboration by auditors can lead to less critical review of subordinates' work (Rich [2004]) and a failure to incorporate relational cues when evaluating specialists' work (Griffith [2018]). Therefore, it is important to understand both the determinants and consequences of auditors' elaboration.

Input variables affecting auditors' elaboration can be organized into four categories: auditor, client, evidence, and environmental characteristics (Griffith et al. [2018]). Prior research has typically focused on how auditor and client-related factors (e.g., motivation, alumni affiliation) or environmental features (e.g., audit risk) affect auditors' elaboration, and therefore their judgments (Bhattacharjee and Brown [2018]; Griffith [2018]; Kadous and Zhou [2019]). In this study, we examine how a characteristic of specialist-provided evidence – the degree of specificity – affects auditors' reliance on that evidence. Valuation specialists often evaluate management's estimation model, inputs, and assumptions (e.g., discount rates, market benchmarks, and general industry or economic trends), all of which can be very subjective (Glover et al. [2017]; Griffith [2020]). These specialists are required to document the work performed, results obtained, and conclusions reached (PCAOB [2018a]). That said, the amount and type of information provided to the auditor about the specialist's methodology and assumptions likely varies in practice, with some specialist reports providing greater specificity than others. The ELM suggests these differences in specialist-provided evidence will likely influence auditors' degree of elaboration. For instance, when auditors are provided with a less-specific explanation from a specialist, this will constrain auditors' capacity to elaborate, whereas a more-specific explanation allows for greater elaboration. As a result, the ELM suggests that lower evidence specificity will cause auditors to rely more heavily on simple cues (i.e., peripheral route) when determining how to incorporate audit evidence into their judgments.

Human versus Non-Human Specialists

Emerging research suggests auditors will use the human (versus non-human) nature of an evidence source as a heuristic cue of evidence quality. Several studies find evidence of algorithm aversion, such that individuals rely more heavily on advice produced by a human source than advice produced by a computer-based source (Önkal, Goodwin, Thomson, Gonul, and Pollock [2009]; Dietvorst, Simmons, and Massey [2015]; Yeomans et al. [2019]; Commerford et al. [2022]). For example, Önkal et al. [2009] show that individuals more heavily weight stock forecasting advice from a human instead of a computer-based model, despite the advice being otherwise identical. Castelo, Bos, and Lehmann [2019] find that consumers are averse to relying on algorithms to perform subjective tasks that are typically done by humans, even though algorithms often perform better than humans. Specific to the audit setting, Commerford et al. [2022] demonstrate that auditors are also susceptible to algorithm aversion, as evidenced by smaller proposed adjustments to management's estimates when receiving contradictory advice from a firm's artificial intelligence-based specialist system relative to a human specialist.

The ELM suggests less-specific specialist explanations will lead to low elaboration, resulting in greater reliance on evidence from human specialists versus non-human specialists.

Consistent with Commerford et al. [2022] this difference in reliance should be evident in auditors' adjustment decisions. Within our setting, when weighing evidence from both the client and a specialist, less-specific specialist explanations should lead to relatively lower audit adjustments when that specialist is an AI-based system instead of a human (Commerford et al. [2022]). In contrast, the ELM suggests that when presented with more-specific explanations, auditors will follow the central route of processing, which entails a more careful consideration of the information and an integration of that information into an overall position (Petty and Cacioppo [1986]). Consequently, auditor's reliance on advice will be based on the true merits of the information presented without the influence of peripheral cues (e.g., human versus non-human source), neutralizing the effects of algorithm aversion. Therefore, if the specialist-provided evidence is identical otherwise, this should yield similar audit adjustments in response to contradictory evidence from human and non-human sources. Accordingly, we predict the following:

> ***H1*****:** Low elaboration, arising from less-specific specialist explanations, will lead to greater reliance on evidence from human specialists than specialist systems, whereas high elaboration will lead to equal reliance on evidence from human specialists and specialist systems.

Importantly, it is possible for higher elaboration (via increased evidence specificity) to increase or decrease reliance on specialist-provided evidence. On one hand, more detailed, specific explanations around a specialist's work could increase auditors' understanding of that work, which may promote increased reliance. On the other hand, as auditors receive more detail around a specialist's methodology and assumptions, they may become more critical of the specialist's work and consequently more hesitant to rely on it. We propose that the directional effect of evidence specificity on reliance will depend on whether that evidence is produced by a specialist system or a human specialist.

Research demonstrates that individuals often use a peripheral route when choosing to rely on the advice of others, and that they are quite willing to rely on the advice of "expert" sources (Petty and Cacioppo [1984], [1986]; Maheswaran, Mackie, and Chaiken [1992]). For instance, people often accept the advice of a doctor with little explanation because an expert gives the advice, not because they have carefully examined and evaluated the advice as being correct (Dijkstra [2001]). Additionally, just knowing that advice comes from an expert can lead individuals to assume that advice will be more valuable. Meshi et al. [2012] use functional magnetic resonance imaging (fMRI) to demonstrate that reward centers in the brain become more active when individuals know they will receive advice from an expert (versus a novice). Interestingly, reward center activity increases when participants discover they will receive advice from an expert, even *before* they see or evaluate that advice. This suggests individuals use expert status as a heuristic cue of advice quality. Thus, without careful elaboration, auditors will likely default into greater reliance on evidence from a human specialist than is warranted because the specialist is believed to be a subject matter expert. Conversely, this heuristic cue should influence auditor judgments less when auditors' ability elaborate is less constrained (i.e., more-specific specialist explanations).

Furthermore, the inclusion of details around testing procedures can induce a more critical analysis (Fischhoff [1995]). For example, Kadous, Koonce, and Towry [2005] conclude that providing additional information around the details of a proposal invites criticism, thereby making a proposal less persuasive. Relatedly, auditors are less comfortable with uncertain estimates when provided with additional evidential support because it presents a challenge to their beliefs (Rowe [2019]). Therefore, as auditors engage in a more thoughtful consideration of the evidence presented by a human specialist, they likely will become more critical of the inherent uncertainty and imprecision contained within the specialists' work, due to the high degree of subjectivity and uncertainty associated with complex estimates. Accordingly, we expect that as human specialists' explanations become more specific, auditors will become *less* likely to rely on the specialist-provided evidence.

In contrast, we theorize auditors likely have an opposite default response to evidence produced by non-human sources. As noted earlier, research shows that individuals' judgments often exhibit algorithm aversion, such that they are particularly hesitant to rely on advice from computer-based systems and algorithms. Interestingly, Yeomans et al.'s [2019] findings suggest that individuals are particularly averse to system-based recommendations when they feel they do not understand the process the system is utilizing to make its recommendations. However, when individuals are provided with more-detailed explanations regarding how the system formulated its recommendations, individuals became less averse to relying on the system's recommendations. Consistent with this notion, there is growing demand for explainable AI, which seeks to reduce the "black box" nature of AI-based systems by adding interfaces within the systems that increase transparency and understanding around system output (Adadi and Berrada [2018]; Došilović, Brčić, and Hlupić [2018]; Samek and Müller [2019]; Jiménez-Luna, Grisoni, and Schneider [2020]). For example, Herlocker, Konstan, and Riedl [2000] provide experimental evidence that explanation interfaces can improve the acceptance of automated systems. These finding suggest that when provided with more-specific explanations, auditors will have a better opportunity to inform themselves about the system's testing procedures and methodologies. Accordingly, auditors will no longer rely on the simple cue of a human versus non-human source to determine the extent of their reliance. As such, we expect that as evidence specificity increases, auditors will become *more* likely to rely on audit evidence that comes from a specialist system.

H2: High (versus low) elaboration will decrease auditor reliance on evidence from human specialists, but increase reliance on evidence from specialist systems.

In summary, H1 predicts that, when provided with less-specific specialist explanations, auditors' elaboration will be restricted. As a result, auditors will rely more on heuristic cues, making their judgments more likely to exhibit algorithm aversion. In contrast, more-specific specialist explanations should induce higher elaboration, leading to a thoughtful consideration of the true merits of the information as opposed to reliance on peripheral cues, neutralizing the effects of algorithm aversion. Furthermore, H2 predicts that as evidence specificity increases, auditors will be less (more) likely to rely on evidence from human specialists (specialist systems). Taken together, our hypotheses form an ordinal interaction in which auditors' reliance on specialist-provided evidence will be highest (lowest) when receiving a less-specific explanation from a human specialist (specialist system). Additionally, when auditors receive more-specific specialist explanations, there evidence reliance will not differ across specialist-type (human versus system). Our predicted pattern of experimental cell means is illustrated in Figure 1.

[Insert Figure 1 here]

III. METHOD

Participants

Participants were recruited with the assistance of the Center for Audit Quality (CAQ) and through the authors' personal contacts. We provided the CAQ with a recruitment email inviting auditors to participate in the study. This email included a brief description of our research objectives along with a hyperlink to the case materials, which auditors accessed electronically through Qualtrics. CAQ personnel forwarded the email to contacts at each of the participating firms. A similar process was followed when recruiting participants through personal contacts.[1] Table 1 presents participants' demographic data. Our final data set consists of 98 responses from audit managers with an average of 6.94 years of public accounting experience.[2] Participants indicated they are very likely to provide input into decisions related to proposed audit adjustments in a typical year (mean of 5.77 on a scale from 1 = "Not at All Likely" and 7 = "Highly Likely"). Given that both audit seniors and managers are responsible for evaluating assumptions related to complex estimates (Pyzoha, Taylor, and Wu [2020]), these participants are well-matched to the task and the goals of our research (Libby, Bloomfield, and Nelson [2002]).

[Insert Table 1 here]

Experimental Audit Case

Participants assume the role of manager on the financial statement audit of AG Insurance (AGI), a hypothetical property and casualty insurance company. Participants are asked to evaluate one of AGI's most significant estimates – the Insurance Claims Reserve. First, participants receive background information about AGI and a general overview of how AGI management estimates the Insurance Claims Reserve. Participants are then informed that a firm specialist will be assisting the audit team in their evaluation of AGI's Insurance Claims Reserve. We manipulate the firm

[1] All participants were assured that their identity and identity of their firm would be kept confidential. Approval for this study was obtained from representatives for each of the participating firms as well as the institutional review board (IRB) at the university from which the online experiment was administered.

[2] We received 112 complete responses. Before analyzing the data, we developed an estimate for the minimum amount of time necessary to read through the case details. For silent reading of non-fiction, most adults fall in the range of 175-300 words per minute (wpm) and retain comprehension (Andrews [2010]; Brysbaert [2019]). We take a conservative approach to our estimate by using rate of 300 wpm (i.e., the "fastest" reading rate) and by focusing on the condition with the fewest words (*System/Less Specific*). Given that there are 3,290 words in the *System/Less Specific* condition, it should take participants a minimum of approximately 11 minutes to read through the case materials. There were 13 responses in which participants spent less than 11 minutes completing the task. These 13 responses were excluded from our analyses as they did not spend a sufficient amount of time on the experimental task. Additionally, following Field's [2018] approach for identifying outliers through Z scores, we identified one response as an extreme outlier based on the total time taken to complete the study (approximately 10,000 minutes). Participants were instructed to complete the task in one sitting and, therefore, we exclude this extreme outlier from our analyses. Thus, our final sample includes 98 responses with a median completion time of about 23 minutes.

specialist as either a human specialist or an AI-based specialist system (this manipulation is described in further detail below). Next, participants read about the audit procedures that have already been performed by the audit team, which include testing the effective of internal controls, testing the calculations and vouching the underlying data to respective sources, and performing a look-back analysis. As part of the audit procedures, the firm's valuation specialist developed an independent estimate for AGI's Insurance Claims Reserve. The specialist's independent estimate suggests the client's Insurance Claims Reserve is understated by $28 million, which is greater than the materiality threshold on the engagement.

Participants then review audit evidence from both AGI's management and the audit firm's specialist. Case materials indicate that the potential audit difference is primarily attributed to differences in the estimated loss projections associated with automotive insurance policies, which are driven by differences in expectations for future mobility trends. Future mobility estimates are a key input into the Insurance Claim Reserve. In general, as mobility increases, accidents (and related claims) become more frequent. Management expects future mobility to be 16 percent lower than historical levels, due to pandemic-related changes in mobility patterns (e.g., work from home, telehealth, lower business travel) that they believe will become largely permanent. The specialist provides a report indicating an expectation for future mobility to return to 90 percent of normal levels (i.e., 10 percent lower than the historical average). This specialist report contains our *Degree of Specificity* manipulation (described in further below). Overall, case details indicate that both AGI's management and the firm specialist expect future mobility to remain lower than historical averages. However, because management's mobility estimate is lower than the specialist's, management's estimate for the Insurance Claim ($560 million) is significantly lower than the specialist's independent estimate ($588 million), yielding a $28 million potential audit difference.

Participants view a comprehensive summary of the issue (including the source of the disagreements in estimated loss projections and the impact of the potential audit difference on AG Insurance's net income) and then recommend a proposed adjustment to the Insurance Claim Reserve. Proposed adjustments provide a link between auditors' attitudes, feelings, judgments, and anticipated actions (Nelson [2009]; Kadous, Nolder, and Peecher [2018]; Rowe [2019]) and is reflective of the auditor's relative weighting of the available evidence (Commerford et al. [2022]). Given the contradictory nature of the evidence, larger (smaller) proposed adjustments indicate higher (lower) reliance on the evidence provided by the specialist (Commerford et al. [2022]). Finally, participants complete a post-experiment questionnaire, which allows us to collect process measures and demographic information. Figure 2 displays the flow of the experimental design.

[Insert Figure 2 here]

Independent Variables

We employ a 2 x 2 between-subjects experimental design, manipulating the source of audit evidence (i.e., *"Specialist Type"*) and the degree of specificity contained within the specialist memo regarding their testing methodology and assumptions (*Less* versus *More Specific*). Thus, participants are randomly assigned to one of four experimental conditions. To manipulate the first factor, *Specialist Type*, participants receive specialist-provided evidence from either a *System* or *Human* source. In the *System* condition, the audit firm utilizes a proprietary AI-based system that incorporates advanced data analytics and adaptive predictive modeling to develop its independent estimate of AGI's Insurance Claims Reserve. In the *Human* condition, the independent estimate of AGI's Insurance Claims Reserve is developed by an internal group of specialized professionals. Following Commerford et al. [2022], the description of the two sources is written in a manner that the accuracy and legitimacy of the two sources are equivalent, which allows any observed effects to be attributed to the human/non-human nature of the source.[3]

We also manipulate *Degree of Specificity* within the specialist report as either *Less Specific* or *More Specific*. In the *Less Specific* condition, the specialist's report contains a brief outline of assumptions and methodology used to develop the independent estimate. In the *More Specific* condition, the valuation specialist's memo provides an identical discussion around the testing methodology and assumptions incorporated by the specialist in the development of its/their independent estimate, but also includes an additional section detailing the specific factors considered when estimating future mobility rates (see Appendix A).

IV. RESULTS

Manipulation Checks

To evaluate our *Specialist Type* manipulation, we asked participants whether the audit team received input from the audit firm's internal valuation group or the valuation system to assist them in their testing of the Insurance Claims Reserve. A large majority of participants (82 percent) correctly identified *Specialist Type*. To assess whether we successfully manipulated the *Degree of Specificity*, we asked participants to indicate the extent to which the respective specialist provided "specific factors for its/their expectation for rising future mobility levels" (1 = "Not at All Specific" and 7 = "Very Specific"). As expected, the perceived specificity was higher for participants in the *More Specific* condition (m = 5.02) relative to participants in the *Less Specific* condition (m = 3.54,

[3] To ensure observed effects can be attributed to differences in the human/non-human nature of the specialist and not to differences in the perceived credentials (e.g., accuracy and legitimacy) between the two sources, we asked participants whether the audit firm views the respective specialist as an "approved source of audit evidence" (1 = "Strongly Disagree" and 7 = "Strongly Agree"). Mean responses in both conditions are significantly higher than the scale midpoint ($p < 0.01$, two-tailed, untabulated), indicating that participants viewed their respective specialist as an approved, reliable source of audit evidence. Interestingly, the mean response in the *System* condition (6.18) is higher than the mean response in *Human* condition (5.72), though this difference is only marginally significant ($p = 0.09$, two-tailed, untabulated). Given that we expect auditor judgments to exhibit algorithm aversion when elaboration is low, this marginal difference across conditions biases against finding our predicted results.

$t_{94} = 5.79$, $p < 0.01$, one-tailed, untabulated), indicating a successful manipulation of *Degree of Specificity*.[4]

Tests of Hypotheses

Table 2, Panel A presents the descriptive statistics for participants' proposed adjustments by experimental condition, which are depicted visually in Figure 3. In conjunction, our hypotheses predict an ordinal interaction (as depicted in Figure 1). Therefore, we use a single planned contrast to test whether the magnitudes of participants' proposed adjustments are in accordance with the hypothesized ordinal pattern (Buckless and Ravenscroft [1990]; Guggenmos, Piercey, and Agoglia [2018]; Mendoza [2020]). Based on our theoretical predictions, we selected the following contrast weights for each experimental condition: +1 Human/Less Specific, -1 System/Less Specific, 0 Human/More Specific, and 0 System/More Specific (corresponding to the labels A, B, C, D in Figure 1, respectively). Consistent with best practices set forth by Buckless and Ravenscroft [1990] and reinforced by Guggenmos et al. [2018], these weights were selected prior to performing our analyses and reflect our theory-driven hypotheses.

Following Guggenmos et al. [2018], we conduct a three-part approach for testing a predicted pattern of means prior to concluding the extent to which our hypotheses are supported. First, we check the visual fit of the observed data and find the mean pattern depicted in Figure 3 strongly resembles the hypothesized mean pattern (see Figure 1). Second, as reported in Table 2, Panel B, we find that the contrast test is significant ($F_{1,94} = 4.19$, $p = 0.04$), while the residual variance after accounting for our planned contrast is not significant ($F_{1,94} = 0.06$, $p = 0.95$, two-tailed).[5] Third, the *q2*, which captures the residual between-cells contrast variance unexplained by

[4] For directional predictions, we report *p*-values that are one-tailed equivalents in the text, unless otherwise noted.
[5] As noted in Table 2, Panel C, the conventional ANOVA does report a significant interaction ($F_{1,94} = 2.90$, $p = 0.05$, one-tailed). However, Buckless and Ravenscroft [1990] suggests that contrast testing is the appropriate method for

the contrast versus the total variance explained in the experiment, is only 2.4 percent. This suggests that 97.6 percent of the systematic (i.e., between-cells) variance is explained by the contrast. Overall, the contrast test provides strong support for the ordinal interaction predicted by H1 and H2.

[Insert Table 2 and Figure 3 here]

For completeness, we also report the traditional ANOVA results in Table 2, Panel C. Consistent with our contrast test, we find support for the predicted interaction using the traditional ANVOA interaction term ($F_{1,94}$ = 2.90, p = 0.05).[6] Additionally, we conduct follow-up simple effects tests (reported in Table 2, Panel D). Consistent with H1, we find that simple effect for *Specialist Type* is significant in the *Less Specific* condition (t_{94} = 2.06, p = 0.02), but not in the *More Specific* condition (t_{94} = 0.31, p = 0.76, two-tailed).

H2 predicts that the directional effect of *Degree of Specificity* depends on *Specialist Type*. Specifically, H2 predicts that high elaboration, arising from more-specific specialist explanations, will decrease (increases) auditors' reliance on evidence from human specialists (specialist systems). For those in the *Human* condition, we find that participants' judgments are directionally consistent with this expectation, with proposed adjustments being lower in the *More Specific* condition than in the *Less Specific* condition explanation (19.20 versus 15.74, respectively).

---

our main analysis given that ANOVA models attribute most of the explained variance for an ordinal interaction to main effects.

[6] Included in our post-experimental questionnaire are several measures aimed at capturing participants' public accounting experience, including their specific experience with public clients, insurance clients, artificial intelligence, and proposing audit adjustments. In order to test if our results are robust to the inclusion of these experience measures, we first run a series of five 2 x 2 analysis of covariance (ANCOVA) models that control for each of the aforementioned demographic measures. Results show that our interactive effect remains directionally consist with our main results and retains at least marginal statistical significance (p-values range from 0.03 to 0.07). Further, we then adapt the 2 x 2 ANOVA model from our main analyses to include a three-way interaction between *Degree of Explanation*, *Specialist Type*, and each of the respective demographic measures (mean-centered for continuous variables). Across all models, the three-way interaction is not significant (all two tailed p-values greater than 0.26), providing additional support that the effects we observe in our main analysis are not influenced by our experience measures.

However, this difference is not significant at a traditional significance level (t_{94} = 1.12, p = 0.13). Within the *System* condition, the effect of *Degree of Specificity* is marginally significant (t_{94} = 1.30, p = 0.10) and directionally consistent with H2. We find that auditors propose larger adjustments in the *More Specific* condition versus the *Less Specific* condition (16.66 versus 12.76, respectively).

Overall, our findings demonstrate that when the opportunity to elaborate is inhibited, individuals rely on the simple cue of a human versus non-human source to determine their extent of reliance on the advice being provided, resulting in behavior indicative of algorithm aversion. However, this effect is mitigated when specialists' explanations do not constrain auditors' ability to elaborate. Furthermore, our results also provide some evidence that when their elaboration is low, auditors will default to reliance on evidence from human specialists, as this evidence comes from an "expert". In contrast, auditors' default response to system-provided evidence is to discount that evidence. Accordingly, higher elaboration (arising from more-specific explanations) decreases auditors' reliance on human specialists, but increases their reliance on specialist systems.

Moderated Mediation Analyses

In forming our hypotheses, we rely on the ELM, which postulates that variations in judgment and decisions depend on variations in a person's level of elaboration (i.e., amount of thought). When individuals have both high motivation and high capacity, attitude changes are based on a thoughtful consideration of issue-relevant information and an integration of that information into an overall position (Petty and Cacioppo [1986]). Specific to the audit setting, we posit that performing an effortful examination of the information available will have differential outcomes, resulting in auditors increasing their level of reliance on evidence provided by a specialist system while decreasing their reliance on information/advice provided by a human specialist.

When evaluating management estimates, audit standards state that auditors should first identify the assumptions that are important to the recognition or measurement of the estimate and then evaluate their reasonableness (PCAOB [2018b]). Further, auditors are required to determine whether the specialist's work provides sufficient appropriate evidence and when in question, are responsible for assessing whether the significant assumptions and methods used by the specialist were appropriate (PCAOB [2018a]). Given the importance of evaluating the underlying assumptions of an accounting estimate, we examine whether the effect of *Degree of Specificity* on auditors' proposed adjustments ("*Proposed*") operates indirectly through auditors' consideration of the assumptions underlying a specialist's independent estimate. Specifically, we ask participants to rate the extent to which they agree with the assumptions used by the firm specialist when evaluating the Insurance Claims Reserve (1 = "Not at All" to 7 = "A Great Deal"). Additionally, because our theory predicts *Degree of Specificity* will have opposite effects, depending on whether that evidence comes from a human or a system, we also examine whether this indirect effect is moderated by *Specialist Type* (i.e., moderated mediation).

Following the procedures described by Hayes [2018], we conduct a moderated mediation analysis using the SPSS macro (Model 8) with participants' agreement with the specialist's assumptions ("*Agree*") as the mediator. We present the results of this model in Figure 4, with results partitioned by *Specialist Type* for ease of interpretation. As reported in Figure 4, Panel A, in the *System* condition, the effect of *Degree of Specificity* on *Proposed* operates indirectly through *Agree* (90% confidence interval of 1.72 to 7.53), consistent with our expectations. In contrast, this indirect effect is not significant in the *Human* condition (see Figure 4, Panel B; 90% confidence interval of -1.77 to 1.19). Furthermore, the index of moderated mediation is significant (90 percent confidence interval of -6.37 to -0.16), which confirms that the indirect effects estimated at the two levels of *Specialist Type* are significantly different from one another.

Examination of the coefficients in the path model reveal that, in the *System* condition, when specialists provide more-specific explanations, auditors' agreement with the specialist's assumptions increases ($\beta_a = 0.68$, $p < 0.01$), leading to larger proposed adjustments ($\beta_b = 6.24$, $p < 0.01$). This is consistent with our theory-based prediction that increased elaboration (due to the provision of a more-specific explanation) results in less aversion to reliance on evidence provided by an AI-based specialist system. However, these relations differ in the *Human* condition. Interestingly, the relation between *Degree of Specificity* and *Agree* is negative, which is directionally consistent with our expectations, though this relation is not statistically significant ($\beta_a = -0.03$, $p = 0.46$). Although this relation is not significant, this provides some evidence that more-specific explanations lead to a more critical assessment of the specialist's work, which can lead to lower reliance and smaller adjustments ($\beta_b = 2.59$, $p = 0.03$).[7]

Supplemental Analyses

Evidence of Elaboration

Our hypotheses are predicated on the notion that less-specific specialist explanations will constrain auditors' ability to elaborate, while more-specific explanations will allow for greater elaboration. To test for differences in elaboration across these conditions, we first examine the amount of time participants spent on the audit case. Time has been used in prior literature to examine various elements of cognitive processing, including the quantity of cognitive processing

[7] In support of this notion, we also find that when the specialist provides a more-specific explanation around the assumptions used, participants rate the assumptions to be significantly more subjective when the valuation specialist is a human (M = 5.63) relative to a system (M = 4.29; $t_{94} = 3.29$, $p < 0.01$, one-tailed, untabulated).

(Bettman, Johnson, and Payne [1990]; Garbarino and Edell [1997]; Nolder, Ratzinger-Sakel, and Theis [2021]) and, specific to the ELM, the degree of elaboration (Chaiken [1980]). Results are as expected, with participants in the *More Specific* condition taking longer to complete the audit case relative to those in the *Less Specific* condition (t_{94} = 1.33, p = 0.09, one-tailed, untabulated), suggesting that the provision of a more specific explanation prompts higher elaboration.

Second, we follow Ratneshwar and Chaiken [1991], who hypothesize and find that message comprehension is an important prerequisite for systematic processing. Accordingly, we ask participants to rate how well they understood the valuation specialist's process for evaluating AGI's insurance claims reserve (*Understanding*, 1 = "Not at All" and 7 = "A Great Deal), with higher ratings indicating higher comprehension. Results reveal that participants in the *More Specific* condition indicate a higher level of *Understanding* relative to *Less Specific* (t_{94} = 1.56, p = 0.06, one-tailed, untabulated), providing additional evidence that participants in the *More Specific* condition had a greater opportunity to engage in a higher level of elaboration.

Lastly, we utilize the Linguistic Inquiry and Word Count (LIWC) text analysis application to analyze participants' open-ended responses where they provided justifications for their proposed audit adjustment. This approach allows for an unobtrusive measure of our participant's psychological processing through the examination of linguistic patterns (Tausczik and Pennebaker [2010]; Pennebaker, Boyd, Jordan, and Blackburn [2015]; Aghazadeh, Hoang, and Pomeroy [2021]).[8] We propose that higher elaboration should be reflected through the increased use of words in the LIWC *interrogative* category (comprised of 48 words such as *how*, *when*, and *what)*.

[8] One advantage of LIWC analysis is the reliability and validity of its dictionaries in producing high quality psychometric measures (Aghazadeh et al. [2021]). Recently, accounting researchers have begun using LIWC to analyze linguistic patterns in conference call transcripts (Hope and Wang [2018]), in written communications of auditors to audit committees (Fiolleau, Hoang, and Pomeroy [2019]), and in written justifications of managers' cooperation with auditors (Hatfield, Hoang, Ricci, and Thomas [2021]).

Interrogatives are words used in asking questions and are identified as an elaboration strategy in associate learning and recall tasks (Turnure, Buium, and Thurlow [1976]; Kestner and Borkowski [1979]; Kurtz, Reid, Borkowski, and Cavanaugh [1982]). The following is an example of an excerpt from a participant's justification for their proposed audit adjustment with a relatively greater use of this word category:

> "*<u>While</u>* management has expertise in making these estimates, the internal valuation group at the firm primarily focuses on these estimates, meaning that they know *<u>what</u>* to expect as far as economic trends and *<u>how</u>* other organizations are impacted by similar economic conditions. Because of this, I think the full adjustment should be made." (Interrogative words underlined and italicized)

Our LIWC analysis is based on word count frequencies, which is calculated by taking the words from the *interrogative* category scaled by the total words in each participant's response. The average length of participants' written justifications for their proposed audit adjustments is 82 (ranging from 2 to 262, untabulated) words. In accordance with best practices set forth by Aghazadeh et al. [2021] and LIWC guidelines, we remove any responses identified as probable outliers based on a Z score analysis (Field [2018]) to ensure that our analysis is not skewed by high word count responses and responses containing less than 25 words. Results are consistent with our expectations, with participants in the *More Specific* condition using relatively more *interrogative* words in their written responses (m = 1.25%) than those in the *Less Specific* condition (m = 0.91%, t_{78} = 1.36, p = 0.09, one-tailed, untabulated). Collectively, the results of these analyses provide convergent evidence that less-specific (more-specific) specialist explanations lead to lower (higher) elaboration – consistent with ELM.

V. DISCUSSION AND CONCLUSION

We report the results of an experiment that informs our understanding of how auditor's elaboration impacts their reliance on evidence provided by valuation specialists in several ways.

First, our study demonstrates, based on the Elaboration Likelihood Model, that when the capacity to elaborate is constrained due to the provision of a less specific explanation around the testing procedures performed by a valuation specialist, auditors rely on the simple cue of a human versus non-human source of evidence and discount their own firm's contradictory advice more heavily when it comes from a specialist system rather than a human specialist (i.e., algorithm aversion). This finding suggests that audit firms' implementation of AI systems to assist auditors in their performance of complex tasks may have unintended consequences that would prohibit potential improvements to audit quality.

Additionally, our study examines how characteristics of specialist-provided evidence influence auditors' ability to elaborate. While prior literature has often focused on how auditor characteristics (e.g., intrinsic motivation; Kadous and Zhou [2019] and epistemic motivation; Griffith [2018]) effect auditors' cognitive processing, our study demonstrates how more (versus less) specific explanations around the work performed by a valuation specialist leads auditors to engage in relatively higher (lower) levels of elaboration. Furthermore, auditors' elaboration on evidence provided by a valuation specialist has differing outcomes predicated on the type of specialist employed. Our study reveals that when given a greater opportunity to elaborate (i.e., when provided with a more specific explanation around the work performed), auditors increase their reliance on the evidence provided by a specialist system, mitigating the aversion exhibited when auditors are provided with a less specific explanation. Conversely, auditor's reliance on evidence provided from a human specialist reduces to a level consistent with that of a specialist system when provided with a more specific explanation.

The results of this study have implications for research related to the auditing of complex estimates. After observing several deficiencies regarding auditors' use of specialists when auditing complex accounting estimates, including overreliance on, and insufficient understanding of, assumptions used by experts (PCAOB [2010]), the PCAOB stated that auditors might not be exercising adequate professional skepticism, particularly with respect to reliance on specialists in the audit of fair values (PCAOB [2009]). Given that auditing standards explicitly require auditors to review and evaluate specialists' work (PCAOB [2019]), these deficiencies may not simply be due to a lack of adequate professional skepticism. Our study highlights how the degree of specificity contained within a valuation specialists' memo can influence the core mechanisms underlying auditors' cognition (i.e., elaboration), leading to differing levels of reliance depending on whether the specialist is a system or human. Further, the results of our moderated mediation analysis suggest that when auditors follow the central route of processing, this elaboration on the information provided by human valuation specialists does not equate to higher levels of agreement with the assumptions contained in the estimate. These results suggest that the deficiencies identified by the PCAOB may stem from insufficient motivation and/or capacity inhibiting auditor's from engaging in a higher level of elaboration rather than simply failing to exercise sufficient levels of professional skepticism. This notion is supported by Griffith [2018], who finds that a relational cue in a specialist's work only improves auditors' problems representation and corresponding judgments about estimates when auditors' epistemic motivation is sufficiently high.

This study also contributes to the growing body of research in psychology examining how individuals perceive, respond, and interact with new, automated forms of technology (Lyons and Havig [2014]; Mercado, Rupp, Chen, Barnes, Barber, and Procci [2016]). The auditing profession is investing billions of dollars to develop and implement AI systems with the goal of improving audit quality (FEI [2017]; EY [2018]; Bloomberg Tax [2020]). Such technological advancements may significantly alter the "timing, nature, and amount of information available to auditors" and "the judgments auditors make in critical areas of their audits" (PCAOB [2018c]). For example, in addition to performing routine, mundane tasks (enhancing audit efficiency), AI systems will also assist auditors with more challenging audit tasks, such as the evaluation of complex estimates (enhancing audit effectiveness) (Commerford et al. [2022]). However, despite the promising improvements to audit effectiveness and efficiency, these benefits will fail to be realized if auditors are hesitant to rely on the evidence produced by AI systems. Our findings echo this concern, with auditors relying more heavily on evidence produced by a human valuation specialist relative to a system specialist (i.e., algorithm aversion) when following the peripheral route of processing.

Our findings also demonstrate that evidence characteristics (e.g., degree of specificity) can influence auditors' elaboration and ultimately their level of reliance on specialist-provided evidence. Specifically, we find that greater elaboration, prompted by the provision of a more specific explanation, results in auditors increasing their reliance on the evidence provided system specialists, answering the call by Commerford et al [2022] to identify theory-grounded interventions designed to mitigate the effects of algorithm aversion. Lastly, while recent studies have begun to examine audit seniors' susceptibility to algorithm aversion and the potential consequences to audit quality (e.g., Commerford et al. [2022]; Johanns and Peecher [2022]), our study is the first to examine whether more experienced auditors (i.e., audit managers) are susceptible to this phenomenon. This is an empirical question given that Yalcin, Klesse, and Dahl [2019] find that individuals with higher (vs. lower) levels of perceived competence in the focal domain value recommendations more when they are generated by algorithms (vs. human experts). Based on the results of our study and prior findings, algorithm aversion appears to span across various levels of audit experience (audit seniors and audit managers), amplifying the importance of identifying interventions designed to mitigate this effect.

Our study is not without limitations. Participants in our study are only provided a limited and predetermined information set. In practice, auditors often have the ability to contact the valuation specialist and obtain additional information or documentation around the testing procedures the specialist utilized. Additionally, this study only investigates how one element (i.e., the degree of specificity) contained in a specialist's memo can influence auditors underlying cognitive processes. Future studies can investigate how specific elements of information content as well as elements of the specialist's memo (i.e., tone, readability, visualizations) influence auditors' levels of elaboration and whether this interaction leads to varying reliance determinations. Lastly, only 18.36% (18 out of 98) of our participants report having experience with AI systems. It is possible that as auditors repeatedly interact with such systems, auditors will become more familiar with and comfortable relying on them. However, Dietvorst et al. [2015] find that individuals with no experience using algorithms are more willing to rely on them, due to the fact that they have not seen the algorithm err. Future research could examine how encountering an error made by an AI system and/or how differing types of errors (e.g., fraud risk assessments, mechanical, conceptual) can influence auditors' willingness to rely on evidence provided by them.

REFERENCES

ADADI, A., AND M. BERRADA. "Peeking Inside the Black-box: A Survey on Explainable Artificial Intelligence (XAI)." *IEEE access* 6 (2018): 52138-52160.

AGHAZADEH, S.; K. HOANG; AND B. POMERY. "Using LIWC to Analyze Participants' Psychological Processing in Accounting JDM Research." *Auditing: A Journal of Practice & Theory* (2021).

ANDREWS, S. "From Inkmarks to Ideas: Current Issues in Lexical Processing." 2010. Psychology Press.

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.

BETTMAN, J. R.; E. J. JOHNSON; AND J. W. PAYNE. "A Componential Analysis of Cognitive Effort in Choice." *Organizational behavior and human decision processes* 45 (1990): 111-139.

BHATTACHARJEE, S., AND J. O. BROWN. "The Impact of Management Alumni Affiliation and Persuasion Tactics on Auditors' Internal Control Judgments." *The Accounting Review* 93 (2018): 97-115.

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

BORITZ, J. E.; N. V. KOCHETOVA; L. A. ROBINSON; AND C. WONG. "Auditors' and Specialists' Views About the Use of Specialists During an Audit." *Behavioral Research in Accounting* 32 (2020): 15-40.

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.

BROWN, J. O.; J. H. GRENIER; J. S. PYZOHA; AND A. REFFETT. "The Effects of Specialist Type and Estimate Aggressiveness on Jurors' Judgments of Auditor Negligence." *Auditing: A Journal of Practice & Theory* 38 (2019): 47-69.

BRYSBAERT, M. "How Many Words Do We Read Per Minute? A Review and Meta-Analysis of Reading Rate." *Journal of Memory and Language* 109 (2019): 1-30.

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

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.

CHAIKEN, S. "Heuristic Versus Systematic Information Processing and the Use of Source Versus Message Cues in Persuasion." *Journal of personality and social psychology* 39 (1980): 752.

COMMERFORD, B. P.; S. A. DENNIS; J. R. JOE; AND J. W. ULLA. "Man Versus Machine: Complex Estimates and Auditor Reliance on Artificial Intelligence." *Journal of Accounting Research* 60 (2022): 171-201.

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.

DIJKSTRA, J. J. "Legal Knowledge-Based Systems: The Blind Leading the Sheep?." *International Review of Law, Computers & Technology* 15 (2001): 119-128.

DOŠILOVIĆ, F. K.; M BRČIĆ; AND N. HLUPIĆ. "Explainable Artificial Intelligence: A Survey." *41st International convention on information and communication technology, electronics and microelectronics (MIPRO)* (2018): 210-215.

DOTY, J. "Statement on Proposed Auditing Standard on Auditing Accounting Estimates, Including Fair Value Measurements, and Proposed Amendments to Auditing Standards for Auditor's Use of the Work of Specialists." 2017. Speech delivered at PCAOB Open Board Meeting, Washington, DC. Available at https://pcaobus.org/news-events/speeches/speech-detail/statement-on-proposed-auditing-standard-on-auditing-accounting-estimates-including-fair-value-measurements-and-proposed-amendments-to-auditing-standards-for-auditor-s-use-of-the-work-of-specialists_652.

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

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

FIELD, A. *Discovering Statistics Using IBM SPSS Statistics 5th ed*. London: SAGE, 2018.

FIOLLEAU, K.; K. HOANG; AND B. POMEROY. "Auditors' Communications with Audit Committees: The Influence of the Audit Committee's Oversight Approach." *Auditing: A Journal of Practice & Theory* 38 (2019): 125-150.

FISCHHOFF, B. "Risk Perception and Communication Unplugged: Twenty Years of Process 1." *Risk analysis* 15 (1995): 137-145.

GARBARINO, E. C., AND J. A. EDELL. "Cognitive Effort, Affect, and Choice." *Journal of consumer research* 24 (1997): 147-158.

GLOVER, S. M.; M. H. TAYLOR; AND Y. WU. "Current Practices and Challenges in Auditing Fair Value Measurements and Complex Estimates: Implications for Auditing Standards and the Academy." *Auditing: A Journal of Practice & Theory* 36 (2017): 63-84.

GLOVER, S. M.; M. H. TAYLOR; Y. WU; AND K. T. TROTMAN. "Mind the Gap: Why Do Experts Have Differences of Opinion Regarding the Sufficiency of Audit Evidence Supporting Complex Fair Value Measurements?." *Contemporary Accounting Research* 36 (2019): 1417-1460.

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.; C. J. NOLDER; AND R. E. PETTY. "The Elaboration Likelihood Model: A Meta-Theory for Synthesizing Auditor Judgment and Decision-Making Research." *Auditing: A Journal of Practice & Theory* 37 (2018): 169-186.

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.

HATFIELD, R. C.; K. HOANG; M. RICCI; AND E. THOMAS. "Auditors’ Customer Relationship Management Activities and Managers’ Cooperation with Auditor Evidence Requests." Working paper, The University of Alabama, The University of Alabama, University of Florida, and Georgia College & State University, 2021. Available at https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3560921.

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

HERLOCKER, J. L.; J. A. KONSTAN; AND J. RIEDL. "Explaining Collaborative Filtering Recommendations." In *Proceedings of the 2000 ACM conference on Computer supported cooperative work* (2000): 241-250.

HOPE, O. K., AND J. WANG. "Management Deception, Big-Bath Accounting, and Information Asymmetry: Evidence from Linguistic Analysis." *Accounting, Organizations and Society* 70 (2018): 33-51.

JIMÉNEZ-LUNA, J.; F. GRISONI; AND G. SCHNEIDER. "Drug Discovery with Explainable Artificial Intelligence." *Nature Machine Intelligence* 10 (2020): 573-584.

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.

JOHANNS, A., AND M. PEECHER. “Algorithm Aversion in Fraud Risk Assessment: Implications of Technological Innovation for Professional Skepticism. Working paper, The University of Illinois Urbana-Champaign, 2022.

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

KADOUS, K.; C. J. NOLDER; AND M. PEECHER. “The role of the attitudes in auditing research and directions for future research.” In Albarracin, D., and B. T. Johnson (Eds.) 2nd ed., *Vol. 2. Handbook of attitudes*. New York: Routledge, 2018.

KADOUS, K., AND Y. ZHOU. "How Does Intrinsic Motivation Improve Auditor Judgment in Complex Audit Tasks?." *Contemporary Accounting Research* 36 (2019): 108-131.

KESTNER, J., AND J. G. BORKOWSKI. "Children's Maintenance and Generalization of an Interrogative Learning Strategy." *Child Development* (1979): 485-494.

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.

KURTZ, B. E.; M. K. REID; J. G. BORKOWSKI; AND J. C. CAVANAUGH. "On the Reliability and Validity of Children's Metamemory." *Bulletin of the Psychonomic Society* 19 (1982): 137-140.

LIBBY, R.; R. BLOOMFIELD; AND M. W. NELSON. "Experimental Research in Financial Accounting." *Accounting, Organizations and Society* 27 (2002): 775-810.

LYONS, J. B., AND P. R. HAVIG. "Transparency in a Human-Machine Context: Approaches for Fostering Shared Awareness/Intent." In *International Conference on Virtual, Augmented and Mixed Reality*. Springer: Cham, 2014.

MAHESWARAN, D.; D. M. MACKIE; AND S. CHAIKEN. "Brand Name as a Heuristic Cue: The Effects of Task Importance and Expectancy Confirmation on Consumer Judgments." *Journal of consumer psychology* 1 (1992): 317-336.

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.

MENDOZA, K. "Reducing Underreporting by Aggregating Budgeted Time." *The Accounting Review* 95 (2020): 299-319.

MERCADO, J. E.; M. A. RUPP; J. CHEN; M. J. BARNES; D. BARBER; AND K. PROCCI. "Intelligent Agent Transparency in Human–Agent Teaming for Multi-UxV Management." *Human factors* 58 (2016): 401-415.

MESHI, D.; G. BIELE, C. W. KORN; AND H. R. HEEKEREN. "How Expert Advice Influences Decision Making." *PLoS One* 7 (2012): e49748.

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. "A Model and Literature Review of Professional Skepticism in Auditing." *Auditing: A Journal of Practice & Theory* 28 (2009): 1-34.

NOLDER, C. J.; N. V. S. RATZINGER-SAKEL; AND J. C. THEIS. "Nudging Auditors' Unconscious to Improve Performance on an Accounting Estimate Task." *International Journal of Auditing* (2021): 1-16.

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.

PENNEBAKER, J. W.; R. L. BOYD; K. JORDAN; AND K. BLACKBURN. “The Development of Psychometric Properties of LIWC 2015.” (2015). Available at: http://liwc.wpengine.com/wp-content/uploads/2015/11/LIWC2015_OperatorManual.pdf.

PETTY, R. E., AND P. BRIÑOL. "The Elaboration Likelihood and Metacognitive Models of Attitudes." *Dual-Process Theories of the Social Mind* (2014): 172-187.

PETTY, R. E., AND J. T. CACIOPPO. "Source Factors and the Elaboration Likelihood Model of Persuasion." *ACR North American Advances* (1984).

PETTY, R. E., AND J. T. CACIOPPO. “The Elaboration Likelihood Model of Persuasion.” *Advances in Experimental Social Psychology* 19 (1986): 123–205.

PETTY, R. E.; J. T. CACIOPPO; AND D. SCHUMANN. "Central and Peripheral Routes to Advertising Effectiveness: The Moderating Role of Involvement." *Journal of Consumer Research* 10 (1983): 135-146.

PETTY, R. E.; D. T. WEGENER; AND L. R. FABRIGAR. "Attitudes and Attitude Change." *Annual Review of Psychology* 48 (1997): 609-647.

PETTY, R. E.; S. C. WHEELER; AND Z. L. TORMALA. "Persuasion and Attitude Change." *Handbook of Psychology* (2003): 353-382.

PUBLIC COMPANY ACCOUNTING OVERSIGHT BOARD (PCAOB). *Auditing Fair Value Measurements and Using the Work of a Specialist*. Standing Advisory Group Meeting. October 14–15. Washington, DC: PCAOB, 2009.

PUBLIC COMPANY ACCOUNTING OVERSIGHT BOARD (PCAOB). *Report on Observations of PCAOB Inspectors Related to Audit Risk Areas Affected by the Economic Crisis*. PCAOB Release No. 2010-006. September 29. Washington, DC: PCAOB, 2010.

PUBLIC COMPANY ACCOUNTING OVERSIGHT BOARD (PCAOB). *Staff Inspection Brief: Preview of Observations from 2015 Inspections of Auditors of Issuers*. Washington, DC: PCAOB, 2016.

PUBLIC COMPANY ACCOUNTING OVERSIGHT BOARD (PCAOB). *Proposed Amendments to Auditing Standards for Auditor’s Use of the Work of Specialists*. Washington, DC: PCAOB, 2017.

PUBLIC COMPANY ACCOUNTING OVERSIGHT BOARD (PCAOB). Supervision of the Audit Engagement. *Auditing Standard 1201.* Washington, DC: PCAOB, 2018a.

PUBLIC COMPANY ACCOUNTING OVERSIGHT BOARD (PCAOB). Auditing Accounting Estimates, Including Fair Value Measurements. *Auditing Standard 2501.* Washington, DC: PCAOB, 2018b.

PUBLIC COMPANY ACCOUNTING OVERSIGHT BOARD (PCAOB). *Strategic Plan 2018-2022*. Washington, DC: PCAOB, 2018c. Available at: https://pcaob-assets.azureedge.net/pcaob-dev/docs/default-source/about/administration/documents/strategic_plans/pcaob-2018-2022-strategic-plan.pdf?sfvrsn=d74236b3_0.

PUBLIC COMPANY ACCOUNTING OVERSIGHT BOARD (PCAOB). *Staff Guidance: Supervising or Using the Work of an Auditor’s Specialist.* Washington, DC: PCAOB, 2019.

PYZOHA, J. S.; M. H. TAYLOR; AND Y. J. WU. "Can Auditors Pursue Firm-Level Goals Nonconsciously on Audits of Complex Estimates? An Examination of the Joint Effects of Tone at the Top and Management's Specialist." *The Accounting Review* 95 (2020): 367-394.

RATNESHWAR, S., AND S. CHAIKEN. "Comprehension's Role in Persuasion: The Case of its Moderating Effect on the Persuasive Impact of Source Cues." *Journal of Consumer Research* 18 (1991): 52-62.

RICH, J. S. "Reviewers' Responses to Expectations About the Client and the Preparer." *The Accounting Review* 79 (2004): 497-517.

ROWE, S. P. "Auditors' Comfort with Uncertain Estimates: More Evidence is Not Always Better." *Accounting, Organizations and Society* 76 (2019): 1-11.

SAMEK, W., AND K. R. MÜLLER. "Towards Explainable Artificial Intelligence." *Explainable AI: Interpreting, Explaining and Visualizing Deep Learning* (2019): 5-22.

TAUSCZIK, Y. R., AND J. W. PENNEBAKER. "The Psychological Meaning of Words: LIWC and Computerized Text Analysis Methods." *Journal of Language and Social Psychology* 29 (2010): 24-54.

TURNURE, J.; N. BUIUM; AND M. THURLOW. "The Effectiveness of Interrogatives for Promoting Verbal Elaboration Productivity in Young Children." *Child Development* (1976): 851-855.

YALCIN, G.; A. K. KLESSE; AND D. DAHL. "The Expert Vs. the Algorithm: Consumers' Perceived Competence Impacts Their Valuation of Product Recommendations." *ACR North American Advances* (2019).

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

Appendix A

Participants in the "*Less Specific*" condition will view the instrument as written without the italicized bold text. Participants in the "*More Specific*" condition will view the instrument with the inclusion of both regular black text and italicized bold text.

Participants in the "*Human*" specialist condition will view the instrument as written without the italicized words in brackets. Participants in the "*System*" specialist condition will view the instrument as written, except with the italicized words in brackets instead of "internal valuation group" or "we".

Excerpts from Jones & Company's Internal Valuation Group's/[*Valuation Systems*] Report

The purpose of this report is to summarize the results of procedures performed by the audit team's internal valuation group/[*Valuation System*] for the audit of AG Insurance for the fiscal year ended March 31, 2021. This report outlines the testing procedures and states the overall conclusions made by the internal valuation group/[*Valuation System*]. For purposes of establishing an appropriate Insurance Claims Reserve, one of the most important factors to consider is the expected frequency of future claims. Changes in auto claim frequency may result from changes in mix of business, miles driven, and/or other macroeconomic factors.

To calculate the impact of such changes on the Insurance Claims Reserve estimate, the internal valuation group/[*Valuation System*] used a "chain ladder" approach, which is a common actuarial technique. In the chain ladder estimation technique, a ratio (development factor) is calculated for each accident year and then compounded over the remaining future periods to calculate an estimate of ultimate losses for each accident year.

Findings

Reserve estimates typically incorporate expectations for future mobility based on mobility patterns observed in previous years. Historically, there is a positive association between mobility patterns and the frequency of claims. However, the COVID-19 pandemic significantly altered mobility patterns. As such, historical mobility data has limited usefulness in predicting the frequency of claims in the coming year. Therefore, to accurately estimate the frequency of claims and the Insurance Claims Reserve, it is necessary to incorporate additional assumptions about future mobility.

In April of 2020, overall mobility in the United States was about 50 percent lower than normal (based on a historical three-year average), but the internal valuation group/[*Valuation System*] is projecting that this decline is temporary. Mobility levels have been consistently increasing since April 2020 and the internal valuation group/[*Valuation System*] expects this increasing trend to continue in 2021. The internal valuation group/[*Valuation System*]'s independent estimate assumes that 2021 mobility levels will be about 10 percent lower than previous years (i.e. 90 percent of normal levels).

The list below details specific factors that lead to the expectation for rising future mobility:

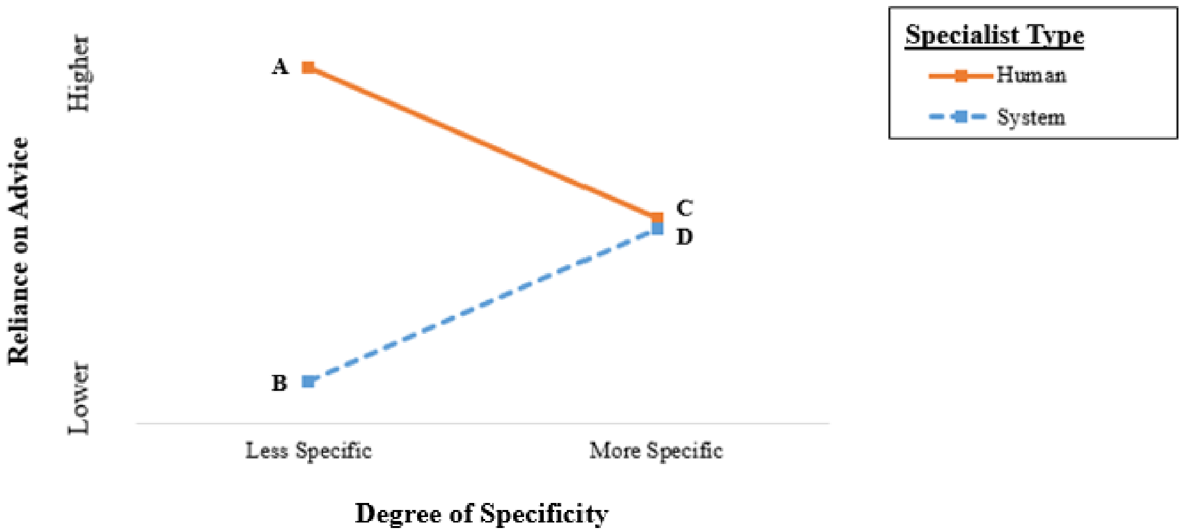
- ***State and local governments will continue to ease restrictions on business operations, school attendance, and recreational travel.***
- ***As travel restrictions ease, some organizations will maintain work-from-home options, but in insufficient numbers to allow overall mobility patterns to remain at such unprecedentedly low levels.***
- ***Usage of public transportation is still 35 percent lower than normal. Lower usage of public transportation is associated with increased utilization of private vehicles. This inherently***

impacts drivers of differing demographics, leading to higher risk drivers taking to the roadways.

- ***The TSA is reporting significant increases in the number of air travelers. In April 2020, the average number of daily travelers was approximately 109,000. In March 2021, the average number of daily travelers was about 1.2 million, indicating increasing demand for business and recreational travel.***
- ***Sales of recreational vehicles have increased 22 percent year-over-year. Further, park visitations were consistently well above the norm during the summer months, peaking at an approximate 64% increase in July 2020 (compared to a 5-week baseline period between January and February 2020.) As such, reduced mobility due to virtual work and learning arrangements is partially offset by increased recreational travel.***

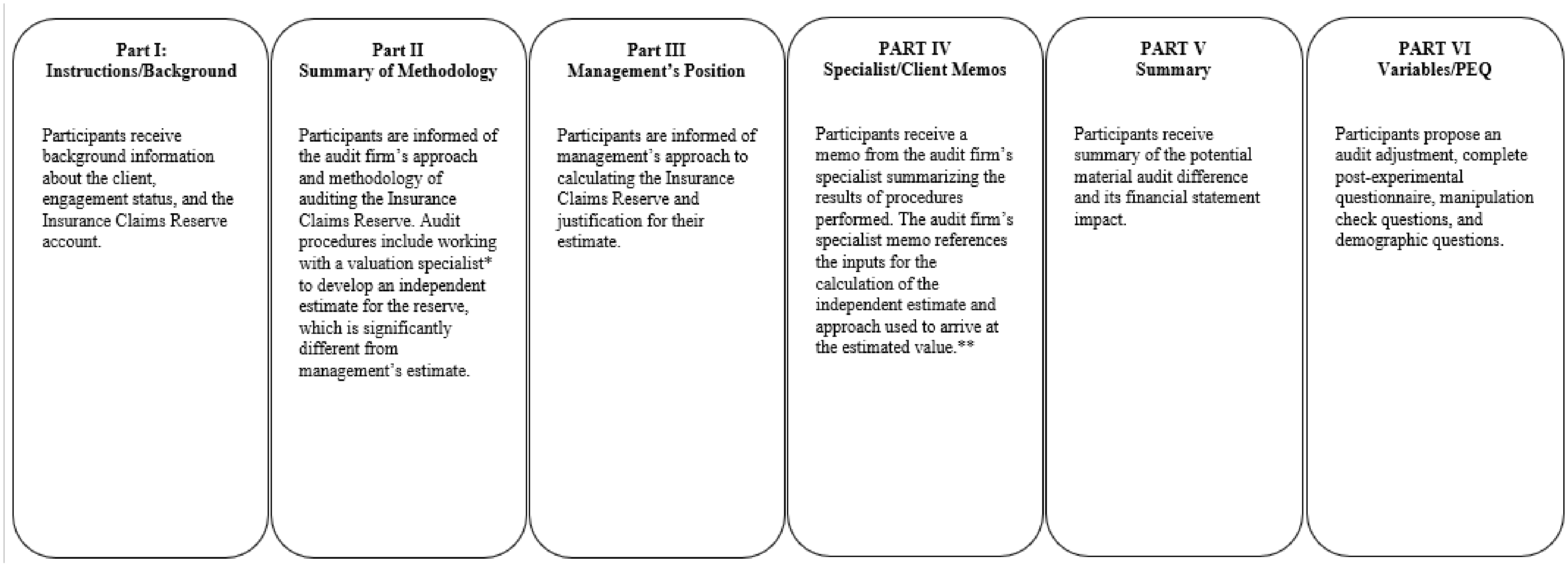
Based on the procedures performed, the internal valuation group/*Valuation System*] estimates that the Insurance Claims Reserve should be $588 million.

FIGURE 1
Graphical Representation of Hypothesized Results



Note: Our hypothesis predicts that the provision of a more specific explanation will cause auditors to decrease their reliance on evidence from human valuation specialists, manifesting in lower proposed audit adjustments than when provided with a less specific explanation. However, when provided with a more (versus less) specific explanation from a specialist system, auditors will be pulled out of their naturally aversive state and increase their reliance on the evidence provided, manifesting in higher proposed audit adjustments.

FIGURE 2
Experimental Design

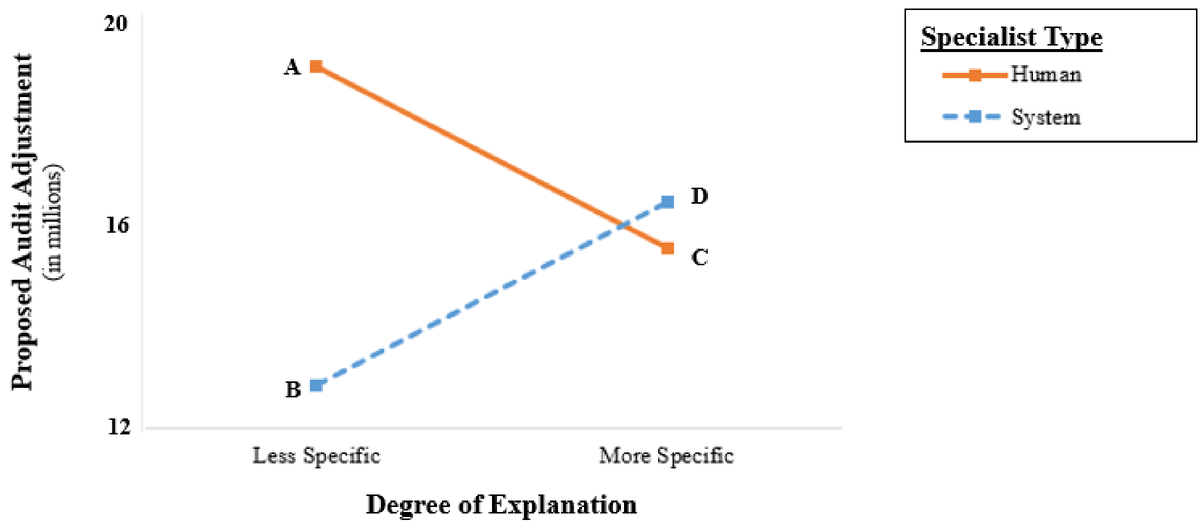


Note: This figure presents the flow of the experimental design

* *Specialist Type* is manipulated as human valuation specialist (i.e., internal valuation group) or specialist system.
** *Degree of Explanation* is manipulated as a less specific or more specific explanation within a specialist's memo related to the development of the specialists' independent estimate of AGI's Insurance Claims Reserve.

FIGURE 3

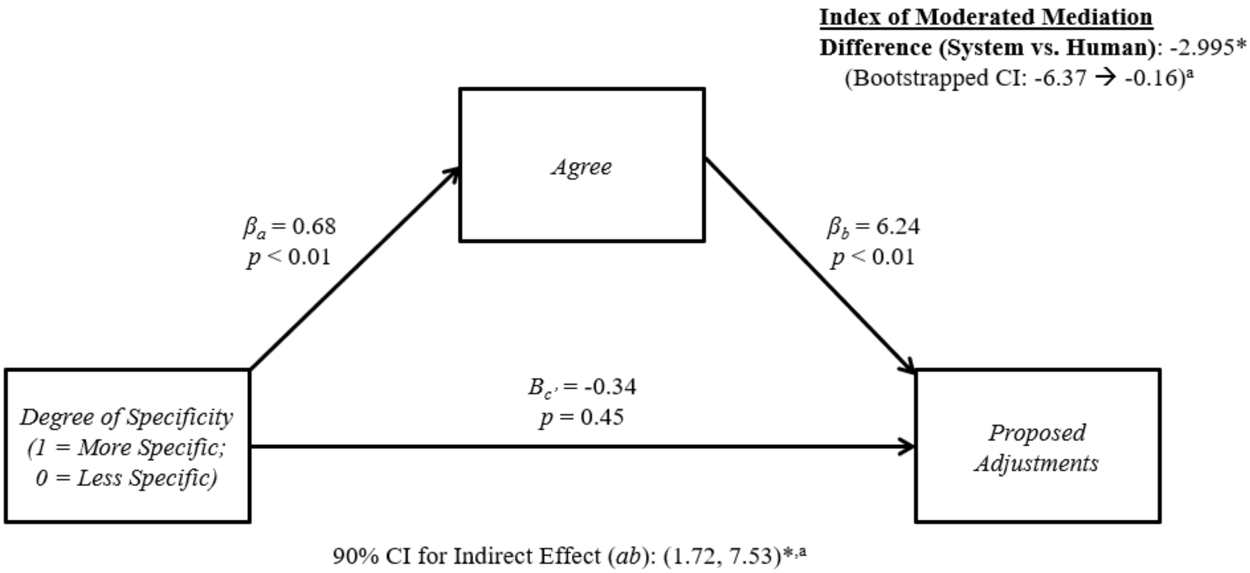
Observed Effects of Degree of Specificity and Specialist Type on Proposed Audit Adjustments



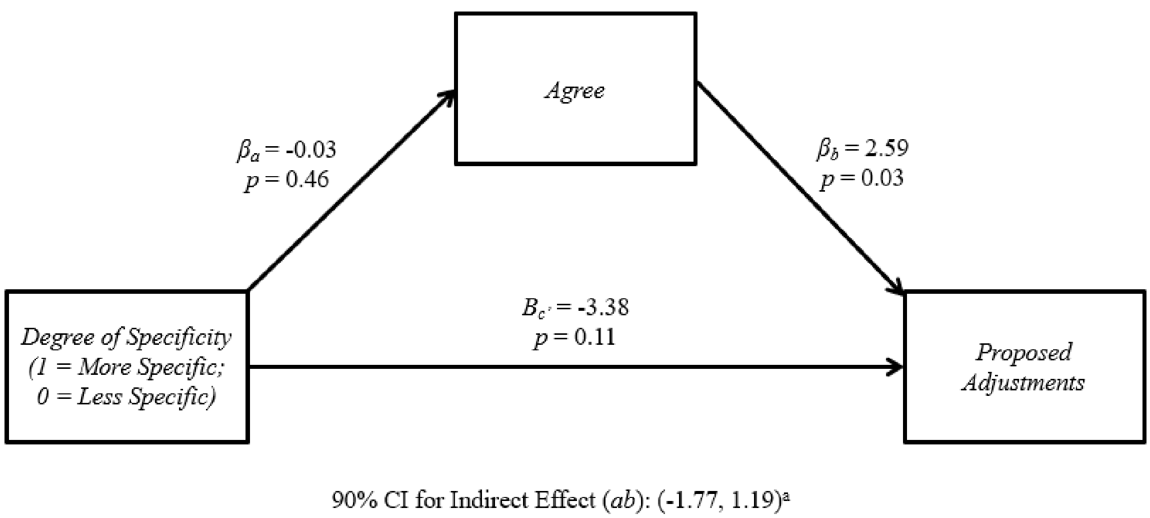
Note: See notes to Table 2 for descriptions of dependent variable and independent factors.

FIGURE 4
Moderated Mediation Analysis

Panel A: *System* Conditions



Panel B: *Human* Conditions



Note: The above diagram represents a moderated mediation model (Hayes [2018]). Analysis was conducted using the SPSS Process macro (Model 8) to examine how the indirect effect of *Degree of Specificity* (X) on *Proposed Adjustments* (Y) through *Agree* (M) differs depending on *Specialist Type* (W). To capture *Agree*, we asked participants "To what extent do you agree with the assumptions used by the Valuation System/[*internal valuation group*] when evaluating the Insurance Claims Reserve?" (1 = "Not at All" and 7 = "A Great Deal"). Results are presented separately for the *System* and *Human* conditions for visual simplicity and all continuous variables are mean-centered to facilitate interpretation of the coefficients.

[a] To test for the significance of indirect effects, we use 90% confidence intervals from bootstrapped sampling distributions (based on 10,000 bootstrap samples) for the product of paths *a* and *b* (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.

* denotes statistical significance equivalent to $p < 0.05$, one-tailed.

TABLE 1
Demographic Profile of Participants

Mean public accounting experience (in years)	6.94
Percentage of participants who have experience with insurance companies	18.36%
Percent of participants' time working on public clients	49.60%
Likelihood of providing input for proposed audit adjustments (1 = "Not at All Likely" and 7 = "Highly Likely")	5.77
Percentage of participants who have experience using an AI system	18.36%

TABLE 2
Proposed Audit Adjustments

Panel A: Descriptive statistics for *Proposed Audit Adjustments* (mean, (standard error), [n])

	Degree of Specificity		
Specialist Type	*Less Specific*	*More Specific*	Overall
Human	19.20 (1.85) [23] A	15.74 (2.29) [24] C	17.44 (1.48) [47]
System	12.76 (2.30) [23] B	16.66 (2.13) [28] D	14.90 (1.57) [51]
Overall	15.98 (1.54) [46]	16.24 (1.54) [52]	

Panel B: Contrast and Residual Between Cells Variance Test[a]

Source of Variation	Sum of Squares	*df*	*F*	*p*
[+1, -1, 0, 0] for [A, B, C, D]	477.46	1	4.19	0.04
Residual between-cells variance	12.49	2	0.06	0.95
Contrast Residual Variance, q^2	2.4%			

Panel C: Conventional ANOVA[a]

Source of Variation	Sum of Squares	*df*	*F*	*p*
Specialist Type	185.71	1	1.63	0.10
Degree of Specificity	1.16	1	0.01	0.92
Specialist Type × Degree of Specificity	329.85	1	2.90	0.05
Error	10,703.57	94		

Panel D: Simple Effects with Dependent Variable – *Proposed Audit Adjustment*[a]

Source of Variation	*df*	*t*	*p*
Effect of *Specialist Type* given *Less Specific* (A vs B)	94	2.06	0.02
Effect of *Specialist Type* given *More Specific* (C vs D)	94	0.31	0.76[b]
Effect of *Degree of Specificity* given *Human* (A vs C)	94	1.17	0.13
Effect of *Degree of Specificity* given *System* (B vs D)	94	1.30	0.10

Note: The dependent variable is participants' proposed audit adjustments. *Specialist Type* and *Degree of Specificity* are both manipulated at two levels (human valuation specialist vs. specialist system and less specific vs. more specific, respectively), between participants.

[a] Consistent with our directional predictions, *p*-values reported in Panel B and Panel C are one-tailed equivalents. Similarly, all *p*-values reported in Panel C are one-tailed, unless noted otherwise.

[b] Given we expect the means in Cell C and Cell D will not significantly differ from each other, the reported *p*-value is equivalent to a two-tailed test.