When does reliance on technology elevate auditor liability?

ABSTRACT

Audit firms continue to make large investments in audit technologies to enhance effectiveness and efficiency, while simultaneously creating the potential for elevated liability due to well-documented psychological aversions to use of technology in judgment and decision-making. We develop a framework that lays out a process for investigating and understanding the nascent research on how technology use affects auditor liability. As an initial study guided by this framework, we predict that reliance on technology will only result in elevated liability when the audit task is relatively subjective, thereby requiring professional judgment, and the effect will be larger when the undetected misstatement is due to management fraud (vs. error). We predict no effect on auditor liability for objective tasks requiring little to no judgment. Results from two experiments with jury-eligible individuals support these predictions as we only observe elevated liability for subjective audit tasks involving management fraud. In other words, jurors are generally receptive to reliance on technology, but do not view technology as an adequate replacement when auditors need to simultaneously exercise professional judgment and detect cues of management deception. Although more research is necessary, our study should give practitioners a degree of comfort that technology reliance will only elevate auditor liability under specific conditions and not elevate unilaterally due to the use of advanced technology during the audit.

Keywords: auditing, technology, liability, negligence, jurors, task subjectivity, fraud, error

Data: available from authors by request

1. Introduction

Advanced technologies such as drones, artificial intelligence, robotic process automation, and advanced data analytics are revolutionizing the audit profession (Forbes Insights 2018; KPMG 2021). Although the effectiveness and efficiency benefits of advanced technologies are rarely disputed (EY 2017), there are concerns about possible escalation of auditor liability if the use of technology promotes second-guessing of the auditors' judgments and decisions (Munoko et al. 2020, Kipp et al. 2022; Grenier et al. 2021). The purpose of our study is *not* to examine the validity of concerns about technologies elevating auditor liability, but to elucidate the conditions under which such escalation (or perhaps reduction) is likely to manifest. That is, we theorize that technology reliance will not either unequivocally escalate, reduce, or not affect auditor liability, but instead will be contextually dependent on a set of predictable factors. Thus, our study is intended to be an important first step towards understanding this context dependence by offering a theoretical framework for predicting the effects of current and future factors of interest and validating our framework by examining two such factors. Specifically, we examine how the degree of subjectivity (i.e., the professional judgment required) and the intentionality of the material misstatement (i.e., error vs. fraud) jointly affect how jurors' evaluations of auditors are influenced by reliance on advanced audit technologies.

Concerns about technology use elevating auditor liability stem from the psychology and emerging accounting literature on algorithm aversion (Promberger and Baron 2006; Önkal et al. 2009; Eastwood et al. 2012; Commerford et al. 2022). Algorithm aversion is the tendency for individuals to rely more on humans than technology, even when doing so results in objectively worse outcomes and such technologies have been shown to be more accurate. Algorithm aversion has been documented in the medical field with people preferring the advice of a human medical practitioner over advanced diagnostic technologies (Promberger and Baron 2006; Longoni et al.

2019), in managerial forecasting (Fildes and Goodwin 2007; Dietvorst et al. 2018), and in audit judgment (Commerford et al. 2022). However, the fact that people tend to prefer humans over technology does not necessarily mean that they will punish others for using technology when evaluating the quality of others' judgments and decisions. In other words, it is likely this "second-order" algorithm aversion is context-dependent (cf. Castelo et al. 2019).

Our theoretical framework is designed to help researchers identify the conditions under which algorithm aversion is likely to manifest in the auditor liability setting. The framework specifies two technology liability determinants (i.e., factors predictive of second-order algorithm aversion): consensus and fit. Consensus is the commonality of the use of the technology for the task by other auditors. Fit is the match between the technology's abilities and the requisite capabilities for the task. Both of these determinants are theorized to be inversely related to auditor liability with stronger consensus and/or fit reducing perceptions of auditor negligence. Our framework then lays out a process for (a) identifying auditor, task, environmental, and technology characteristics that affect consensus and fit, (b) connecting the resulting effects to the predominant theoretical lenses used in the auditor liability literature, and (c) formulating expectations.

As an initial test of our theoretical framework, we conduct a study that tests the predictions of our framework when we vary two aspects of fit. We hold consensus constant as Kadous and Mercer (2012) find that jurors provide harsher evaluations of auditors who do not follow the consensus, which provides us some information about the likely relation between consensus, technology use, and auditor liability. However, to our knowledge, the impact of task-technology fit has not been investigated in the auditor negligence literature. Task-technology fit highlights the importance of matching appropriate technological functionalities to the demands imposed by a given task (Goodhue and Thompson 1995) and suggests that a better match yields improved outcomes. We posit that two distinct aspects of fit will jointly impact jurors' counterfactual reasoning (our chosen theoretical lens from the auditor liability literature), helping to predict when and why we see evidence of algorithm aversion in jurors' negligence assessments.

We specifically investigate how the degree of subjectivity of the task (objective vs. subjective) and intentionality of the misstatement (error vs. fraud) affect jurors' judgments. For objective audit tasks, jurors likely view technology as a suitable substitute for humans given such tasks are logical, rules-based tasks requiring minimal to no human judgment (e.g., counting inventory). This strong fit will make it difficult for jurors to think about what human auditors could have done differently from technology (and vice versa) to detect a misstatement in this scenario. As such, negligence assessments are unlikely to vary with technology use for objective audit tasks in which the technology possesses the requisite capabilities.

Conversely, for subjective audit tasks, jurors likely do not view technology as a sufficient substitute due to its inability to exercise professional judgment. In other words, there is a poor fit between technology and a subjective audit task that requires judgment. Because jurors are less likely to trust technology that lacks the abilities required to perform the task (Castelo et al. 2019), jurors will likely experience more intense counterfactual thoughts when auditors use technology for subjective audit tasks.

Further, we expect the effect of auditor use of technology on jurors' counterfactual intensity to be larger for fraudulent misstatements. Individuals tend to believe that technology lacks certain key capabilities that only humans possess (Gray et al. 2007; Haslam et al. 2008). For example, understanding non-verbal behavioral cues and emotion, as well as detecting context-based verbal cues, would seem to fall into this category and may be perceived as critical to detecting deceptive fraudulent behavior. Given that counterfactual intensity leads to harsher evaluations of auditors (Reffett 2010), we expect that jurors' evaluations of auditor negligence will increase with auditors' use of technology for subjective tasks, and the increase will be greater when auditors fail to detect a material misstatement due to fraud rather than an error.

To test our predictions, we conduct two experiments where jury-eligible individuals evaluate the potential negligence of an audit firm who failed to detect a material misstatement in a hypothetical civil suit. Both experiments employ 2 x 2 between-participants designs with manipulations of technology reliance (human vs. technology) and misstatement intentionality (error vs. fraud). In Experiment 1, the audit firm completes the objective audit task of counting inventory using either a human auditor or a drone. Experiment 2 centers on the subjective audit task of evaluating management's estimate of a lease liability. The audit firm relied on either artificial intelligence technology or a human auditor to identify the relevant terms of the lease and calculate the associated liability.

Results are consistent with our predictions. As expected, we do not find that technology use affects perceptions of negligence for the objective audit task in Experiment 1. Jurors thought about what could have been done differently (i.e., using humans), but did not believe humans would have been any more reliable than the technology, consistent with the objective nature of the task. With respect to Experiment 2's subjective task, we find support for our prediction of increased liability due to technology use when the misstatement arises from fraud as opposed to error. When auditors relied on technology, the intensity of jurors' counterfactual thoughts about what could have been done differently is significantly higher following an undetected fraud compared to an undetected error. However, when auditors did not rely on technology, jurors do not believe another auditor could have done better and, therefore, do not experience more intense counterfactual thoughts following an intentional versus unintentional misstatement. Taken together, these results indicate that technology reliance only elevates liability when such technology is used for subjective audit tasks and the undetected misstatement was intentionally made by management. In these cases, jurors do not seem to view technology as a suitable replacement for human professional judgment.

Our study contributes to the auditor liability literature by offering a theoretical framework that predicts the conditions under which the use of advanced audit technologies is likely to elevate perceptions of auditor negligence. In doing so, we simultaneously help reconcile the seemingly conflicting results of early studies and guide future studies. There is a growing literature documenting mixed effects of audit technology on auditor liability. For example, Kipp et al. (2022) find elevated negligence likelihoods for auditors utilizing artificial intelligence to identify risky transactions. While the use of this technology is a high consensus activity and there are no allegations of fraud, this task requires professional judgment thereby weakening fit and potentially explaining the resulting algorithm aversion. Two studies find that use of audit data analytics has no effect on auditor liability and actually provides liability protection under certain conditions (Ballou et al. 2021; Barr-Pulliam et al. 2022), but these studies focus on the relatively objective audit task of substantive testing where use of advanced technology is likely viewed as high in both consensus and fit, thereby explaining the lack of algorithm aversion and some evidence of algorithm appreciation. Although these early studies have highlighted several important factors that affect how jurors evaluate auditor technology use, there are likely others that our framework will help researchers identify and test.

Our study also makes a theoretical contribution to the psychological literature on algorithm aversion (Promberger and Baron 2006; Önkal et al. 2009; Eastwood et al. 2012; Commerford et al. 2022) by examining the conditions under which the phenomenon extends to the evaluation of others' judgments and decisions. That is, the extant algorithm aversion literature examines the extent to which individuals are willing to rely on technology-based advice over advice from humans, and find that under most conditions, they prefer human-based advice. However, before our study, it is unclear whether this also applies to evaluations of others' judgment and decision quality, such as jurors' evaluation of professional negligence. Consistent with concurrent research (Kipp et al. 2022, Cui 2021), we find that algorithm aversion does extend to such evaluations, but only in the presence of certain contextual factors.

2. Background and Theoretical Framework

Background

Audit firms are investing heavily in the use of technology (Kapoor 2020), with the goal of enhancing both audit efficiency and effectiveness (Harris 2017).[1] The use of technology, though, is not standardized across audit firms or audit engagements. This is in part because auditing standards do not require the use of technology-based audit tools coupled with the reality that the audit firms differ, especially across firms of different sizes/resources, as to the type and extent of technology-based tools available for use on audit engagements.

Further, even with the best available technology, auditors are unable to provide absolute assurance that the financial statements are free of material misstatements. Consequently, audit failures are still possible. Audit failures occur when material unintentional errors and/or intentional fraud goes undetected. The adverse outcome associated with such audit failures serves as motivation for investors and creditors to bring lawsuits against the auditors. If the suit goes to trial, jurors must decide whether the audit firm acted negligently and, if so, whether the audit firm should pay damages to the injured parties.

[1] For example, some tools are being developed to independently analyze board meeting minutes in order to identify risks and requesting relevant supporting information (AICPA 2020).

Concerns have been raised about the use of technology-based audit tools potentially elevating auditors' litigation due to enhanced second-guessing of auditors' judgments and decisions (Kipp et al. 2022; Cui 2021; Grenier et al. 2021; Munoko et al. 2020). Such second-guessing may be the result of jurors' algorithm aversion, the tendency for individuals to prefer the advice of humans over technology when making judgments and decisions. However, this personal aversion may not always translate to harsher evaluations of others' judgments and decisions who choose to rely on technology. In other words, it is likely that this second-order algorithm aversion is context-dependent.

Theoretical Framework

This section develops our theoretical framework for predicting the conditions under which algorithm aversion is likely to exist in the auditor liability setting (see Figure 1). The framework is a melding of relevant psychology theories, including algorithm aversion, and the extant auditor liability literature. Although the framework can be used retrospectively to reconcile the results of previous studies, we feel the greater value will be in stimulating and guiding future research. To that end, we include a host of potential factors for future scholarly inquiry and make the framework adaptable to any theoretical lens, while also laying out potential connections to the current predominant lenses. While discrete steps are identified, we expect using the framework to be a recursive process where scholars often return to earlier steps after consideration of later steps.

[INSERT FIGURE 1 HERE]

Step 1 involves identifying the independent factors of interest. Some factors may be clear from practice or theory, and thus already identified, for a given research question. However, other scholars may benefit from consideration of the host of factors that could potentially affect how jurors evaluate auditor technology use. We categorize these potential factors consistent with

Bonner's (1999) categories of person (change to "auditor" consistent with Nelson and Tan 2005), task, and environment characteristics. We then added a fourth category for technology characteristics.

Step 2 involves assessing the potential importance of a given factor (or set of factors) by considering how a factor potentially affects one or our two technology liability determinants: consensus and fit. Although there could conceivably be a factor that does not affect either of these determinants but still affects auditor liability, we feel that the vast majority of liability-relevant factors would affect one or both of these determinants. Consequently, if a given factor would not conceptually affect either of these determinants, scholars should consider returning to Step 1. In fact, they could use these drivers to inform Step 1 by thinking of factors relevant to these determinants.

We identified the two determinants by collectively considering the relevant aspects of the algorithm aversion and auditor liability literatures, including the nascent studies on the effect of technology use on auditor liability. *Consensus* is the extent to which other auditors use the technology. Actions that deviate from behavioral norms are often perceived as salient causes of the adverse outcome as it is easier for jurors to engage in counterfactual reasoning about how the outcome could have been avoided in the presence of such deviations (Kahneman and Miller 1986). Reffett (2010) finds that jurors' assessments of auditors' negligence increase as the intensity of their counterfactual thoughts increases. This is consistent with Kadous and Mercer's (2012) finding that auditors are penalized for failure to comply with norms when the accounting standards are imprecise.

Fit is whether the technology has the perceived requisite capabilities to effectively complete the task. Task-technology fit is defined as "the degree to which a technology assists an individual in performing his or her portfolio of tasks" (Goodhue and Thompson, 1995, p. 216). The best outcomes occur when one matches the appropriate technological functionalities to the demands imposed by a given task. Castelo et al. (2019) discuss how consumer trust in algorithms depends on the capabilities that they perceive the algorithm to possess. We expect this to apply to second order reactions to algorithm use as well with jurors being less (more) likely to punish auditors for using technology that does (does not) have the requisite capabilities. An example is the extent to which the technology is asked to exercise professional judgment. Tasks fall on a range from application of logic and rules (i.e., cognitive tasks) to intuition and gut feelings (i.e., emotional tasks; Inbar et al. 2010). Humans are less likely to trust technology that is asked to perform emotional tasks (Haslam et al. 2008; Castelo et al. 2010). As professional judgment entails intuition and emotion (Ranzilla et al. 2011), we would expect jurors to punish auditors for using technology for professional judgment.

Step 3 involves scholars choosing a theoretical lens for their study. Although Steps 1 and 2 may be sufficient to formulate hypotheses, we encourage scholars to connect technology liability determinants to existing theoretical drivers of liability. Doing so will enable a richer theoretical process description and promote comparability to the literature, thereby clarifying and enhancing a study's contribution. In Figure 1, we have listed and briefly summarized the three predominant theoretical lenses in the auditor negligence literature: motivated reasoning (Kadous 2000), Culpable Control Model (Backof 2015), and counterfactual thinking (Reffett 2010). We have also presented potential relationships of the technology liability determinants with the primary drivers of liability under each theoretical lens. It is important to note that although we believe these relationships are intuitive, they should be empirically tested by future research. After considering the interplay of the technology liability determinants within the chosen theoretical lens, Step 4 concludes with formulation of a study's hypotheses.

3. Hypotheses Development

Given that the prior auditor negligence literature has shown that *consensus* matters (Kadous and Mercer 2012) and thus provides us with an initial starting point for consensus in the framework, we conduct a study to validate the other key determinant of auditor liability listed in our framework. Specifically, we examine how two determinants of fit (i.e., subjectivity of the task and the intentionality of the misstatement) affect jurors' assessments of auditor negligence through the theoretical lens of counterfactual reasoning.

Task Subjectivity

Subjectivity of an audit task is defined as the amount of judgment required to complete the task. Relatively objective audit tasks are those that require little, if any, judgment and vice versa. For example, counting easily identifiable physical inventory (e.g., aluminum cans) is an audit task that is relatively objective. On the other hand, relatively subjective audit tasks require the auditors to exercise their professional judgment. For example, when auditing a client's lease classification, auditors required to use their judgment to evaluate the classification of a lease whose terms do not clearly lead to one classification decision over another is a relatively subjective task.

A recent study by Wright and Wu (2018) finds that audit task subjectivity may affect jurors' evaluations of auditor negligence through the effect of such subjectivity on jurors' attributions. In particular, they find evidence that suggests jurors recognize the difficulty associated with completing subjective audit tasks (e.g., auditing complex estimates) which, in turn, increases the likelihood that jurors attribute the audit failure to external causes rather than the auditor as the subjectivity inherent in the audit task increases. What is unclear from this prior work, though, is whether the subjectivity of the audit task affects how jurors react to auditors' *use of technology* when completing those tasks. This is an important question given audit firms' large investments in advanced audit technologies capable of performing tasks that are objective, as well as those that are relatively more subjective.

Objective Tasks

Prior research notes that trust in and use of algorithms depends on the kinds of abilities individuals typically believe that algorithms possess (Castelo et al. 2019). Starting with relatively objective tasks, we posit that jurors likely view technology as a suitable replacement for humans in tasks that require minimal to no human judgment. Objective audit tasks may include gathering and aggregating data, matching data based on a set of specific criteria, classifying data based on a clear set of decision rules, etc. Technology-based audit tools can easily be programmed to perform these objective tasks, enabling them to complete such tasks as effectively as human auditors with the added benefits of increased efficiency.

For objective audit tasks, jurors likely view technology as a suitable substitute for humans given such tasks are logical and rules-based requiring minimal to no human judgment (e.g., counting inventory). Based on this strong fit, we predict that algorithm aversion, or the tendency for individuals to prefer the advice of humans over technology when making their judgments and decisions, is unlikely to exist when individuals are evaluating the work of others on a relatively objective task. This is because we do not expect jurors to believe that human auditors, compared to a technology-based audit tool, would have found it easier to detect a similar misstatement. Put another way, the strong fit will make it difficult for jurors to think about what human auditors could have done differently from technology (and vice versa) to detect the misstatement.

Consequently, we do not expect jurors' counterfactuals or jurors' evaluations of auditors to differ based on the inclusion of technology on a relatively objective audit task.

It is worth noting, though, that the intentionality of the material misstatement missed by either the auditor or the technology-based audit tool may vary. Both human auditors and technology may fail to detect an unintentional error or an intentional, fraudulent misstatement. While recent audit failures associated with undetected frauds (e.g., Lehman Brothers, Satyam, Luckin Coffee) receive a great deal of media attention, restatements are almost three times as likely to occur due to unintentional errors (Hennes et al. 2008).[2] However, we do not expect the intentionality of the misstatement to interact with auditors' approaches to testing the area in question given that it makes intuitive sense to utilize technology for relatively objective audit tasks. This leads to the following formal hypothesis:

> **H1: When the audit task is relatively objective, auditors' use of technology will not significantly affect jurors' evaluations of auditor negligence regardless of the intentionality of the misstatement.**

Subjective Tasks

In contrast to objective audit tasks, more subjective audit tasks are those that require relatively more professional judgment. For example, determining whether assumptions underlying a complex estimate are reasonable or ensuring that the classification of a lease captures the underlying economic substance of the transaction are both relatively subjective tasks. Such audit tasks tend to have guidance or criteria that is <u>open to interpretation</u>, requiring a higher-level of professional judgment than a relatively objective task.

[2] Although not the focus of our study (and thus not formally stated as a hypothesis), it is reasonable to expect a main effect of misstatement intentionality. Given the deliberate nature of frauds combined with efforts taken to conceal them, we expect that jurors have an appreciation for the fact that frauds occur less frequently and are often more difficult to detect than unintentional errors. Thus, jurors are likely to hold auditors more accountable and, thus, be more likely to find auditors negligent for failing to detect an error versus fraud.

A growing body of research studying algorithm aversion finds that both audit and non-audit decision makers are averse to using algorithms in highly subjective settings, opting instead to rely on potentially less accurate human judgments (Castelo et al. 2019; Yeomans et al. 2019; Commerford et al. 2022). Research has shown that lay people see subjective tasks as requiring intuition and "gut instincts" (Inbar, Cone, and Gilovich 2010). Thus, this aversion can be attributed to decision makers' beliefs that algorithmic information sources lack the necessary cognitive capabilities to conduct the subjective task (i.e., poor fit). This suggests that in the auditor negligence setting, jurors will believe that human auditors would have found it easier to detect a similar misstatement compared to the technology-based audit tool utilized. This, in turn, leads us to believe that it will be easier for jurors to engage in counterfactual thinking about what could have been done differently to detect the misstatement when the audit firm relied on a technology-based audit tool (i.e., they could or should have used human judgement) rather than a human auditor (i.e., they could or should have used technology).

Importantly, though, we expect that the intentionality of the misstatement will moderate this predicted effect whereby the increase in assessed negligence associated with auditors' use of technology will be greater when they fail to detect a material misstatement due to fraud rather than an error. This expectation relies on the idea that technology is not generally equipped to reason, much less detect cues of deception indicative of potential fraud, thereby further worsening the perceived fit. Individuals tend to believe that technology lacks non-cognitive skills (Gray et al. 2007; Haslam et al. 2008). Understanding non-verbal behavioral cues and emotion, as well as detecting context-based verbal cues would seem to fall into this category and could be perceived by jurors as critical to detecting deceptive fraudulent behavior. Consequently, when the fraud occurs in an area for which the use of technology is already perceived to be a deviation from the norm, we expect jurors to experience more intense counterfactual thoughts when technology is relied upon rather than humans, and this effect to be stronger for frauds (vs. errors). We predict that the increase in litigation risk associated with auditors' use of technology will be greater when they fail to detect a material misstatement due to fraud rather than an error. Stated formally, we predict the following:

> **H2: When the audit task is relatively subjective, the increase in jurors' evaluations of auditor negligence associated with auditor use of technology will be greater when they fail to detect a material misstatement due to fraud rather than an error.**

4. Experiment 1

Method

In Experiment 1, we examine jurors' assessments of auditor negligence in a setting where task-technology fit is *high* and consensus is held constant. Specifically, we examine how auditors' use of drone technology to aid in the count of observable gravel inventory affects jurors' judgments when the undetected misstatement is either intentional or not. We chose this setting as drones are commonly used in practice to facilitate inventory counts (EY 2021) and the audit task of counting inventory is a relatively objective task.

Participants

Following prior litigation research, we recruited 280 jury-eligible participants (above the age of 18 and citizens of the United States) from Amazon Mechanical Turk (cf. Grenier et al. 2018). Mechanical Turk workers represent a more diverse cross-section of the jury-eligible population than do undergraduate students (Grenier et al. 2015; Maksymov and Nelson 2017), have been found to be reliable (Burhmester et al. 2011; Farrell et al. 2017), and have been commonly used as proxies for jurors (Grenier et al. 2015; Brasel et al. 2016; Maksymov and Nelson 2017). Consistent with prior research, we excluded participants who previously served on financial crime juries, worked as either accountants or lawyers, or worked in a field similar to the audit client in the case.

To help ensure data quality as well as participant understanding of a technical accounting lawsuit, we embedded 18 review questions in the case information. We required participants to maintain an 80-percent accuracy rate as they answered the review questions. The online instrument automatically removed participants from the survey who fell below the required accuracy rate. Participants received $2.00 compensation for completing the instrument. We did not note significant differences in any demographic variables across conditions.

Case

We adapted Kadous' (2000, 2001) experimental case to our setting and manipulations of interest. Participants first read general information about the audit process to familiarize themselves with basic auditing terms and concepts, emulating court efforts to educate jurors (Peecher and Piercey 2008). They learned about what auditors do and the importance of the audit function. They also read about material misstatements and the importance of discovering material misstatements during the audit process. Participants next read about the audit firm, Jones & Company, and its inventory valuation of gravel piles for a client, Big Time Gravel Company. In all conditions, participants learned that the firm audited five of Big Time's seven inventory sites. The auditors concluded that the inventory count of gravel was not materially misstated, and that Big Time's financial statements as a whole were not materially misstated. The plaintiff, an outside lender, relied on the firm's audit report in making its decision to lend funds to Big Time.

The participants then learned that, following an SEC investigation, it came to light that Big Time's gravel inventory was materially overstated. This overstatement hid the company's financial difficulties. Unfortunately, the audit firm's failure to detect the misstatement had severe consequences including the plaintiff's inability to collect funds related to its outstanding loan to

Big Time, employees of Big Time losing their jobs, and shareholders of the audit client suffering significant losses.

Independent variables

We first manipulated the intentionality of the misstatement at two levels. In the *fraud condition*, participants learned that Big Time intentionally distorted (i.e., increased) the apparent size of some of the gravel piles used in their inventory count calculation. In the *error condition*, the distortion was unintentional. Second, we manipulated the auditors' approach to testing the inventory count at two levels by indicating that either a human auditor (*human condition*) or a drone (*technology condition*) tested the gravel inventory balance. Further, all participants were told that the utilized approach was industry best practice in the calculation of the inventory. This enables us to hold consensus constant at a high level so we can focus on how our fit characteristics affect jurors' judgments. These manipulations resulted in a 2 x 2 between-participants design.

Dependent variables

As our primary dependent variable, we asked the participants how they would vote if the jury took a poll before deliberations. Participants responded either yes, Jones & Company was negligent, or no, Jones & Company was NOT negligent. For robustness and process testing, we also measure jurors' assessments of auditors' negligence on a continuous scale, with participants indicating the likelihood of auditor negligence on a scale of 1 ("Extremely unlikely") to 5 ("Extremely likely").

Post-experimental questionnaire and demographic questions

Following the dependent variables, we asked participants a series of questions to better understand jurors' decision-making process. Consistent with Reffett (2010), we asked participants to rate how easy it was to think about what the audit firm could have done differently to detect the material misstatement (1 = Extremely easy, 5 = Extremely difficult). Their responses serve as our measure of the intensity of jurors' counterfactual thoughts, with lower responses on the scale indicating more intense counterfactual thoughts. We further asked participants to rate how difficult they believed it would be for an auditor to detect a similar misstatement (1 = Extremely easy, 5 = Extremely difficult). This measure helps us assess their perceptions of the fit between the use of a technology-based tool (or human) and the audit task (i.e., relatively more objective or subjective, with lower responses on this scale indicating a poorer fit between the task and the utilized audit approach. Participants also rated the competence demonstrated by the audit firm (1 = Extremely incompetent, 5 = Extremely competent), as well as their agreement with the statement that the audit process used was reliable (1 = Strongly disagree, 5 = Strongly agree). The survey concluded with participants responding to basic demographic questions.

Results

Manipulation checks

We included a manipulation check question for each of our independent variables. We first asked participants whether Jones & Company (the audit firm) used human auditors or a drone to estimate Big Time's gravel inventory. Only five of the 280 participants (0.02 percent) answered this question incorrectly. We next asked whether Big Time (the audit client) intentionally or unintentionally distorted the apparent size of the gravel piles. Sixty-two participants (22 percent) answered this question incorrectly. As another gauge of whether our participants attended to our intentionality manipulation, we also asked whether, in the participants' opinions, not informing the audit firm about the material information was an error or a fraud. Participants responded on a 5-point Likert scale (1 = Definitely an error, 5 = Definitely a fraud). The responses are directionally consistent with our manipulations ($mean_{error}$ = 2.73 versus $mean_{fraud}$ = 4.29) and significantly different (t_{278} = -10.97, two-tailed *p*-value < 0.001), providing support that the participants successfully attended to our intentionality manipulation. Exclusion of participants failing at least one of the manipulation checks does not affect our inferences. Thus, our tests include responses from all 280 participants.

Test of Hypothesis 1

Panel A of Table 1 reports the percentage of jurors finding the audit firm negligent, Panel B reports the results of the general linear model with a logit link for the verdict decision, and Panel C reports the simple main effects. Hypothesis 1 predicts that an auditor's litigation risk will not be affected by technology when the audit task is relatively objective.

<INSERT TABLE 1 HERE>

To test this hypothesis, we regressed our experimental conditions on the negligent verdict dependent variable. As shown in Panel A of Table 1, jurors' verdict decisions do not depend on auditors' use of technology-based audit tools as part of their audit approach for a relatively objective audit task ($mean_{technology}$ = 41.48 percent vs $mean_{human}$ = 36.55 percent, $\chi^2 = 0.685$, p = 0.408).[3] Also, consistent with expectations, the interaction of *Audit Approach* and *Intentionality* is not significant ($\chi^2 = 0.074$, p = 0.785). Thus, we find support for Hypothesis 1 suggesting that jurors do not penalize auditors more for an audit failure following the use of technology when the task in question is relatively objective, regardless of misstatement intentionality.[4]

Counterfactual intensity and fit

---

[3] All p-values are two-tailed unless otherwise noted (i.e., for directional predictions).

[4] In untabulated results, we also estimate an ANOVA model of jurors' negligence likelihood assessments. We again find that jurors view auditors utilizing drones as having a higher negligence likelihood than those not utilizing such technology as part of their audit approach, but that difference is not significant ($F_{1,276}$ = 2.410, two-tailed p=0.122). While we report the results of this variable, it is important to note that negligence likelihood assessments capture jurors' beliefs about auditor negligence, but it does not take into account their tolerance for negligence. As such, the same raw likelihood assessment may translate to different binary verdict decisions for two different jurors. Thus, we focus on the binary verdict decision consistent with prior research (cf. Grenier et al. 2018).

The predictions for our hypotheses centered around counterfactual intensity (or lack thereof in the case of Hypothesis 1) resulting from perceptions of the fit between the audit task and the technology used. Our primary measure of counterfactual intensity comes from responses to the question, "When considering the material misstatement described in this case, how easy is it for you to think about what the audit firm could have done differently to detect the material misstatement?" Participants responded to this question on a five-point Likert scale anchored at 1 (Extremely easy) and 5 (Extremely difficult). As such, lower-numbered responses indicate higher levels of counterfactual intensity.

Table 2 displays the mean counterfactual intensity for each of our four experimental conditions. We ran an ANOVA to test for differences in counterfactual intensity and, seemingly counter to our theory, found a significant main effect for audit approach ($F = 4.884$, $p = 0.028$, untabulated). Counterfactual intensity is significantly higher in the technology conditions (mean = 3.12) than in the human conditions (mean = 3.42, p-value for difference 0.027). We test for the impact of differences in counterfactual intensity further in the next section.

5. Experiment 2

Method

In Experiment 2, we examine jurors' assessments of auditor negligence in a setting where task-technology fit is *low*. Specifically, we examine how auditors' use of artificial intelligence (AI) technology to aid in classification of leases affects jurors' judgments when the undetected misstatement is either intentional or not. We chose this setting as AI is commonly used in auditing lease accounting (EY 2017) and the audit task of classifying leases can require significant judgment.

Participants

We recruited 196 jury-eligible participants (above the age of 18 and citizens of the United States) from Amazon Mechanical Turk. Everything concerning participant inclusion and participation was identical to Experiment 1, including the series of comprehension check questions that required an 80 percent pass rate.

Case

Participants in Experiment 2 received basic information about the audit process. The major difference between Experiment 1 and 2 is the audit task. In this second experiment, the audit failure occurred in the estimation of the lease liability, which is more subjective in nature than the relatively objective task found in Experiment 1. In this case, the plaintiff alleges auditor negligence during the audit of LocalPharm, a large drugstore chain located primarily in the Northeast. All participants are told that LocalPharm sought to expand into the Midwest and needed to borrow money in order to facilitate the expansion. LocalPharm typically rents already existing buildings to use for its retail stores rather than build new ones, resulting in leases representing a large portion (about 70%) of the company's total liabilities. Under the accounting rules, the liability associated with any one lease increases with the length of that lease. However, participants are told that the determination of the length of a given lease can require significant judgment when there is the option to renew that lease because one must determine whether or not to include the renewal term in the overall term of the lease.

Our participants were told that, according to the SEC, it became clear in hindsight that LocalPharm understated its lease liability by $12 million. This understatement occurred because the renewal option embedded in the contract was not a clear extension of the length of the contract and, thus, not included in the calculation of the liability. The understatement hid the company's financial difficulties from the plaintiff. Similar to Experiment 1, the audit firm's failure to detect the misstatement had severe consequences, including the plaintiff's inability to collect funds related to its outstanding loan to LocalPharm.

Independent variables

We first manipulated whether LocalPharm intentionally (*fraud condition*) or unintentionally (*error condition*) understated its lease liability. The manipulation centered around whether LocalPharm intentionally or unintentionally omitted information about the agreement. Participants in the fraud (error) condition were told that the "understatement was the result of [LocalPharm's] management intentionally (unintentionally) structuring the new lease contracts so that the renewal option embedded in the contract was not a clear extension of the length of the contract." Second, we manipulated whether a human auditor (*human condition*) or an artificial intelligence program (*technology condition*) reviewed information about the leases and determined the reasonableness of LocalPharm's reported lease liability. Like Experiment 1, all participants were told that the utilized approach for estimating the lease liability was industry best practice in the calculation of the inventory which helps us to hold consensus constant. These manipulations resulted in a 2 x 2 between-participants experimental design.

Dependent and process variables

Identical to Experiment 1, we gathered participants' binary verdicts by asking them how they would vote if the jury took a poll before deliberations (0 = No, Jones & Company was NOT negligent, 1 = Yes, Jones & Company was negligent). Participants also assessed the audit firm's negligence likelihood (1 = Extremely unlikely, 5 = Extremely likely). Consistent with Experiment 1, we also measured the intensity of jurors' counterfactual thoughts, how difficult they believed it would be for an auditor to detect a similar misstatement, and other process-related measures. The survey concluded with participants responding to basic demographic questions.

Results

Manipulation Checks

We included a manipulation check question for each of our independent variables. We first asked participants whether Jones & Company (the audit firm) used human auditors or an artificial intelligence program to read the leases, extract relevant terms from the leases, classify the leases, and compute the lease liabilities. Only two of the 196 participants (1.0 percent) answered this question incorrectly. We next asked whether LocalPharm (the audit client) intentionally or unintentionally neglected to tell the audit firm about information material to the lease classification and computation decision. Twenty-seven of the 196 participants (13.78 percent) answered this question incorrectly. To provide further support that our participants attended to our intentionality manipulation, we also asked jurors to classify whether management's failure to inform the audit firm about the lease renewal agreements was an error or a fraud on a scale from 1 ("Definitely an error") to 5 ("Definitely a fraud"). The responses are directionally consistent with our manipulations ($mean_{error} = 2.49$ versus $mean_{fraud} = 4.47$) and significantly different ($t_{194} = -12.650$, $p < 0.001$). Thus, our tests described below include responses from all 196 participants.[5]

Test of Hypothesis 2

To test our second hypothesis, we examine the effect of our variables of interest on jurors' binary negligence verdicts. Panel A of Table 2 reports the percentage of jurors finding the audit firm negligent, Panel B reports the general linear model with a logit link for the verdict decision, and Panel C reports the simple main effects.

<INSERT TABLE 2 HERE>

[5] We reran the tests described in the next section with the manipulation check failures removed. All reported results are the same with the exception that the significant interaction described falls to marginal significance.

Hypothesis 2 predicts that, for relatively subjective audit tasks, the increase in litigation risk when using technology over human auditors will be greater when the firm fails to detect a material misstatement due to fraud rather than an error. As an initial test of this difference-in-differences hypothesis, we regressed our experimental conditions on negligence judgments. As shown in Panel B of Table 2, we find a statistically significant interaction ($\chi^2 = 5.672$, two-tailed $p = 0.017$). Further analysis of the simple effects offers support for our hypothesis. In particular, we find that jurors penalize auditors for utilizing technology-based audit tools as part of the audit of a subjective audit task when the subsequently detected misstatement was intentionally made ($mean_{human/fraud} = 34.1$ percent vs $mean_{technology/fraud} = 55.4$ percent, $\chi^2 = 4.486$, $p = 0.034$). The use of technology has no such effect on jurors' verdicts when the misstatement was unintentional ($mean_{human/error} = 46.3$ percent vs $mean_{technology/error} = 33.3$ percent, $\chi^2 = 1.646$, $p = 0.200$). We find similar results using jurors' assessments of auditors' negligence likelihood whereby auditors are more likely to be found negligent when they utilize a technology-based audit tool as part of the approach to auditing a relatively subjective audit task and subsequently fail to detect an intentional misstatement ($mean_{human/fraud} = 2.45$ vs. $mean_{technology/fraud} = 3.16$, $t_{98} = 2.476$, $p = 0.015$), but not when that misstatement is unintentional ($mean_{human/error} = 2.74$ vs. $mean_{technology/error} = 2.55$, $t_{94} = 0.746$, $p = 0.457$). Collectively, these two tests provide support for Hypothesis 2, suggesting that auditors' use of technology-based audit tools are more likely to increase perceptions of negligence when management scienter is present.

Counterfactual intensity and fit

Recall that the predictions for our hypotheses centered around counterfactual intensity resulting from perceptions of the fit between the audit task and the technology used. We asked the same question about counterfactual intensity that we did in Experiment 1: "When considering the material misstatement described in this case, how easy is it for you to think about what the audit firm could have done differently to detect the material misstatement?" Participants responded to this question on a five-point Likert scale anchored at 1 (Extremely easy) and 5 (Extremely difficult). As such, lower-numbered responses indicate higher levels of counterfactual intensity.

Table 4 displays the mean counterfactual intensity for each of our four experimental conditions in Experiment 2. We used ANOVA to test for differences in counterfactual intensity across conditions and found a significant interaction between audit approach and misstatement intentionality ($F = 4.238$, $p = 0.041$, untabulated). The highest counterfactual intensity came in the technology/fraud condition while the lowest appears in the human/fraud condition.

Part of our prediction in Hypothesis 2 is that the impact of auditor use of technology in a subjective audit task will be higher when the misstatement arises from fraud relative to arising from an error. We ran a mediation analysis to test whether the difference in counterfactual intensity affected participant judgment. Figure 2 represents the outcome of the analysis. As shown, counterfactual intensity mediates the effect of audit approach on participant judgment of negligence in the fraud conditions.[6]

Recall that in Experiment 1 we found a significant difference in counterfactual intensity between audit approach conditions. For sake of comparison, we ran the same mediation analysis described above using the data from Experiment 1. As shown in Figure 3, counterfactual intensity does not mediate the effect of audit approach on participant judgment of negligence for the objective audit task like it did for the subjective audit task. However, we do find a mediation effect when we check for mediation in the *error* conditions, providing initial evidence that the effect of counterfactual intensity varies based on misstatement intentionality.

---

[6] Figure 2 shows the outcome of the analysis using our negligence verdict dependent variable. We find similar results using the likelihood of negligence dependent variable.

6. Conclusion

This study elucidates the conditions under which using advanced technology in audit tasks affects perceptions of auditor liability through development of a theoretical framework and applying the framework to two important factors: the subjectivity of the audit task and the intentionality of the undetected material misstatement. Our theoretical framework melds psychological theories on attribution and counterfactual thinking to predict the conditions under which algorithm aversion applies to the retrospective evaluation of others (e.g., jurors' evaluations of auditor negligence). Leveraging this framework, we predict and find that technology reliance will have no effect on liability for objective tasks. On the other hand, we predict and find an escalation of liability for subjective tasks, particularly when management fraud is the cause of the misstatement rather than management error. In sum, reliance on advanced technologies only elevated auditor liability when detecting the misstatement involved professional judgment.

Our study has implications for both practice and research. For practice, our study will help practitioners assess the audit risk implications of adopting new technologies. Although the effectiveness and efficiency benefits of such technologies likely far outweigh potential litigation costs, it is still important for firms to understand their litigation exposure for effective risk management and audit pricing. To that end, firms could use our study to better assess audit risk at the engagement-level based on the presence or absence of various contextual factors. Our study should also give firms a degree of comfort that technologies will not pervasively escalate liability for undetected misstatements as we only documented escalation under certain important, but relatively rare, conditions (i.e., subjective audit tasks failing to detect fraud). Although we urge caution in over-relying on this result in terms of assessing the overall effect of technology on liability, firms may be able to use our empirical evidence to lower their professional malpractice premiums. For research, we both reconcile seemingly conflicting results in prior studies, and hope to stimulate future research examining other relevant factors using our theoretical framework as a guide.

Our study is subject to limitations beyond those typically associated with experimental research. First, although we examine an important aspect of liability risk, jurors' negligence judgments, we cannot speak to the overall effect on liability. Most notably, advanced technologies may reduce the incidence rate of undetected misstatements (and the resulting litigation) to an extent that outweighs the increase in perceptions of negligence for the remaining cases. Although firms extol the effectiveness benefits of technology (EY 2017), future research should attempt to empirically document the corresponding reduction in audit failures.

Second, as we have often mentioned, our study should not be considered definitive evidence on the overall effect of technology on auditor liability, nor that our independent variables are completely determinative. There are almost certainly other important contextual factors that would moderate our observed effects. For instance, the manner in which defense teams present and defend audit procedures including the use of technology are likely to influence jurors' perceptions of quality and ultimate judgments. Future research should examine such defense tactics.

Third, we may not obtain the same pattern of results if we examined different technologies than drone technologies for objective tasks and artificial intelligence for subjective tasks. Although we felt these technologies were best matched with the audit task of interest and high in consensus, it is possible that other technologies or future improvements of the technologies we examine will be evaluated differently by jurors. Finally, it is possible that the algorithm aversion that we document in juror judgments may fade with time as these technologies improve and/or become more commonly understood by lay evaluators, perhaps turning into algorithm appreciation. In that case, auditors may actually be rewarded (rather than punished) in terms of auditor liability for using advanced technologies.

Finally, we held consensus constant in our initial experiments because the results of Kadous and Mercer (2012) gave us a likely starting point for consensus in our theoretical framework. Future research can test the effect of varying degrees of consensus on judgments of auditor use of technology in audit tasks. Studies can also examine different determinants of consensus.

References

American Institute of CPAs (AICPA). 2020. *The Data-Driven Audit: How Automation and AI are Changing the Audit and the Role of the Auditor*. Available at: https://www.aicpa.org/content/dam/aicpa/interestareas/frc/assuranceadvisoryservices/downloadabledocuments/5617589-02484-he-data-driven-audit.pdf.

Backof, A. G. 2015. The impact of audit evidence documentation on jurors' negligence verdicts and damage awards. *The Accounting Review* 90(6): 2177-2204.

Ballou, B., Grenier, J.H., and A. Reffett. 2021. Stakeholder perceptions of data and analytics based auditing techniques. *Accounting Horizons* 35(3): 47-68.

Barr-Pulliam, D., Brown-Liburd, H., and K. Sanderson. 2022. The effect of the internal control opinion and use of audit data analytics on perceptions of audit quality, assurance, and auditor negligence. *Auditing: A Journal of Practice & Theory* 41(1): 25-48.

Bonner, S. E. 1999. Judgment and decision-making research in accounting. *Accounting Horizons* 13(4): 385-398.

Brasel, K., Doxey, M.M., Grenier, J.H., and A. Reffett. 2016. Risk disclosure preceding negative outcomes: The effect of reporting critical audit matters on judgments of auditor liability. *The Accounting Review* 91 (5): 1345-1362.

Buhrmester, M., Kwang, T., and S. D. Gosling. 2011. Amazon's Mechanical Turk: A new source of inexpensive, yet high-quality, data? *Perspectives on Psychological Science* 6(1): 3-5.

Byrne, B. M. 2010. *Structural Equation Modeling with AMOS: Basic Concepts, Applications, and Programming*. edited by L. Harlow. 2 ed. New York, NY: Routledge.

Castelo, N., Bos, M.W., and D. R. Lehmann. 2019. Task-dependent algorithm aversion. *Journal of Marketing Research* 56(5): 809-825.

Commerford, B.P., Dennis, S.A., Joe, J.R., and J. Ulla. 2022. Man versus machine: Complex estimates and auditor reliance on artificial intelligence. *Journal of Accounting Research* 60(1): 171-201.

Cui, J. 2021. The effects of the use of natural language processing and task complexity on juror's assessments of auditor negligence. Dissertation, University of North Texas.

Dietvorst, B. J., Simmons, J.P., and C. Massey. 2018. Overcoming algorithm aversion: People will use imperfect algorithms if they can (even slightly) modify them. *Management Science* 64(3): 1155-1170.

Eastwood, J., Snook, B., and K. Luther. 2012. What people want from their professionals: Attitudes toward decision-making strategies. *Journal of Behavioral Decision Making* 25(5): 458-468.

Ernst & Young (EY). 2017. *How AI will Enable Us to Work Smarter* (July 7). Available at: https://www.ey.com/en_gl/assurance/how-ai-will-enable-us-to-work-smarter-faster.

Ernst & Young (EY). 2021. EY institutes drone use in manufacturing and retail warehouse audits. Available at: https://interdrone.com/news/ey-institutes-drone-use-in-manufacturing-and-retail-warehouse-audits/

Farrell, A. M., Grenier, J.H., and J. Leiby. 2017. Scoundrels or stars? Theory and evidence on the quality of workers in online labor markets. *The Accounting Review* 92(1): 93-114.

Fildes, R. and P. Goodwin. 2007. Against your better judgment? How organizations can improve their use of management judgment in forecasting. *Interfaces* 37(6): 570-576.

Forbes Insights. 2018. Audit 2025: The future is now. Available at: https://i.forbesimg.com/forbesinsights/kpmg_audit2025/KPMG_Audit_2025.pdf

Goodhue, D. L. and R. L. Thompson. 1995. Task-technology fit and individual performance. *MIS Quarterly* 19(2): 213-236

Gray, H.M., Gray, K., and D. M. Wegner. 2007. Dimensions of mind perception. *Science* 315(5812): 619.

Grenier, J. H., Reffett, A., Simon, C.A., and R. C. Warne. 2018. Researching juror judgment and decision making in cases of alleged auditor negligence: A toolkit for new scholars. *Behavioral Research in Accounting* 30 (1): 99-110.

Grenier, J. H., Holman, B.A., Lowe, D.J., and J. W. Ulla. 2021. The ticking time bomb: Population testing and jurors' assessments of auditor negligence. Working paper, Miami University.

Grenier, J. H., Lowe, D.J., Reffett, A. and R. Warne. 2015. The effects of independent expert recommendations on juror judgments of auditor negligence. *Auditing: A Journal of Practice & Theory* 34(4): 157-170.

Harris, S. B. 2017. Technology and the audit of today and tomorrow. Available at: https://pcaobus.org/news-events/speeches/speech-detail/technology-and-the-audit-of-today-and-tomorrow_644.

Haslam, N., Kashima, Y., Loughnan, S., Shi, J., and C. Suitner. 2008. Subhuman, inhuman, and superhuman: Contrasting humans with nonhumans in three cultures. *Social Cognition* 26(2): 248-58.

Hennes, K. M., Leone, A.J., and B. P. Miller. 2008. The importance of distinguishing errors from irregularities in restatement research: The case of restatements and CEO/CFO turnover. *The Accounting Review* 83(6): 1487-1519.

Inbar, Y., Cone, J., and T. Gilovich. 2010. People's intuitions about intuitive insight and intuitive choice. *Journal of Personality and Social Psychology* 99(2): 232-247.

Kadous, K. 2000. The effects of audit quality and consequence severity on juror evaluations of auditor responsibility for plaintiff losses. *The Accounting Review* 75(3): 327-341.

Kadous, K. 2001. Improving jurors' evaluations of auditors in negligence cases. *Contemporary Accounting Research* 18(3): 425-444.

Kadous, K. and M. Mercer. 2012. Can reporting norms create a safe harbor? Jury verdicts against auditors under precise and imprecise accounting standards. *The Accounting Review* 87(2): 565-587.

Kahneman, D. and D. T. Miller. 1986. Norm theory: Comparing reality to its alternatives. *Psychological Review* 93(2): 136-153.

Kapoor, M. 2020. Big four invest billions in tech, reshaping their identities. *Bloomberg*. Available at https://news.bloombergtax.com/financial-accounting/big-four-invest-billions-in-tech-reshaping-their-identities.

Kipp, P., Olvera, R., Robertson, J.C., and J. Vinson. 2022. Audit data analytics and jurors' assessment of auditor negligence: The effects of follow-up procedures and the lack of a standard. Working paper, University of North Texas.

KPMG. 2021. Available at: https://audit.kpmg.us/emerging-technologies.html.

Longoni, C., Bonezzi, A., and C. K. Morewedge. 2019. Resistance to medical artificial intelligence. *Journal of Consumer Research* 46(4): 629-650.

Maksymov, E. M., and M. W. Nelson. 2017. Malleable standards of care required by jurors when assessing auditor negligence. *The Accounting Review* 92(1): 185-202.

Munoko, I., Brown-Liburd, H.L. and M. Vasarhelyi. 2020. The ethical implications of using artificial intelligence in auditing. *Journal of Business Ethics* 167: 209-234.

Nelson, M. and H. T. Tan. 2005. Judgment and decision making research in auditing: A task, person, and interpersonal interaction perspective. *Auditing: A Journal of Practice & Theory* 24(s-1): 41-71.

Önkal, D., Goodwin, P., Thomson, M., Gönül, S., and A. Pollock. 2009. The relative influence of advice from human experts and statistical methods on forecast adjustments. *Journal of Behavioral Decision Making* 22(4): 390-409.

Peecher, M. E. , and M. D. Piercey. 2008. Judging audit quality in light of adverse outcomes: Evidence of outcome bias and reverse outcome bias. *Contemporary Accounting Research* 25(1): 243-274.

Promberger, M., and J. Baron. 2006. Do patients trust computers? *Journal of Behavioral Decision Making* 19(5): 455-468.

Ranzilla, S., Chevalier, R., Herrmann, G., Glover, S. and D. Prawitt. 2011. Elevating professional judgment in auditing and accounting: The KPMG professional judgment framework. Montvale, NJ: KPMG.

Reffett, A. B. 2010. Can identifying and investigating fraud risks increase auditors' liability? *The Accounting Review* 85(6): 2145-2167.

Wright, A. M., and Y. J. Wu. 2018. The impact of auditor task difficulty and audit quality indicators on jurors' assessments of audit quality. *Behavioral Research in Accounting* 30(2): 109-125.

Yeomans, M., Shah, A., Mullainathan, S., and J. Kleinberg. 2019. Making sense of recommendations. *Journal of Behavioral Decision Making* 32: 403-414.

Zeisel, H., and S. Diamond. 1976. The jury selection in the Mitchell-Stans trial. *American Bar Foundation Research Journal* 1: 151-174.

FIGURE 1
FRAMEWORK FOR TECHNOLOGY USE ON AUDITOR NEGLIGENCE

<table>
<tr><td colspan="4">Step 1: Identify independent factors of interest</td></tr>
<tr><td>Auditor Characteristics</td><td>Task Characteristics</td><td>Environment Characteristics</td><td>Technology Characteristics</td></tr>
<tr><td><u>Experience</u>: General, industry, technology expertise

<u>Firm</u>: Size, resources, specialization, independence, PCAOB Inspection history</td><td>Complexity
Subjectivity
Timing (interim vs. final)
Type (analytical, control, or substantive procedure)
Risk of Material Misstatement/Weakness
Fraud Risk</td><td><u>Client</u>: industry, technological sophistication, historical and current financial/non-financial performance, prior SEC infractions

<u>Country-level</u>: political, regulatory, and legal factors

<u>Economy-wide</u>: market performance, interest rates, growth, inflation.</td><td>Reliability
Complexity
Age
Developer (internal vs. external)
Reputation
Cost</td></tr>
<tr><td colspan="4">Step 2: Assess the factors with respect to the technology liability determinants</td></tr>
<tr><td>Technology Liability Determinants</td><td colspan="2"><u>Consensus</u>:
Do other auditors commonly use the technology for the audit task?

Consensus is inversely related to negligence with an increase (decrease) in the perception that the technology is frequently used by other auditors decreasing (increasing) perceptions of auditor negligence when using the technology.</td><td><u>Fit</u>:
How well suited is the technology for the task?

Fit is inversely related to negligence with an increase (decrease) in perceptions that the technology is well-suited for the task decreasing (increasing) perceptions of auditor negligence when using the technology.</td></tr>
<tr><td colspan="4">Step 3: Choose theoretical lens based on interplay with technology liability determinants</td></tr>
<tr><td>Predominant Theoretical Lenses in Auditor Liability Research</td><td>Motivated Reasoning (Kadous 2000, 2001)

To maintain their belief in a just world, jurors are motivated to blame auditors for undetected material misstatements and evaluate evidence in a biased manner consistent with that motivation, subject to reasonable constraints.</td><td>Culpable Control Model (Backof 2015)

Jurors’ attributions of blame are a function of their perceptions of auditors’ causal influence over the misstatement, the foreseeability of the misstatement, and auditors’ intention to perform a quality audit.</td><td>Counterfactual Thinking (Reffett 2010)

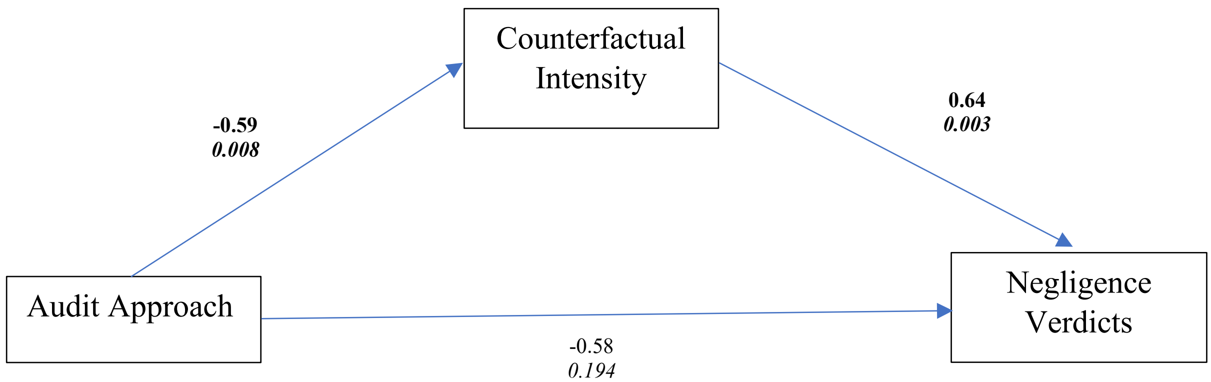
Jurors are prone to consider other ways that the auditors could have detected the undetected material misstatement. The intensity of counterfactual thinking (i.e., the ease of thinking about alternative course of action) is positively associated with auditor negligence.</td></tr>
<tr><td>Potential Linkages of Lens with Technology Liability Determinants</td><td><u>Consensus:</u>
• Inversely related to motivation to blame the auditor
• High consensus could impose reasonable constraints

<u>Fit:</u>
• Inversely related to motivation to blame the auditor
• Strong fit could impose reasonable constraints</td><td><u>Consensus:</u>
• Inversely related to causality
• Positively related to intention

<u>Fit:</u>
• Inversely related to causality and foreseeability
• Positively related to intention</td><td><u>Consensus:</u>
• Inversely related to intensity of counterfactual thinking

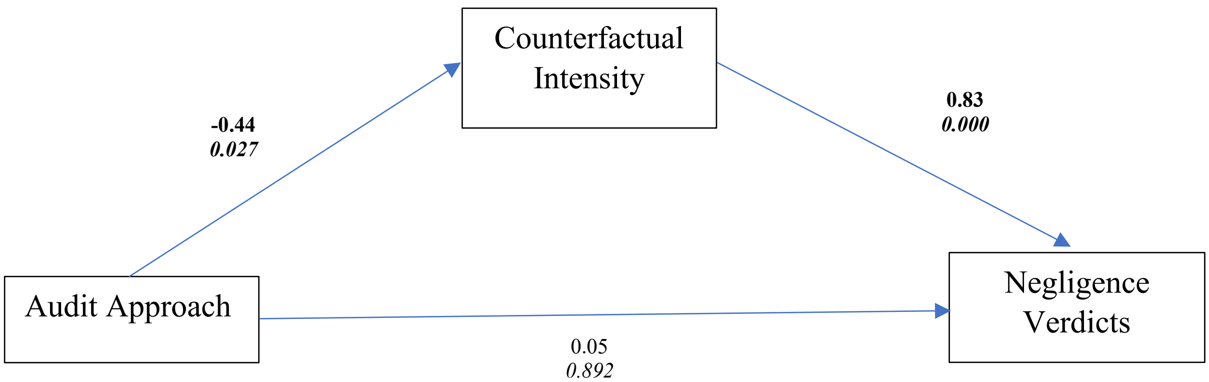
<u>Fit:</u>
• Inversely related to intensity of counterfactual thinking</td></tr>
<tr><td colspan="4">Step 4: Formulate Expectations</td></tr>
<tr><td colspan="4">Develop testable hypotheses based on how the factors of interest, individually and jointly, affect the technology conceptual liability determinants and correspondingly affect the components of the chosen theoretical lens.</td></tr>
</table>

FIGURE 2
AUDITORS' USE OF TECHNOLOGY-BASED AUDIT TOOLS IN A SUBJECTIVE AUDIT TASK – FRAUD CONDITIONS



This figure shows the results of a mediation analysis for the fraud conditions. T-stats and corresponding *p-values* (two-tailed) are shown on all links.
Audit Approach refers to whether the audit firm utilized artificial intelligence (=1) or not (=0) as part of the audit of the estimated lease liability.
Counterfactual Intensity is measured as participants' responses to the post-experimental question "When considering the material misstatement described in this case, how easy is it for you to think about what the audit firm could have done differently to detect the material misstatement?" (0 = Extremely easy, 5 = Extremely difficult).
Negligence Verdicts refers to participants' responses to "If the jury took a poll before deliberations, how would you vote? Would you vote that Jones & Company (i.e. the auditor) was or was not negligent in their performance of LocalPharm's audit?" (1 = Yes, Jones & Company was negligent, 2 = No, Jones & Company was NOT negligent).

FIGURE 3
AUDITORS' USE OF TECHNOLOGY-BASED AUDIT TOOLS IN AN OBJECTIVE AUDIT TASK – ERROR CONDITIONS



This figure shows the results of a mediation analysis for the fraud conditions. T-stats and corresponding *p-values* (two-tailed) are shown on all links.

Audit Approach refers to whether the audit firm utilized a drone (=1) or not (=0) as part of the audit of the gravel inventory.

Counterfactual Intensity is measured as participants' responses to the post-experimental question "When considering the material misstatement described in this case, how easy is it for you to think about what the audit firm could have done differently to detect the material misstatement?" (0 = Extremely easy, 5 = Extremely difficult).

Negligence Verdicts refers to participants' responses to "If the jury took a poll before deliberations, how would you vote? Would you vote that Jones & Company (i.e. the auditor) was or was not negligent in their performance of LocalPharm's audit?" (1 = Yes, Jones & Company was negligent, 2 = No, Jones & Company was NOT negligent).

TABLE 1
Negligence Verdicts – Experiment 1

Panel A: Percent of Jurors Finding Audit Firm Negligent: # of negligent verdicts, percent of negligent verdicts, [n]

	Fraud	Error	Overall
Human	25 34.72% [72]	28 38.36% [73]	53 36.55% [145]
Technology	25 37.88% [66]	31 44.93% [69]	56 41.48% [135]
Overall	50 36.23% [138]	59 41.55% [142]	

Panel B: General Linear Models (Logit Link, Binomial Distribution) for Juror Verdicts

Source of Variation	**df**	**χ^2**	**p-value**
Audit Approach	1	0.685	0.408
Intentionality	1	0.828	0.363
Audit Approach * Intentionality	1	0.074	0.785
Constant	1	13.439	<0.001

Panel C: Simple Main Effects of Technology vs. Human for Negligence Verdicts

	Contrast	**χ^2**	**p-value**
Error conditions	-0.07	0.633	0.426
Fraud conditions	-0.03	0.148	0.700

Audit Approach refers to whether the audit firm utilized a drone (= 1) or not (= 0) as part of the audit of the inventory count.
Intentionality refers to whether management intentionally (=1) or unintentionally (= 0) distorted the apparent size of the gravel piles, including piles at sites visited by the auditors.
Negligence Verdicts refers to participants' responses to "If the jury took a poll before deliberations, how would you vote? Would you vote that Jones & Company (i.e. the auditor) was or was not negligent in their performance of LocalPharm's audit?" (1 = Yes, Jones & Company was negligent, 0 = No, Jones & Company was NOT negligent).

TABLE 2
Counterfactual Thinking - Experiment 1
Means and Standard Deviations

	Error	Fraud	Overall
Human	3.49	3.35	3.42
	[1.156]	[1.140]	[1.147]
	n = 73	n = 72	n = 145
Technology	3.06	3.18	3.12
	[1.162]	[1.080]	[1.120]
	n = 69	n = 66	n = 135
Overall	3.28	3.27	
	[1.175]	[1.111]	
	n = 142	n = 138	

Counterfactual thinking is derived from the question, "When considering the material misstatement described in this case, how ***easy*** *is it for you to think about what the audit firm could have done differently to detect the material misstatement?" Responses can range from 1 - "Extremely easy" to 5 - "Extremely difficult". As such, lower numbers represent higher levels of counterfactual thinking.*

TABLE 3
Negligence Verdicts – Experiment 2

Panel A: Percent of Jurors Finding Audit Firm Negligent: # of negligent verdicts, percent of negligent verdicts, [n]

	Error	Fraud	Overall
Human	25	15	40
	46.30%	34.10%	40.82%
	[54]	[44]	[98]
Technology	14	31	45
	33.33%	55.36%	45.92%
	[42]	[56]	[98]
Overall	39	46	
	40.63%	46.00%	
	[96]	[100]	

Panel B: General Linear Models (Logit Link, Binomial Distribution) for Juror Verdicts

Source of Variation	df	χ^2	p-value
Audit Approach	1	0.306	0.580
Intentionality	1	0.445	0.505
Audit Approach * Intentionality	1	5.672	**0.017**
Constant	1	4.656	0.031

Panel C: Simple Main Effects of Technology vs. Human for Negligence Verdicts

	Contrast	χ^2	p-value
Error conditions	-0.13	1.646	0.200
Fraud conditions	-0.21	4.486	0.034

Audit Approach refers to whether the audit firm utilized artificial intelligence (= 1) or not (= 0) as part of the audit of the estimated lease liability.

Intentionality refers to whether management intentionally (=1) or unintentionally (= 0) structured the new lease contracts so that the renewal option embedded in the contract was not a clear extension of the length of the contract.

Negligence Verdicts refers to participants' responses to "If the jury took a poll before deliberations, how would you vote? Would you vote that Jones & Company (i.e. the auditor) was or was not negligent in their performance of LocalPharm's audit?" (1 = Yes, Jones & Company was negligent, 0 = No, Jones & Company was NOT negligent).

TABLE 4
Counterfactual Thinking - Experiment 2
Means and Standard Deviations

	Error	Fraud	Overall
Human	3.46	3.93	3.67
	[1.177]	[0.950]	[1.101]
	n = 54	n = 44	n = 98
Technology	3.52	3.34	3.42
	[1.042]	[1.180]	[1.121]
	n = 42	n = 56	n = 98
Overall	3.49	3.60	
	[1.114]	[1.119]	
	n = 96	n = 100	

Counterfactual thinking is derived from the question, "When considering the material misstatement described in this case, how ***easy*** *is it for you to think about what the audit firm could have done differently to detect the material misstatement?" Responses can range from 1 - "Extremely easy" to 5 - "Extremely difficult". As such, lower numbers represent higher levels of counterfactual thinking.*