“Auditor Deception Detection from Earnings Conference Calls: The Role of Auditor Experience and Management Cognitive Dissonance”

By

Jessen L. Hobson
Department of Accountancy, University of Illinois at Urbana-Champaign

William J. Mayew*
Fuqua School of Business, Duke University

Mark Peecher
Department of Accountancy, University of Illinois at Urbana-Champaign

Mohan Venkatachalam
Fuqua School of Business, Duke University

Thursday, June 4th, 2015
1:20 – 2:50 p.m.
Room C06

*Speaker

Paper Available in Room 447
Auditor Deception Detection from Earnings Conference Calls: The Role of Auditor Experience and Management Cognitive Dissonance

Jessen L. Hobson*
William J. Mayew†
Mark Peecher*
Mohan Venkatachalam†

*Department of Accountancy, University of Illinois at Urbana-Champaign
†Duke University – Fuqua School of Business

Preliminary. Please do not cite, distribute, or post without permission.

May 2015

We are grateful to the accounting professionals from various audit firms who participated in the study. We appreciate helpful comments from Spencer Anderson, Scott Asay, Clara Chen, Jon Davis, Brooke Elliott, Scott Emett, Kevin Jackson, Joseph Johnson, Tracie Majors, Terence Ng, Matt Pickard, Mike Ricci, Chad Simon, Jason Smith, Hun-Tong Tan, Mark Zimbelman, reviewers and participants at the Auditing Section Midyear Meetings, the Brigham Young University Accounting Symposium, the 2014 Accountancy Conference the University of São Paulo, the 2015 EDEN Doctoral Seminar in Audit Research (London), the Financial Accounting and Reporting Section Midyear Meeting, and the International Symposium on Audit Research, and workshop participants at Lehigh University, Nanyang Technological University, and the University of Illinois at Urbana-Champaign. We thank EB Altiero and Chris Calvin for their helpful research assistance.
We examine how extensive audit experience and a prompt to attend to the CEO’s cognitive dissonance individually and jointly influence auditors’ detection of financial deception from earnings conference calls. We predict and find that experienced auditors outperform both chance and inexperienced auditors, especially when prompted to attend to managers’ cognitive dissonance. This is encouraging, as meta-analyses from psychology find that experts generally outperform neither chance nor novices in detecting deception. Also as predicted but more worrisome, unprompted experienced auditors’ performance edge over novices arises predominantly from fewer false positives. While false positives about fraud could strain an auditor’s relationship with management, false negatives jeopardize audit effectiveness and increase the risk that financial statement users will suffer loss from fraud. In supplemental process analysis, we find that experience enables auditors to identify more, and more accurate, red flags in CEO’s narratives. This red flag advantage, when experienced auditors are prompted to attend management’s cognitive dissonance, translates into more accurate fraud detection. Finally, in exploring whether adding audio to written conference call transcripts improves accuracy, we observe it does so only for inexperienced auditors.
Auditor Deception Detection From Earnings Conference Calls: The Role of Auditor Experience and Management Cognitive Dissonance

1. Introduction

We use an experiment to examine whether and how extensive audit experience (over 20 years on average) and a prompt to attend to the CEO’s cognitive dissonance individually and jointly influence auditors’ detection of financial deception from CEO question and answer (Q&A) portions of earnings conference calls. We also explore whether the medium used to convey these Q&A portions, transcript alone or transcript plus audio, improves auditors’ detection accuracy.

Our experimental examination is important because detection of material financial statement fraud is a long-standing responsibility of public-company financial statement audits. Despite their responsibility to provide high assurance that management’s financial statements are free of material fraud, auditors seldom have been the first party to detect frauds, especially prior to the Sarbanes-Oxley Act (Dyck et al. [2010]). Further, there is concern that considerably more material financial statement fraud goes undetected than is ever detected (Dyck et al. [2013]). Investors, regulators, standard setters, and auditors all agree that it would be helpful if auditors were more effective in detecting fraud (e.g., Christensen, et al. [2015], Hogan et al. [2008]), particularly given recent increases in accounting-related class action lawsuits (Heller [2015]).

Our examination is also important for three, more specific reasons. One, it helps build theory about the determinants of auditors’ ability to detect fraud from relatively unscripted...
management narratives about their financial statements. We view Q&A portions of management’s earnings conference calls as a prominent example of relatively unscripted narratives. Other examples pervade audits, including management responses to formal and informal inquiries to help auditors gain an understanding of and test internal controls as well as to plan and conduct substantive tests (see, e.g., PCAOB AS No. 12, No. 13, and No. 15).

Second, this study of auditor deception detection contributes to the larger psychological literature on deception detection. Prior studies using participants ranging from experienced judges and police officers to lay persons and college students across numerous contexts only rarely attain above-chance performance in detecting deception, and in many instances experts are no better than non-experts (Bond and DePaulo [2006]; Bond and DePaulo [2008]; Vrij et al. [2006]). Unlike experts commonly studied in the extant literature, such as police officers, our “experts” have somewhat unique incentives in that they are detecting deception from a party (management) who can strongly influence their own compensation. Additionally, our potential deceivers (CEOs and CFOs) can be very experienced and even coached in communicating in a self-interested manner.

Third, this study a pragmatic benefit of improving our understanding of whether and how auditors can detect fraud using earnings conference calls, per se. Recent evidence suggests automated analysis of conference call speech can assist in detecting financial misreporting at better than chance levels (Larcker and Zakolyukina [2012], Hobson et al. [2012]). However, we know of no research examining whether and when auditors judgmentally can do so. Assessing auditor performance in this regard is important given current audit standards recommend reviewing earnings conference calls to help assess risk of misstatement (AS No. 12, PCAOB [2010]). AS 12 neither requires nor recommends that auditors use conference calls to help detect
warning signs of fraud, however, and the results of our study add to mounting evidence suggesting that audit standard setters may want to consider doing so.2

The theory we rely on warrants predicting that many years of experience will help auditors avoid false positives (i.e., predicting fraud when it is not present) but will help significantly less, if at all, in avoiding false negatives (i.e., predicting no fraud when it exists). An implication is that very experienced auditors will outperform inexperienced auditors in evaluating non-fraud companies, but that they will do so to a lesser extent when evaluating fraud companies. It also predicts that a prompt to attend to management’s cognitive dissonance will help very experienced auditors overcome this imbalance, enabling them to use knowledge acquired over many years to avoid false negatives.

These theory-based predictions are derived from prior psychology and auditing research. Over time, decision makers adaptively attend to and selectively interpret information so as to avoid committing what they experientially learn to be the “primary error” in their natural decision environments (Friedrich [1993]; Friedrich et al. [2005]). That is, rather than just learning to be accurate, decision makers–arguably quite rationally–learn to avoid more costly types of errors. In the language of motivated reasoning theory, decision makers adaptively learn to pursue directional goals. For auditors, commitment to directional goals biases them towards acceptance of management’s accounting methods, even more so when auditors are explicitly asked to objectively assess the quality of these methods (Kadous, et al. [2003]).

2 In contemplating new standards on how auditors should consider using conference calls to help assess the risk of misstatement due to fraud, standard setters may also want to consider the stream of research comparing accuracy rates of statistical linear models unaided by human judgment, statistical linear models of human judgment, and human judgments themselves (e.g., Dawes, et al. [1989]), but we are focused on unaided auditors’ professional judgments in this paper. Swets et al. [2000, 5] observe that, in complex professional judgment contexts (e.g., medicine), a “prevailing practice” is to supply output from statistical linear models to human decision makers who then make final judgment calls.
Evidence suggests experience increases auditor commitment to management-preferred directional goals, whereby inexperienced auditors (audit staff) are relatively aggressive skeptics while experienced auditors (audit managers and partners) are relatively reluctant skeptics in terms of beliefs and actions (Shaub and Lawrence [1999]). While university courses on auditing likely amplify inexperienced auditors’ skepticism by covering infamous frauds (e.g., Worldcom), even seasoned auditors rarely directly experience fraud. One recent archival study (Dyck et al. [2013]) finds that the rate of detected frauds, by any party, is less than 3% for public companies, even as it warns that detected frauds could be the “tip of the iceberg”.

By contrast, auditors regularly encounter fee pressure to not expand audit testing, jeopardizing audit quality (Ettredge, et al. [2014]). While post-SOX regulatory inspections may reduce fee pressure, Peecher et al. [2013], in reviewing auditor incentives, conclude that while auditors are penalized for failing to find fraud, they are only ambiguously (if at all) rewarded by regulators, management, or investors for actually detecting fraud. What is salient is that when auditors detect fraud, their expected litigation costs immediately increase, and they often lose that client (Doty [2014]). Thus, over many years of experience, false positives likely become auditors’ “primary error” to avoid, predisposing them to explain away (embrace) cues suggestive of fraud (no fraud), and reinforcing their susceptibility to false negatives.

We next examine a potential remedy to help experienced auditors overcome this learned behavior: prompting attention to management’s cognitive dissonance. Motivated reasoning theory holds that people pursue directional goals only if they reasonably can maintain an illusion of objectivity (Pyszczynski and Greenberg [1987]; Kunda [1990]). Our prompt, randomly assigned to half of our participants, states that research has shown cognitive dissonance to be a

---

3 Dyck et al. [2013] estimate that, if one considers undetected fraud, the ongoing fraud rate is about 13% in US public companies.
correlate of fraudulent reporting, and that CEOs who say things they believe to be untrue likely experience cognitive dissonance—a negative emotion that triggers unease, discomfort, annoyance or botheredness. In addition, only prompt condition participants are asked to assess how much cognitive dissonance the CEO felt during the conference call.

Because people are considerably more successful in detecting others’ emotional states than in detecting deception (Ambady and Weisbuch [2010]), we expect that prompted auditors would be able to detect unpleasant emotions in CEOs experiencing cognitive dissonance. Detection of these emotions coupled with knowledge that dissonance is correlated with deception, would make it unreasonable for very experienced auditors to rationalize away signs of fraud. Thus, the prompt frees very experienced auditors to apply their knowledge to fraud companies, instead of applying it predominantly to non-fraud companies. As a result, we predicted that prompted, very experienced auditors would commit fewer false negatives than unprompted, very experienced auditors. Regarding inexperienced auditors, we expected them to have accuracy goals instead of management-preferred directional goals, so that the prompt would yield less, if any, benefit for their evaluation of fraud companies.

Finally, we explore whether the medium—transcript only or transcript plus audio—used to process the earnings calls matters. While PCAOB AS No. 12 encourages auditors to review earnings conference calls, it does not identify a preferred medium. A priori, our theory does not warrant predicting the medium, per se, to matter overall or to matter differentially for very experienced versus inexperienced auditors. While an intuitively appealing possibility is that audio facilitates conveyance of emotional cues correlated with deception, several psychology studies on deception detection in which participants fail to outperform chance use live or
recorded audio. Exploring medium effects also serves as a robustness check for our main findings.

Our experiment is conducted online. Thirty-one auditors from multiple large public accounting firms with an average audit experience of 24 years participated, with 21 (68%) being current or retired U.S. partners. In addition, 180 inexperienced auditors participated (i.e., accounting students from a large state university). Each participant provides deception judgments for four publicly traded companies, using excerpted CEO responses to analyst questions during quarterly conference calls. Programmed software randomly draws excerpts from a population of five fraud and five non-fraud companies (total of 10), and we inform participants of this 50% fraud rate. We classify excerpts as fraudulent if the company’s quarterly financial statements later were restated and linked to fraud, regulator investigation, or class-action litigation. We also dichotomously measure participants’ experience and manipulate, between subjects, the presence of a cognitive dissonance prompt (absent versus present) and the medium of the conference call (transcript only versus transcript plus audio). The experiment begins with instructional videos, examples, and practice. For each company, participants receive background information and financial statements. Using CEO answers to analyst questions, participants decide whether or not they think the results being discussed are fraudulent. This decision is our primary dependent measure.

We find that, while experienced auditors’ fraud accuracy rates (67%) exceed those of students (53%) and chance overall, experienced and inexperienced auditors are equally prone to false negatives. That is, we observe no evidence that many years of experience alone helps auditors identify fraud companies. In fact, despite being told they would most likely evaluate two fraud companies and two non-fraud, very experienced auditors pick fraud only 40% of the time.
(relative to 46% for inexperienced auditors). Though consistent with a predicted tendency to minimize false positives, this finding is disconcerting, as false negatives increase the risk of investor loss due to undetected financial statement fraud.

Encouragingly, however, our prompt to attend to management’s cognitive dissonance significantly improves how accurately very experienced auditors assess fraud companies, from 43% to 70%. Further, very experienced auditors pick fraud only 31% of the time without the cognitive dissonance prompt, but 50% of the time with the prompt. Thus, a simple prompt to consider management’s cognitive dissonance significantly improves experienced auditors’ sensitivity to fraud.

In exploratory findings, we observe that adding audio to conference call transcripts improves the accuracy rates of inexperienced auditors (from 49% to 57%), but has no effect on very experienced auditors. Finally, we perform several robustness tests and examine process measures underlying the auditors’ fraud judgments. During training, participants were asked to identify specific CEO sentences that they perceived to include a red flag, before rendering their final judgment. We find that very experienced auditors are more accurate at identifying fraudulent statements in conference calls as red flags and that our cognitive dissonance prompt increases this accuracy for very experienced auditors.

In summary, this study contributes to the scholarly audit literature as well as to practice. Practitioners likely care that earnings conference calls can help auditors judgmentally assess fraud risk, and that many years of experience helps auditors more accurately interpret CEO narratives during conference calls, particularly when prompted to consider management cognitive dissonance. Practitioners also may be interested to learn that inexperienced auditors’ performance is improved they are provided with audio.
From a theory perspective, we develop and empirically test two new hypotheses about the individual and joint effects of many years of experience and a prompt to attend to management’s cognitive dissonance on auditors’ ability to detect deception. We observe new evidence consistent with very experienced auditors’ adaptation to their natural decision environments causing them to outperform both chance and inexperienced auditors in avoiding false positives, but also leading them to be significantly more concerned about false positives than false negatives. However, the significant interaction between experience and our cognitive dissonance prompt is encouraging and advances theory on how to improve auditors’ fraud detection capabilities.

2. Prior Research and Hypotheses

2.1 Auditor Judgments about the Presence and Absence of Deception in Earnings Conference Calls

As part of the requirement to understand conditions that might cause material misstatement in management’s financial statement, Auditing Standard No. 12 asks auditors to consider observing or reading the earnings conference call (PCAOB [2010]). The Q&A portion of the earnings conference call is a relatively unscripted narrative about current company operations, and recent archival evidence suggests that unique information exists in the Q&A dialogues between analysts and management (Blau, Delisle, and Price [2015], Hollander et al. [2010], Matsumoto et al. [2011], Mayew and Venkatachalam [2012], Price et al. [2012]). Research extending this line of work suggests that conversations and discussions during earnings conference calls contain information useful for identifying financial misreporting. For example,

---

4 Very experienced practitioners from several audit firms with whom we have spoken noted that, for that last several years, it has been common for experienced audit team members to read or listen to earnings conference calls to gather evidence about misstatement risks and business risks. However, they indicated that generally there were no formal audit procedures about conference calls in their audit methodologies during the same time span.
Larcker and Zakolyukina [2012] estimate linguistic-based classification models and document that CEO and CFO narratives during conference calls help identify deceptive discussions significantly better than chance levels. Hobson et al. [2012] use CEO speech samples from earnings conference call Q&As and document that vocal cues predict financial restatements beyond the predictive ability of financial accounting and linguistic-based predictors.

However, judgment based attempts to detect the presence of deception are quite challenging for individuals (DePaulo et al. [2003], Bond and DePaulo [2006]). Prior theoretical and empirical work in social psychology has shown that the accuracy rates of deception judgments made by experienced professionals are rarely better than those of inexperienced professionals or laypersons (Bond and DePaulo [2006] and Bond and DePaulo [2008]; Vrij [2008]). In auditing, Jamal et al. [1995] (see also Johnson et al. [2001]) find that few of the audit partners they examine detect seeded fraud in case materials. Lee and Welker [2007, 2008] find that upper-level accounting majors detect deception from inquiries about real-estate property values at rates only slightly better than chance.

Nevertheless, some evidence suggests that very experienced auditors may outperform inexperienced auditors in the domain of financial fraud detection (e.g., Bonner and Lewis [1990]; Knapp and Knapp [2001]). Brazel et al. [2010] show greater success in fraud brainstorming sessions when audit partners or forensic specialists, versus lower level staff, lead the session. Carcello and Nagy [2002] find a negative relationship between auditor industry specialization and undetected fraud. Knapp and Knapp [2001] show a positive effect for audit experience on

5 Experts examined include police officers, detectives, judges, interrogators, customs officials, mental health professionals, polygraph examiners, job interviewers, etc. One exception to this finding is that of so-called “lie detection wizards” (O'Sullivan and Ekman [2004]), who do systematically outperform novices.

6 We use the terms very experienced and inexperienced rather than expert and novices because part of what we test is whether experience helps auditors move from novice to expert (or to at least better) performance levels.
the effectiveness of analytical procedures in detecting fraud. Similarly, Bernardi [1994] shows that managers outperform seniors in fraud detection, but only when managers have relatively high moral development.

This conflicting evidence is at least partially resolved, however, if one considers that auditors’ experience may improve fraud detection accuracy only conditionally. Shaub and Lawrence [1999] find that experienced auditors are somewhat reluctant skeptics while inexperienced auditors are aggressive skeptics. Thus, audit experience could bias auditors away from predicting fraud for at least two reasons. First, auditors generally experience low rates of fraud occurrence during their careers. Even when auditors discover a material misstatement, that misstatement is usually due to unintentional error. While the true rate of financial statement fraud is unknown, detected fraud occurs at less than 3% of U.S. public companies (Dyck et al. [2013]). Prior psychology research finds that when individuals make decisions from experience they tend to underweight rare occurrences (Hertwig et al. [2004]).

Second, auditors’ incentives are structured such that there are disincentives to find fraud. Peecher et al. [2013] highlight that while auditors can be penalized for failing to find fraud, they are not rewarded for work to detect and prevent fraud. Nor are there any public regulatory rewards for performing audits of particularly high quality. Further, auditors who blow the whistle for fraud on their own client not only lose that client 50% of the time (Doty [2014]; Dyck et al. [2010]), but also are specifically excluded from the set of persons eligible to receive monetary rewards from the whistleblowing provisions of the Dodd-Frank Wall Street Reform and Consumer Protection Act.

Despite few rewards for finding fraud, both internal and external to the audit firm, the costs of predicting fraud loom large, particularly since fraud is a relative low-probability event.
These costs include fee pressure, budget issues, and client relationship issues, as auditors are motivated to meet budget and as audit fees are generally stable or decreasing (Audit Analytics Staff [2014]; Doty [2014]). Prior research finds that fee pressure and budget issues are salient motivators for audit teams (Willett and Page [1996]; Kelley and Margheim [1999]) that affect audit quality (e.g., Houston [1999]; Asare et al. [2000]; Ettredge et al. [2014]) and, potentially, fraud detection (Braun [2000]). Moreover, voiced suspicion of fraud likely will require additional audit procedures, causing tension, delays, and budget overruns.

We posit that these incentives, coupled with the rarity of fraud have profound psychological effects on experienced auditors, making it both desirable and seemingly likely that any particular company they are investigating is not committing fraud. In particular, two substantially related theoretical accounts exist with regard to the psychological processes experienced auditors likely adaptively learn to employ over time. One, auditors likely follow a primary error detection and minimization (PEDMIN) testing strategy (Friedrich [1993]). When doing so, they strive to minimize errors that they experientially have learned to be more important or salient. Since salient errors from making a “fraud exists” judgment are personally costly, for the reasons discussed above, false positive error (Type I) minimization—mislabeled a clean company as fraudulent—is likely the experienced auditor’s primary concern. Thus, very experienced auditors likely learn to not notice or to explain away red flag indicators of fraud since taking them at face value would increase the likelihood of making an experientially costly error. Importantly, this strategy of minimizing false positives likely comes at the expense of committing more false negatives, potentially creating disutility for regulators, investors, and other financial statement users.
Two, motivated reasoning theory holds that when decision makers have preferred outcomes, they (often subconsciously) activate directional goals, triggering skepticism of contrary evidence and ready acceptance of supportive evidence (Kunda [1990]). Auditors use motivated reasoning (Nelson [2009]), and it can impair their objectivity in assessing the quality of management’s preferred accounting treatments (Kadous et al. [2003]). To the extent experience enables auditors with directional goals to attend to more red flags, these auditors will attempt to explain them away in benign terms (Ditto and Lopez [1992]).

In sum, very experienced auditors will be more successful at correctly classifying non-fraud companies than fraud companies. Accuracy rates of inexperienced auditors, on the other hand, are subject to two forces. First, they are less likely to pursue management-preferred directional goals than are experienced auditors, having yet to experientially learn the relative costs of false positives and false negatives. Second, extant literature on deception detection suggests that their accuracy rates are unlikely to be better than chance (Zuckerman et al. [1981]; Bond and DePaulo [2006]; Vrij [2008]). Thus, inexperienced auditors’ accuracy rates will likely be similar and relatively low across fraud and non-fraud companies.

Overall, very experienced auditors are likely to outperform inexperienced auditors when classifying non-fraud companies but this accuracy difference is expected to dissipate when classifying fraud companies. These predictions lead to an ordinal interaction with the following two simple main effects: a simple main effect of experience given non-fraud companies and a simple main effect of company type (i.e., non-fraud versus fraud) given very experienced auditors.

**H1:** *Accuracy rates will be highest for very experienced auditors assessing non-fraud companies, significantly lower for very experienced auditors assessing fraud companies and similarly low for inexperienced auditors assessing both fraud and non-fraud companies.*
2.2 COGNITIVE DISSONANCE MANIPULATION

If auditors experientially learn from their decision environments that it is adaptive to explain away ambiguous fraud indicators to avoid false positives, identifying remedies that nudge auditors to reinterpret these cues likely will help prevent false negatives. In identifying a remedy, we consider a key principle of motivated reasoning theory. Specifically, this theory holds that decision makers use one-sided reasoning to attain directional goals only if they reasonably can maintain an illusion of objectivity (Pyszczynski and Greenberg [1987], Kunda [1990], Kadous, et al. [2003]).

To make it harder for experienced auditors to maintain this illusion, we devised a prompt. This prompt informed auditors that prior research has shown that cognitive dissonance is predictive of fraud and explained that people who say things that they believe to be untrue usually experience cognitive dissonance. Participants receiving the prompt also were asked to assess the degree to which managers felt cognitive dissonance on the calls that they analyzed. Thus, viewed exhaustively, this prompt contains three parts—the definition of cognitive dissonance (including its effect on individuals), a link from cognitive dissonance to deception, and encouragement to observe cognitive dissonance in the conference call CEO.

This prompt was desirable for several reasons. One is that deception frequently leads to cognitive dissonance (DePaulo et al. [2003], Harmon-Jones [2000]); Ekman [1985], [1992]). Another is that prior research shows dissonance markers in the speech of the CEOs in our sample were associated with financial misreporting (Hobson et al. [2012]), and a third reason is that people generally are better at detecting states of emotion than at detecting deception. As Ambady and Weisbuch [2010, 483] observe, “Whereas social perceivers do not exhibit particular intelligence in deception detection, emotion recognition is a different story….f” People often
accurately identify emotion at well above chance levels from vocal and linguistic cues, even without observing facial expressions, which also improves emotion detection (Elfenbein and Ambady [2002], Alm [2008]). Finally, psychology research finds that peak performance for the ability to perceive emotion occurs between the ages of 40 and 60 (Hartshorne and Germine [2015]), which is the likely age range of our experienced participants given that they have 24 years of audit experience on average.

We therefore conjecture that the cognitive dissonance prompt will help experienced auditors better detect deception in fraud companies. We expect that very experienced auditors’ repeated exposure in casual and formal interview settings with management will enable them to successfully detect the heightened emotions of CEOs experiencing cognitive dissonance, and that once the cognitive dissonance prompt is given, experienced auditors will no longer be able to reasonably ignore these cues. By contrast, we expect that for inexperienced auditors, the cognitive dissonance prompt will be less, if at all, helpful in improving their accuracy in detection of deception. While the prompt should make inexperienced participants more aware of cognitive dissonance and the link from dissonance to deception, inexperienced participants have not had repeated dialogs with management that facilitate detecting emotion and dissonance from speech. Thus, for inexperienced auditors, a cognitive dissonance prompt is less likely to improve the accuracy of their deception judgments. In combination, therefore, we predict the following ordinal interaction in which the cognitive dissonance prompt is significantly more helpful for very experienced than for inexperienced auditors.

H2: Accuracy rates for fraud companies will be highest for very experienced auditors who are prompted about management cognitive dissonance, significantly lower for very experienced auditors without a cognitive dissonance prompt and similarly low for inexperienced auditors, regardless of whether or not they receive a cognitive dissonance prompt.
2.3 EXPLORING THE EFFECTS OF CONFERENCE CALL MEDIUM

A priori, our theory does not compellingly warrant a directional prediction of how the medium of the call will affect auditor accuracy rates at detecting deception⁷. Nevertheless, we explore the medium through which auditors consume the conference call in part because PCAOB AS No. 12 suggests that auditors consider “observing or reading transcripts of earnings calls”. Observing conference calls could refer to reading or listening to conference calls (or possibly watching them, which we do not test herein). Given that the standard does not supply explicit guidance, it is unclear which media the auditor should use, and which of the two we examine is more beneficial.

Reading transcripts stresses the linguistic component of the call whereas listening stresses both the linguistic and vocal components. Interestingly, however, commercial providers of conference call transcripts typically purge speech hesitations (ah’s, um’s) and other linguistic features that likely provide important linguistic clues about cognitive dissonance, deception, or both. Our experiment holds this constant as we manually reinsert these speech hesitations into our written transcripts. This design choice likely biases against our analyses finding differences between our two design mediums, but provides a cleaner manipulation of medium.

Prior research is mixed as to the benefits of adding audio on detection accuracy rates. One the one hand, many psychology studies finding detection rates that fail to surpass chance also include live or taped audio. It could be that audio distracts receivers from linguistic features of interpersonal communication, as prior research suggests linguistic features are useful in detecting deception in both general and business contexts (e.g., Zhou et al. [2004]; Larcker and

---

⁷ The literature on the role of medium (commonly referred to as modality) is scarce. See Qin et al. (2005) for a listing of studies that are inconclusive on the role of modality. See Sweeney and Cici (2014) for more recent investigation of how modalities impact deception detection among subjects of different ages.
Zakolyukina [2012]). On the other hand, prior research suggests social information is contained in sender’s voices (Burgoon et al. [2008]), and hearing the CEO’s voice could facilitate comprehension of the linguistic cues themselves. Intuitively, this may be especially the case for novices who likely are less familiar with industry-specific performance metrics or jargon common in earnings conference calls, which would tax their working memory sources they otherwise could devote to deception detection. Since we have no strong a priori basis for a directional prediction we posit the following exploratory research question:

RQ1: Does the medium used to convey management’s responses during Q&A portions of earnings conference calls—transcript alone or audio plus transcript—affect auditors’ accuracy rates in fraud detection?

3. Method

3.1 PARTICIPANTS

To examine our hypotheses and provide a strong test (Kerlinger and Lee [2000], 459) of the effect of experience on fraud detection, we gather data from very experienced audit professionals and inexperienced audit students. We examine very experienced auditors because research suggests that expertise often takes very extensive, deliberate, and practical experience, in general and professional contexts (Ericsson et al. [1993]). We also recognize that our molar experience variable almost certainly encompasses several molecular constructs known to influence auditor task performance, such as technical knowledge, tacit knowledge, general diagnostic reasoning, etc. We intentionally use a coarse proxy for this study, however, as it is a relevant variable of interest in the natural audit ecology (many of these constructs are confounded in the real world) and because our study is a first step in examining determinants of
auditors’ ability to detect the presence versus absence of fraud from the Q&A sessions of earnings conference calls.\textsuperscript{8}

Our thirty-one very experienced participants are current or retired audit professionals from multiple large public accounting firms with an average of 24 years in the audit, assurance, and/or fraud/forensic services, and 22 years in public company audit. Twenty-one of them are partners or retired partners, four are managers, senior managers, or directors, two are seniors, and four are staff.\textsuperscript{9} All but three are CPAs. In addition, 180 students at a large public university enrolled in an auditing class as accounting majors provided usable judgments.

3.2 SPEECH CORPUS SELECTION

Each participant provided deception judgments to excerpts from the question and answer portion of a quarterly earnings conference call for four companies. Due to time constraints and our desire to solicit audit partners, participants could only reasonably evaluate four companies. We draw each company from a population of 10 public company quarter-end earnings conference calls containing five companies with deceptive discussions and five without. This 50\% rate of fraud follows the vast majority of deception detection experiments in the literature (Levine et al. 2014). The allocation of four companies to each participant is random with the stipulation that participants have an 80 percent chance of evaluating two fraud companies, a nine percent chance of one fraud company and three non-fraud companies, a nine percent chance of one non-fraud company and three fraud companies, a one percent chance of all fraud companies, and

\textsuperscript{8} We encourage future research that disentangles main and interactive effects of these various constructs. We do not have a sufficiently specified theory to a priori predict specific ranks at which there are significant increases in auditors’ ability to detect the absence of fraud or, when coupled with a prompt to consider manager’s cognitive dissonance, to detect the presence of fraud. Nor are we able to predict how precipitously and gradually the increases occur. We leave empirical tests of these matters to future research.

\textsuperscript{9} Though we requested responses from only auditors at only the partner level, we used all responses received from auditors, whether from partners, staff, etc. However, our results are robust (all p < 0.06) to excluding all staff, all staff and seniors, and any participant that identifies themselves as a forensic specialist.
and a one percent chance of no fraud companies.\textsuperscript{10} Additionally, adjustments were made to minimize any one company being evaluated much more frequently than another, and to ensure that any one company was not frequently presented in any one specific order position (e.g., always first).

All participants are made aware that the expected rate of fraud in the companies they will evaluate is 50%. The 10 company quarters are a subset of the 1,572 company-quarter earnings conference calls studied in Hobson et al. [2012], which originally were broadcast during calendar year 2007. We use two selection criteria. First, the fraudulent companies do not systematically differ from non-fraud companies in terms of financial statement predictors of fraud, and second, the companies are not generally well known, as described in more detail below.

Each of the 10 conference call excerpts are the first five minutes of CEO responses to analyst questions in the quarterly earnings conference call, including any analyst questions that elicit those first five minutes of CEO dialogue. We characterize excerpts as deceptive/fraudulent and code them as DECEP\_COMPANY equals 1 if the company’s quarterly financial statements being discussed in the conference call audio were restated (i.e. the fiscal quarter end falls between RES\_BEGIN\_DATE and RES\_END\_DATE on Audit Analytics via WRDS), the restatement adversely impacted the financial statements (RES\_ADVERSE = 1 on Audit Analytics via WRDS) and any of the following “irregularity conditions” hold: the restatement was deemed fraudulent (RES\_FRAUD = 1 on Audit Analytics via WRDS), a regulatory investigation followed the restatement (RES\_SEC\_INVESTIGATION = 1 on Audit Analytics

\textsuperscript{10} We use this distribution and multiple observations per participant for two reasons. First, we use a high-risk population because this is our population of interest. A natural-world equivalent is a client acceptance decision for a pool of risky potential clients. Second, we maximize value from the scarce resource of audit partner participants. A more realistic rate of detected fraud requiring restatement would dramatically increase the number of participants needed to draw useful inferences for fraud companies. We hold ex ante fraud risk constant by matching companies on F-Score, as discussed below. Also, see section 4.5 for a discussion of our selected fraud rate.
via WRDS), or a class action lawsuit followed it (DAYS_TO_SECURITIES_CLASS_ACTION > 0 on Audit Analytics via Audit Analytics online feed).

To meet the two criteria we have for selection of our 10 companies, we first calculate F-Scores (Dechow et al. [2011]) for all 1,572 observations in Hobson et al. [2012]. We then sort all observations by F-Score and for each irregularity quarter, we select the observation from the same two-digit industry with the closest F-Score, without replacement. If no such match is available within 10 observations in either direction, we use one digit SIC code, and if that fails, we take the closest F-Score without matching on industry. We then eliminate fraudulent observations and their related pair based on survey responses from accounting doctoral students that indicate the fraud was familiar and likely to be known to a general audience. Among the remaining companies, we chose the 10 with the widest absolute difference in vocal cognitive dissonance as measured in Hobson et al. [2012] such that fraud companies have higher levels of cognitive dissonance. This helps ensure the potential vocal markers of fraud are sufficiently salient to be considered by participants.

Of the five pairs that comprise the 10 observations, three are matched on two-digit SIC code, one is matched on 1 digit SIC code, and one is not matched on SIC code. The average (median) F-Score is 1.67 (1.90) for the five fraud companies and 1.94 (2.15) for the five non-fraud companies ($t_8 = 0.43, p = 0.68$).

3.3 SPEECH CORPUS PREPARATION

We manually transcribe each conference call excerpt rather than rely on commercial transcription as used in prior literature (Larcker and Zakolyukina [2012]). Manual transcription is necessary to ensure a clean manipulation of medium given commercial transcripts are purged of speech hesitations (ah’s, um’s etc.) that potentially provide useful information. Additionally,
while Hobson et al. [2012] isolated only the voice of the CEO in the conference call, thereby purging any important context from the CEO-analyst exchange, we include the analyst question(s) to which the CEO is responding so as to provide appropriate context. Analyst questions are read by a generic, computerized male or female voice. We use excerpts instead of the full question and answer period given practical limits on participant time.

3.4 PROCEDURE & VARIABLES

We focus on two key variables. First, we measure auditor experience at two levels (inexperienced audit student or very experienced audit professional). Second, between participants, we manipulate whether auditors are prompted to consider management’s cognitive dissonance (COGDIS). This manipulation has three parts. First, initial instructions for all participants preceding each company evaluation stated “Note: Research indicates that certain cues in what a CEO says and how s/he says it can help in the detection of deception.” Next, half of our participants were made aware of cognitive dissonance and additionally received the following:

One cue found to be useful in detecting deception in these CEO responses is cognitive dissonance. Cognitive dissonance is the negative, uncomfortable emotion a person feels when they are saying something that they know is not true. Those experiencing cognitive dissonance feel uncomfortable, uneasy, and bothered.

Next, after answering our principal dependent measure, participants prompted about management’s cognitive dissonance are asked to assess “how much cognitive dissonance the CEO felt during this excerpt of the conference call.” Overall, then, this prompt defines cognitive dissonance (including its effect on individuals), links cognitive dissonance to deception, and encourages participants to observe cognitive dissonance in the conference call.

For exploratory purposes and robustness, we also manipulate conference call medium at two levels. Some participants receive the conference call excerpt only as a transcript (AUDIO =
0) while others receive the transcript and accompanying audio (AUDIO = 1). We include the transcript in both conditions to isolate the effect of audio and because transcripts often are available long after audio feeds have been removed from company web sites.

Appendix A provides a timeline of the online experiment, which on average took about 1.75 hours. First, participants view a brief orientation video that instructs them to complete the experiment in one sitting without accessing outside information. Next, participants complete a guided tour and practice evaluating a fictional company. Written instructions and three narrated videos aid in this process. For example, we tell participants that their task is to use the CEO responses to analyst questions to (1) determine whether they think the results being discussed are fraudulent and to (2) identify potential red flags in the audit of the company. We provide time and encouragement for participants to practice the experimental tasks during this example.

The example and each of the four real companies participants evaluate consist of three parts. Part 1 consists of background company information listing the company name, the quarter being reported, a short business overview, and the four basic financial statements. Part 2 presents the transcript or the transcript with the audio and an area to record red flags. Participants in the transcript only (audio) condition are told that their task is to read the transcript (listen to the audio and follow along on the transcript) and identify red flags. We tell them that a red flag, in the context of the present experiment, exists any time they feel that the CEO's comments are suspicious, give them pause, or require additional investigation. We tell them that they must read the entire transcript in the transcript only condition and listen to the full audio in the audio only condition.

---

11 IRB approval was provided for the experiment.
In Part 3, we collect responses from participants. The first question posed to participants, and our primary dependent measure is, “Next, provide an overall judgment of whether it is more likely than not that fraud was being committed at this company during this quarter. That is, did this company later restate this quarter’s financial results due to one or more of the following: fraudulent financial results, a regulatory investigation, or a class action lawsuit?” Participants respond, “Yes, fraud was likely being committed during this quarter” or “No, fraud was not likely being committed during this quarter.” We then ask participants to state how confident they are about this judgment. We present all four evaluated companies in this manner. After evaluating the fourth company, participants complete a post-experiment questionnaire that asks several questions about work experience and experience detecting deception.

4. Results

4.1 DESCRIPTIVE STATISTICS

4.1.1 General Descriptive Statistics. Each of the 211 (31 very experienced and 180 inexperienced) participants provided four judgments, one per conference call, yielding 844 responses. We eliminate 17 observations because participants indicated they were familiar before participating in the study with fraud at the company, leaving 827 total observations, 211 auditors. Table 1 provides descriptive statistics for the entire sample and for very experienced and inexperienced auditors separately, as well as univariate tests of differences between very experienced and inexperienced. Overall, average audit experience (AUD_EXP) is 3.48 years,

---

12 Additional questions ask the participant whether they thought the CEO was lying, what areas of the financial statements appear problematic, and how familiar the participant was with the company before starting the experiment.

13 We eliminated observations when the participant responded “Yes” to the following question: “This company may or may not have had to restate this quarter’s financial results due to one or more of the following: fraudulent financial results, a regulatory investigation, or a class action lawsuit. Before participating in this study, were you aware of any financial improprieties for this company?” Seventeen responses (2.01% of all responses) were omitted: 14 from inexperienced auditors (1.94% of the inexperienced auditor responses) and three from experienced auditors (2.42% of experienced responses).
with 121 (14.63\%) of the responses from very experienced auditors (EXP). Average audit experience for very experienced auditors is 23.62 years, which is by definition greater than the average of 0.02 years for inexperienced auditing students.\textsuperscript{14} The average length of the conference call information provided, in audio time (\textsc{audio\_time}), was 7.24 minutes; the financial statement based fraud score (\textsc{f\_score}) was 1.82, and 51\% of the conference calls participants reviewed were deceptive (\textsc{decep\_company}). There was no statistical difference in \textsc{audio\_time} (7.26 vs. 7.24, \(p = 0.71\)), \textsc{f\_score} (1.89 vs 1.81, \(p = 0.32\)), or \textsc{decep\_company} (0.52 vs. 0.51, \(p = 0.79\)) between very experienced and inexperienced auditors.

4.1.2 Accuracy Rates. We first empirically test whether auditors, overall, detect the presence and absence of fraud from earnings conference calls at better than chance rates. Using a simple t-test adjusted for repeated measures, we observe that the overall accuracy rate (\textsc{accuracy}) of 55\% (see Figure 1 and Table 1) is statistically greater than chance levels of 50\% (\(p < 0.01\)), and in line with the average accuracy rate of 54\% that is commonly obtained when subjects face a 50\% fraud rate (Levine [2014]). Next, as a preliminary test of H1, we examine whether very experienced auditors are more accurate than inexperienced auditors in detecting the absence versus presence of fraud. We find that experience improves overall accuracy as very experienced auditors detect fraud at better than chance rates (67\%, \(p < 0.01\)), while inexperienced auditors do not (53\%, \(p = 0.14\)). Additionally, the overall accuracy rate is higher for very experienced than inexperienced auditors (\(p < 0.01\), see Table 1). Experienced auditors’ accuracy rate (67\%) is remarkably good, given reported accuracy rates for deception detection by experts in prior literature (Bond and DePaulo [2006]; [2008]).

\textsuperscript{14} A small fraction of inexperienced auditors had worked as audit interns, yielding a non-zero value of \textsc{aud\_exp}. 
4.2 H1—EFFECT OF EXPERIENCE ON ACCURACY RATES FOR FRAUD & NO FRAUD COMPANIES

H1 predicts a specific pattern of results that has two key parts, resulting in an ordinal interaction between experience (EXP) and company type (fraud versus non-fraud, DECEP.Company). First, we predict a positive simple main effect for EXP given a non-fraud company. Second, we predict a negative simple main effect for DECEP.Company given very experienced auditors, i.e., a lower accuracy rate for fraud versus non-fraud companies.

Univariate analysis supports these predictions, as indicated in Table 1. Participants judged 45% of companies to be deceptive (DECEP_JUDG). However, experienced auditors predict fraud only 40% (49/121) of the time, which is statistically less than the disclosed fraud rate of 50% ($p = 0.03$). While directionally consistent with our theory, this is somewhat surprising since we informed participants of the 50% fraud rate. Inexperienced auditors predict fraud 46% (326/706) of the time which differs significantly from the disclosed 50% rate ($p = 0.01$) but not from the experienced auditors’ judged deception rate ($p = 0.19$, see Table 1).

Figure 2 and Panel A of Table 2 show that while experienced auditors are very accurate for non-fraud companies (78%, specificity, true negative rate), they are considerably less accurate for fraud companies (57%, sensitivity, true positive rate). In fact, their accuracy rate for fraud companies is no better than chance ($p = 0.23$), and no better than that of inexperienced auditors ($p = 0.20$). Put another way, the proportion of false positive judgments/Type I errors are

---

15 During training participants are told that, “[The] four companies [you will evaluate] were taken from a larger set of companies. In this larger set, 50% of the companies had to restate their earnings due to fraud. Specifically, in half of the companies in this larger set, the quarterly and/or yearly financial results being discussed in the conference call were later restated….Since you have a sample of only four companies, you will not know for sure how many of these companies committed fraud. The most likely scenario is that you will evaluate two fraud companies and two clean companies. However, you could have any combination of fraud and clean companies, including all fraud companies or all clean companies.” Also, see section 4.5 for a discussion of our selected fraud rate.

16 This is very accurate in the sense of O’Sullivan and Ekman’s “wizard” or genius level of deception detection that they expect only one to two percent of individuals in expert populations to achieve (O’Sullivan and Ekman [2004], 271).
almost half as prevalent for experienced compared to inexperienced auditors (0.11 vs. 0.21, \( p < 0.01 \)), yet false negatives/Type II errors are not statistically different between very experienced and inexperienced auditors (\( p = 0.25 \)).

To formally test the ordinal interaction hypothesized in H1, we estimate a multivariate, 3-way, repeated measures logistic model, in which ACCURACY is the dependent variable, EXP, DECEP_COMPANY, and COGDIS and their interactions are the independent variables of interest, and AUDIO is a covariate.\(^{17}\) Figure 2 and Panel A of Table 2 present mean rates of accuracy, collapsed across COGDIS, and are consistent with H1. Panels B and C of Table 2 present analysis results. We directly test the pattern of results specified in H1 in a linear contrast. Specifically, collapsing across COGDIS, we test that accuracy rates will be highest for very experienced auditors assessing non-fraud companies, significantly lower for very experienced auditors assessing fraud companies and similarly low for inexperienced auditors assessing both fraud and non-fraud companies using weights of 3 (Experienced/Non-Fraud Company), -1 (Experienced/Fraud Company), -1 (Inexperienced/Non-Fraud Company), and -1 (Inexperienced/Fraud Company). Panel C displays the results of this contrast, which is significant, confirming H1 (\( X^2(1) = 10.84, p < 0.01 \) one tailed).

This table also tests the two necessary parts of the prediction in H1. First, Panel C shows a significant simple main effect for EXP given Non-Fraud Company, indicating that very experienced auditors have higher accuracy rates than those of inexperienced auditors when evaluating companies that did not commit fraud (\( t = 3.04, p < 0.01 \) one tailed). Second, Panel C shows that the simple effect of DECEP_COMPANY given Very Experienced is significant (\( t =

\(^{17}\) Specifically, we estimate generalized linear models (GLIMMIX via SAS 9.4) with random effects by subject to account for within subject correlation, using the logit link function given the dichotomous nature of the outcome variable.
2.34, $p = 0.01$, one tailed), indicating that very experienced auditors’ accuracy significantly decreases when they evaluate fraud companies relative to non-fraud companies.

Additional untabulated examination reveals that, collapsing across prompt conditions, the accuracy rate of very experienced auditors evaluating fraud companies is not statistically different from that of inexperienced auditors when they are evaluating fraud companies ($t_{206} = 1.23, p = 0.22$) nor that of inexperienced auditors evaluating non-fraud companies ($t_{206} = 0.13, p = 0.90$). Finally, inexperienced auditors’ accuracy rates also decrease significantly when evaluating fraud companies relative to non-fraud companies ($t_{206} = 2.02, p = 0.04$). That inexperienced auditor accuracy drops from above 50% for non-fraud companies to below 50% for fraud companies is consistent with inexperienced auditors exhibiting truth bias, which is the general tendency of more accurately detecting truth relative to lies (Levine et al. 1999). Overall accuracy, however, for inexperienced auditors is not statistically different from 50% as discussed earlier. This occurs because the 50/50 fraud rates in the research design prevents truth bias from confounding inferences on overall accuracy (Levine et al. 2014; Levine et al. 1999).

In summary, very experienced auditors’ superior accuracy, relative to that of inexperienced auditors, predominantly stems from their ability to correctly classify clean companies (i.e. high specificity), thus avoiding false positive (Type I) errors. When evaluating fraud companies, highly experienced auditors do not outperform audit students.

4.3 H2—COGNITIVE DISSONANCE PROMPT: DOES IT HELP?

In H2, we propose that a remedy that nudges auditors to reinterpret cues that are ambiguously suggestive of fraud will help improve sensitivity and prevent false negatives. We predict that accuracy rates for fraud companies will be highest for very experienced auditors that are prompted about management cognitive dissonance, significantly lower for very experienced
auditors without a cognitive dissonance prompt and similarly low for inexperienced auditors, regardless of whether or not they receive a cognitive dissonance prompt.

Univariate analysis supports these predictions. Very experienced participants judged just 30% of companies to be deceptive (DECEP_JUDG) when they did not receive a cognitive dissonance prompt, which is significantly less than 50% ($p < 0.01$, untabulated). However, when experienced auditors received the cognitive dissonance prompt, they predicted fraud 50% of the time, which is significantly higher than the rate when no prompt is given ($p = 0.02$). Figure 3 and Panel A of Table 2 present means that show a significant improvement in accuracy for fraud companies when very experienced auditors are (70%, which is greater than 50% or chance, $p < 0.01$) versus are not (43% which is not statistically different from 50% or chance, $p = 0.42$) given a cognitive dissonance prompt.

We test the ordinal interaction forwarded in H2 directly we again estimate a multivariate, repeated measures logistic model. Figure 3 and Panel A of Table 2 present mean rates of accuracy that are consistent with H2. Panels B and C of Table 2 present analysis results. We directly test the pattern of results specified in H2 using two linear contrasts. First, we fit linear contrast weights to each mean in the model, such that, very experienced auditors receive a 2.5 when evaluating a non-fraud company whether or not a prompt is provided and when evaluating a fraud company when a cognitive dissonance prompt is given. All five other means are weighted with -1.50. This contrast is significant, confirming H2 ($X^2(1) = 17.46, p < 0.01$ one tailed). Next, we use a linear contrast to test only judgments of fraud companies. Specifically, very experienced auditors are given a weight of 3 when they receive a cognitive dissonance prompt (and judge a fraud company). All other judgments of fraud companies are given a weight of -1. This contrast is significant, confirming H2 ($X^2(1) = 5.70, p = 0.01$ one tailed).
Additionally, we find that the accuracy rate of 0.70 for very experienced auditors given a cognitive dissonance prompt and evaluating a fraud company, is greater than each of the other three accuracy rates evaluating a fraudulent company (all \( p < 0.03 \), one tailed).

4.4 RESEARCH QUESTION — AUDIO VERSUS TRANSCRIPT

Collapsing across COGDIS, we examine our research question, RQ1, through an untabulated, repeated measures logistic analysis, in which ACCURACY is the dependent variable and EXP and AUDIO are the independent variables. The pattern of means in this analysis—0.49 (Inexperienced / Text), 0.55 (Inexperienced / Audio + Text), 0.69 (Experienced / Text), and 0.65 (Experienced / Audio + Text)—appears consistent with an ordinal interaction such that accuracy rates are highest for very experienced auditors when audio is and is not available, equal or lower for inexperienced auditors when audio is available, and lowest for inexperienced auditors when only text is available. Indeed, we find a significant simple main effect for EXP given text (\( t = 3.11, p < 0.01 \) one tailed), and a significant simple main effect of AUDIO given Inexperienced (\( t = 2.17, p = 0.02 \) one tailed). Further, the simple effect of AUDIO given Very Experienced is not significant (\( t = 0.48, p = 0.63 \)). Additional analysis reveals that when audio is provided, accuracy rates of very experienced auditors (65%) and inexperienced auditors (55%) are indistinguishable at conventional levels (\( t = 1.14, p = 0.26 \)). In summary, providing audio to inexperienced auditors improves their accuracy more so than it improves that of experienced auditors. Additionally, that experienced auditors exhibit consistently high accuracy levels in both the audio and text only condition suggests our earlier findings on the superiority of experienced auditors is not an artifact of the conference call medium.

4.5 ROBUSTNESS TESTS
4.5.1 Time  Participants spent an average (median) of 19 (11) minutes consuming each conference call, with inexperienced participants (16 / 11) using significantly less time per call than that of experienced participants (39 / 15, \(t = 2.21, p = 0.03\)). To assess whether time spent plays a significant role in accuracy, we rerun our analyses for H1 and H2 with time as a covariate (control variable, untabulated). We estimate the natural logarithm of time spent as the variable is right skewed. The coefficient on time spent is positive but insignificant (\(p = 0.39\)), and our inferences remained unchanged.

4.5.2 Familiarity with the Company  We ask the following question to assess participants’ familiarity with each company they are evaluating: “What general level of familiarity did you have with this company before doing this study?”, where 0=Very Low and 10=Very High. Overall, familiarity is quite low, at 0.05 and does not significantly differ between inexperienced and very experienced auditors (\(p = 0.28\)). Our inferences are not changed when we add familiarity as a covariate (control variable) in our analysis.

4.5.3 50-50 Fraud Rates  The ideal deception detection study maximizes external validity by using a fraud rate that matches the fraud rate in the field (Levine et al. 2014). Unfortunately, as is the case in many deception detection settings, we do not have precise estimates on the true rate of fraud an auditor observes overall. As such, we follow the vast majority of deception judgment experiments and use a 50/50 fraud rate. Could this research design choice alone drive the pattern of results we observe? Suppose that despite being informed that the fraud rate is 50/50, all audit subjects to some extent simply apply the fraud rate they are accustomed to. Very experienced auditors’ may have difficulty overcoming a lifetime of experiencing relatively low rates of fraud while inexperienced auditors, on the other hand, would be less subject to this issue given that fraud is frequently discussed in audit lectures and audit
text and case books. In such a case, for any given judgment, very experienced auditors would tend to choose truth (deception) more (less) often than inexperienced auditors.

Empirical evidence consistent with this conjecture would include very experienced auditors selecting fraud with lower frequency than inexperienced auditors as well as exhibiting higher (lower) accuracy for non-fraud (fraud) firms. The evidence in Table 1 reveals no statistical difference between very experienced and inexperienced auditors in the proportion of judgments deemed fraudulent. Moreover, in Figure 2 we observe that while very experienced auditors did exhibit higher accuracy on non-fraud companies relative to fraud companies, they were not simultaneously less accurate on fraud firms.

Additionally, it is not clear how the choice of 50/50 fraud rate would somehow account for the results we observe with respect to the cognitive dissonance manipulation. Collectively, then, we reject the choice of a 50/50 fraud rate as a mechanical driver of the results we observe. One benefit of the 50/50 fraud rate is that very experienced auditors provide large number of judgments pertaining to fraud companies, allowing for the analysis of the process by which auditors detect financial deception.

4.6 SUPPLEMENTAL PROCESS ANALYSIS

To better understand the process that underpins the differences we find between experienced and inexperienced auditors, and to further explore the effect of the dissonance prompt, we examine data related to auditor identification of red flags.

4.6.1 Auditors’ Red Flagged Issue Notations  We begin by categorizing the open-ended notations auditors provided pertaining to perceived red flag issues. To do so, two separate raters with knowledge of accounting but with no knowledge of the hypotheses being tested nor the experience level of the auditor providing the response, independently coded the open-ended
notes auditors made when identifying red flags. The coders identified 14 categories and counted
the number of issues identified in each of the 14 categories. The two coders’ counts in each
category were highly, positively correlated (all p < 0.01).18 We average their responses together
and then arrange the 14 categories into three broad classifications: financial statement items
(FIN), speech (SPEECH), and suspect knowledge (SKNOW). The FIN group contains issues
pertaining to financial statements, financial results, company strategy, and topics regarding the
appropriate application of judgment when applying accounting standards given the operating
environment. The SPEECH group captures speech hesitations, filled pauses (ah’s and um’s),
language use that avoids answering questions or deflects blame, and speech that suggests the
speaker was nervous. The SKNOW group contains mentions of instances where the executive
appeared to lack knowledge of the subject matter by giving answers that seemed incomplete,
inaccurate, internally inconsistent, or lacking appropriate level of confidence. The sum of FIN,
SPEECH and SKNOW represents the total number of red flagged issues identified
(RF_ISSUES).

We first observe that while the number of characters provided in the text box
(ISSUE_CHARS) from which the red flagged issues were coded are almost identical across
experience level (242 versus 243 for inexperienced versus very experienced auditors), very
experienced auditors provide 3.07 issues per company, on average, which is statistically larger
than the 2.32 inexperienced auditors provide (t = 2.72, p = 0.01, see Table 3). Thus, very
experienced auditors are more efficient at detecting potential problems in conference calls. Next,
when we compare the number of red flagged issues found based upon whether a company did or

---

18 All but six of the 14 categories have Cronbach’s alphas greater than 0.70. Inferences are identical if these six
categories are omitted with the exception that RF_ISSUES significantly increase for fraud companies relative to
non-fraud companies for inexperienced auditors.
did not commit fraud, we find that very experienced auditors find significantly more issues overall and in each of the three categories (all \( p < 0.05 \)) for fraud versus non-fraud companies. However, inexperienced auditors do not find more issues overall for fraud versus non-fraud companies (\( p = 0.19 \)), and only find statistically more issues for fraud companies in the FIN category (\( p < 0.01 \)). Additionally, experienced auditors seem better able to identify differences regarding ”softer” issues (e.g., the issues categorized into the SPEECH and SKNOW categories).

[Table 3 about here]

4.6.2 Dissonance and Red Flag Sentence Level Sensitivity. We next examine whether we can obtain a deeper understanding of the enhanced sensitivity among experienced auditors. Earlier we observed the dissonance prompt resulted in higher accuracy for experienced auditors among fraudulent firms (i.e. increased sensitivity). Is this increased sensitivity an overall sense by the experienced auditor that the firm is likely misreporting or can it be traced to actual fraudulent statements in conference call narrative?

To address this question, we assess auditor sensitivity at the sentence level unit of analysis. We execute this analysis in two steps. First, we identify the set of sentences pertaining to fraud and, among this set of sentences, assess the extent to which auditors indicated the sentence as a red flag. We label the ratio of accurately identified red flag sentences to total fraudulent sentences red flag accuracy (RF_ACCURACY). Second, we examine whether the dissonance manipulation results in higher sensitivity (i.e. higher levels of RF_ACCURACY) for experienced auditors.

To execute the first step, each CEO sentence in the transcript of each fraudulent firm is classified as fraudulent or not fraudulent, depending on whether the topic of the sentence pertains to the topic of the fraud. This subjective sentence classification of fraud topic is done by a
coauthor and several teams of research assistants with no knowledge of the red flag results, using hand collected data about each of the five frauds. The subjective coding of sentence topics is compared with whether the auditor subject identified the same sentence as a “red flag.” We define red flag accuracy (RF_ACCURACY) as 1 if the auditor subject flags a fraudulent sentence, and 0 if the auditor does not flag a fraudulent sentence. In Panel A of Table 4, we observe that there were 3,117 fraudulent sentences observed in total by auditor subjects, with 19% flagged as fraudulent overall, and 27% (17%) flagged as fraudulent by experienced (inexperienced) auditors. These sentence level sensitivities are an order of magnitude lower than the firm level sensitivities reported in Table 2 of 57% (49%) for experienced (inexperienced) auditors due to differential fraud rates. The overall firm level rate of fraud in Table 2 is 50% by design. The measured rate of fraud at the sentence level is 25% (3,117 fraud topic sentences out of a total of 12,293 sentences), and is based on CEOs actual discussion of fraud in the transcript excerpts we utilized.

To execute the second step, we again use a multivariate, repeated measures logistic analysis to assess whether audit experience and the cognitive dissonance prompt influences auditor sensitivity. RF_ACCURACY is the dependent variable and EXP, COGDIS, and their interaction are the explanatory variables. Table 4 panels B and C provides the estimation results. We find that very experienced auditors exhibit statistically greater sensitivities (0.27) than inexperienced auditors (0.17, $X^2 (1) = 12.30, p < 0.01$, two tailed).

---

19 Just over 100 second-year audit students were divided into small groups and asked to discover and examine the background information behind the fraud for each fraud company, and then to categorize each CEO sentence based upon how directly the fraud topic was discussed. Six groups were assigned to each fraud company in our sample. Of the five fraud firms, here there were a total of 146 sentences, of which 40 were identified as pertaining to the fraud topic.
Next, we find that the cognitive dissonance prompt marginally significantly increases very experienced auditors’ red flag accuracy (0.21 to 0.32, \( t = 1.78, p = 0.08 \), two tailed), but not the accuracy of inexperienced auditors (0.17 to 0.17, \( t = 0.43, p = 0.67 \), two tailed). A linear contrast testing the ordinal interaction in which the accuracy of very experienced auditors receiving the cognitive dissonance prompt is given a weight of 3 while the other three means are given a weight of -1 is significant (\( p < 0.01 \), two tailed). This effect is summarized visually in Figure 4, whereby experienced auditors receiving the dissonance prompt exhibit higher sensitivities than both experienced, unprompted auditors and inexperienced auditors whether prompted or not (sensitivities of 0.32, 0.21, 0.17 and 0.17 respectively). This pattern of sentence level sensitivities mimics the pattern observed in Figure 3 at the firm level (sensitivities of 0.70, 0.43, 0.51, 0.47, respectively), suggesting the dissonance manipulation does indeed translate into experienced auditors improving fraud detection at the sentence level.

5. CONCLUSION AND IMPLICATIONS

We use an experiment to examine how extensive audit experience and a prompt to attend to management’s cognitive dissonance individually and jointly influence auditors’ ability to detect deception from question and answer (Q&A) portions of earnings conference calls. We also explore which medium of the call, audio or transcript, is most helpful to the auditor in that effort. We ask very experienced auditors (average experience of 24 years) and audit students to evaluate the conference calls of four companies, and to predict whether the companies were committing fraud. We identify fraud by whether the financial results released at the time of the conference call are eventually restated due to an irregularity.

We find both good and bad news. Examining overall accuracy, we find that very experienced auditors predict accurately 67% of the time, and significantly outperform both
inexperienced auditors and chance. This is good news as meta-analyses document that experts rarely outperform chance in detecting deception (e.g., Bond and DePaulo [2006]; [2008]). This accuracy rate increases even more, to 78%, when experienced auditors judge non-fraud companies. This rate is approaches O'Sullivan and Ekman’s “wizard” or genius level of deception detection that they expect only one to two percent of individuals in expert populations to achieve (O'Sullivan and Ekman [2004], 271).

The bad news, however, is that this accuracy stems principally from fewer false positives. That is, very experienced auditors have high specificity but low sensitivity, and in turn outperform chance and inexperienced auditors only for companies that do not commit fraud, contrary to the likely preference of financial statement users. Nevertheless, this pattern of findings is consistent with psychology-based theory. We theorize that auditors’ experiential learning with respect to fraud does not primarily make it adaptive for them to avoid or minimize false negatives but rather to avoid or minimize false positives (Friedrich [1993]).

We also develop a psychology-based remedy to help experienced auditors overcome their experientially learned avoidance of false positives: a prompt to attend to management’s cognitive dissonance. With this prompt, experienced auditors increase sensitivity and are significantly more accurate in detecting fraud among fraud firms (from 43% to 70%) and pick fraud more often (31% to 50%). In process analysis, we find that very experienced auditors identify more conference call red flags among fraudulent companies relative to non-fraudulent companies. Moreover, we find that the enhanced sensitivity observed as a result of the dissonance prompt translates to the sensitivity of fraud detection at the sentence level within conference transcripts. Finally, providing the audio of conference calls improves the accuracy rates of inexperienced auditors but provides no benefit to experienced auditors.
Our research is subject to limitations. First, our results generalize most readily to situations in which the auditor is unfamiliar with the client. Future research could examine the robustness of our effects to situations in which the auditor assesses their own client. We expect familiarity would exacerbate experienced auditors’ avoidance of false positives and so present a stronger challenge for our cognitive dissonance prompt remedy (Bowlin et al. [2015]). Second, we do not examine interactions between auditor judgments and other systems audit firms may have in place for detecting fraud. The extent to which auditors view such systems as threats to their own expertise, in turn influencing their detection accuracy, remains unexplored (Elkins et al. [2013]).

This research contributes to both audit practice and research. From a practice perspective, we show that conference calls are a potentially useful source for auditors to find cues to potential fraud. Conference calls are plentiful and low-cost avenues of information about a current or potential client. For example, while the prospect of growth in audit revenues is high in BRIC (Brazil, Russia, India, and China) and other developing countries, auditors have expressed to us hesitancy accepting prospective clients in some parts of the world with higher corruption indices (etc.) due to unknown and potentially elevated levels of fraud risk. To the degree that our theory and experimental findings helps auditors to better discriminate which clients to accept/retain versus reject because of heightened risk of fraud, it will be useful to the profession and to investors.

Further, we begin to address the question of who at the audit firm should consume the conference call. While highly experienced audit personnel are seemingly ideal for this task, inexperienced auditors provided with the audio version of the call can predict fraud from conference calls at better than chance rates.
From a research perspective, we demonstrate a setting in which individuals successfully detect deception from dialogue. In particular, we show a setting in which very experienced auditors (experts) perform well in avoiding false positives. We also highlight the need for additional research examining experienced auditors’ reluctance to predict fraud or “truth bias.” Experienced auditors’ elevated Type II error rates should be of concern to both academics, regulators, investors, and other financial statement users. Future research could examine the particular dispositional and institutional factors that are most influential in this bias. Finally, we identify a potentially useful remedy—prompting auditors to consider the cognitive dissonance of the speaker—to aid experienced auditors in mitigating these Type II errors.
Appendix A: Time Line

Video guided tour through the mechanics of the experiment, including practicing with a hypothetical company.

Evaluate four companies, in turn

Exit survey

For each company, participants receive the following:

Part 1: background company information, the quarter being reported, a short business overview, balance sheet, income statement, and statement of cash flows

Part 2: transcript (transcript and audio) if Audio = 0 (Audio = 1) and an area to enter red flags

Part 3: response variables, including, first, our main dependent measure: "...was fraud being committed during this quarter?"
Appendix B: Variable Definitions

ACCURACY: Indicator variable that identifies whether the participant correctly classifies the conference call: equals 1 if \( \text{DECEP\_COMPANY} = 1 \) and \( \text{DECEP\_JUDG} = 1 \) or \( \text{DECEP\_COMPANY} = 0 \) and \( \text{DECEP\_JUDG} = 0 \) and 0 otherwise.

AUD_EXP: Number of years of experience as an assurance professional reported by the participant.

AUDIO: Indicator variable that equals 1 if the conference call excerpt was provided both as a transcript and as audio, 0 if as a transcript only.

AUDIO_TIME: Duration in minutes of the audio version of the conference call assigned to the participant.

COGDIS: Cognitive dissonance prompt is a between subjects manipulated variable that has two parts. First, participants receiving this manipulation are given additional initial instructions that said the following: “One cue found to be useful in detecting deception in these CEO responses is cognitive dissonance. Cognitive dissonance is the negative, uncomfortable emotion a person feels when they are saying something that they know is not true. Those experiencing cognitive dissonance feel uncomfortable, uneasy, and bothered.” Second, after answering our principal dependent measure, these participants are asked “how much cognitive dissonance the CEO felt during this excerpt of the conference call.” \( \text{COGDIS} = 1 \) when both of these components are present and 0 otherwise.

DECEP\_COMPANY: Indicator variable that equals 1 if conference call pertains to a quarter that was eventually restated due to an irregularity (Fraud Companies), and 0 otherwise (Non-Fraud Companies).

DECEP\_JUDG: Indicator variable that equals 1 if the participant judges the conference call to be related to a company quarter that will eventually be restated due to an irregularity, and 0 otherwise.

EXP: Indicator variable that equals 1 if the response is from a very experienced audit professional, and 0 if the response is from an inexperienced accounting major enrolled in an auditing course.

FSCORE: Fraud score of the company for the fiscal quarter being discussed on the earnings conference call.

FIN: For each of the four companies they evaluated, we ask auditors to read (listen to) the transcript and write down the time stamp (line number from the transcript) of each red flag and as much detail about the red flag as possible. Two separate raters with knowledge of accounting but with no knowledge of the hypotheses being tested nor the experience level of the auditor providing the response, independently coded the open-ended notes auditors make when noting red flags. The coders counted the number of issues and sorted them into 14 categories. We average their responses together and then arrange the 14 categories into three broad groups:
financial statement items (FIN), speech (SPEECH), and suspect knowledge (SKNOW). The FIN group contains red flagged issues pertaining to financial statements, financial results, company strategy, and topics regarding the appropriate application of judgment when applying accounting standards given the operating environment.

**SPEECH**
The SPEECH group captures speech hesitations, filled pauses (ah’s and um’s), language use that avoids answering questions or deflects blame, and speech that suggests the speaker was nervous.

**SKNOW**
The SKNOW group contains red flagged issues identifying instances where the executive appeared to lack knowledge of the subject matter by giving answers that seem incomplete, inaccurate, internally inconsistent, or lacking appropriate level of confidence.

**RF_ISSUES**
The sum of FIN, SPEECH and SKNOW represents the total number of red flagged issues identified (RF_ISSUES).

**ISSUE_CHARS**
ISSUE_CHARS is the number of characters provided in the text box from which the red flagged issues were coded.

**RF_ACCURACY**
Red flag detection accuracy, analyzed for only the five fraud companies, and is constructed in three steps. First, each participant is given a 1 (0) if they identify (do not identify) a CEO’s sentence as a red flag. This is done for each CEO sentence. Second, each CEO sentence is classified as fraudulent or not fraudulent. This subjective classification is done by a coauthor and several teams of research assistants with no knowledge of the red flag results, using hand collected data about each of the five frauds. Finally, RF_ACCURACY receives a 1 if the auditor correctly flags a fraudulent sentence or does not flag a non-fraudulent sentence, and a 0 otherwise.
References


AUDIT ANALYTICS STAFF 'Analysis of Audit Fees by Industry Sector,' in Book Analysis of Audit Fees by Industry Sector, edited by Editor. City, 2014.


HELLER, M. 'Accounting-related Securities Class Actions Climb 47%,' CFO.com, 2015.


PUBLIC COMPANY ACCOUNTING OVERSIGHT BOARD. Observations on auditors' implementation of PCAOB standards relating to auditors' responsibilities with respect to fraud. 2007.


FIGURE 1
Descriptive Statistics

Overall Accuracy Rates

![Accuracy Rate (ACCURACY)]

EXP is an indicator variable that equals 1 if the response was from a very experienced audit professional (24 years in audit on average), and 0 if the response was from an inexperienced accounting major enrolled in an auditing course. Fraud and non-fraud companies are identified using DECEP_COMPANY, which is an indicator variable that equals 1 if conference call pertains to a quarter that was eventually restated due to an irregularity (Fraud Company), and 0 otherwise (Non-Fraud Company). DECEP_JUDG is an indicator variable that equals 1 if the participant judges the conference call to be related to a company quarter that will eventually be restated due to an irregularity, and 0 otherwise. The dependent measure is ACCURACY, which is an indicator variable that identifies whether the participant correctly classifies the conference call: equals 1 if (DECEP_COMPANY = 1 and DECEP_JUDG = 1 or DECEP_COMPANY = 0 and DECEP_JUDG = 0) and 0 otherwise.
Figure 2
Accuracy of Very Experienced and Inexperienced Auditors for Fraud and Non-Fraud Companies

Very experienced and inexperienced auditors are identified using EXP, which is an indicator variable that equals 1 if the response was from a very experienced audit professional (24 years in audit on average), and 0 if the response was from an inexperienced accounting major enrolled in an auditing course. Fraud and non-fraud companies are identified using DECEP_COMPANY, which is an indicator variable that equals 1 if conference call pertains to a quarter that was eventually restated due to an irregularity (Fraud Company), and 0 otherwise (Non-Fraud Company). DECEP_JUDG is an indicator variable that equals 1 if the participant judges the conference call to be related to a company quarter that will eventually be restated due to an irregularity, and 0 otherwise. The dependent measure is ACCURACY, which is an indicator variable that identifies whether the participant correctly classifies the conference call: equals 1 if (DECEP_COMPANY = 1 and DECEP_JUDG = 1 or DECEP_COMPANY = 0 and DECEP_JUDG = 0) and 0 otherwise.
FIGURE 3
The Effect of Audit Experience and Cognitive Dissonance Prompt on Accuracy for Fraud and Non-Fraud Companies

Very experienced and inexperienced auditors are identified using EXP, which is an indicator variable that equals 1 if the response was from a very experienced audit professional (24 years in audit on average), and 0 if the response was from an inexperienced accounting major enrolled in an auditing course. Fraud and non-fraud companies are identified using DECEP_COMPANY, which is an indicator variable that equals 1 if the conference call pertains to a quarter that was eventually restated due to an irregularity (Fraud Company), and 0 otherwise (Non-Fraud Company). AUDIO is an indicator variable that equals 1 if the conference call excerpt was provided both as a transcript and as audio, 0 if as a transcript only. DECEP_JUDG is an indicator variable that equals 1 if the participant judges the conference call to be related to a company quarter that will eventually be restated due to an irregularity, and 0 otherwise. The dependent measure is ACCURACY, which is an indicator variable that identifies whether the participant correctly classifies the conference call: equals 1 if (DECEP_COMPANY = 1 and DECEP_JUDG = 1) or (DECEP_COMPANY = 0 and DECEP_JUDG = 0) and 0 otherwise. Cognitive dissonance prompt is a between-subjects manipulated variable that has two parts. First, participants receiving this manipulation are given additional initial instructions that said the following: “One cue found to be useful in detecting deception in these CEO responses is cognitive dissonance. Cognitive dissonance is the negative, uncomfortable emotion a person feels when they are saying something that they know is not true. Those experiencing cognitive dissonance feel uncomfortable, uneasy, and bothered.” Second, after answering our principal dependent measure, these participants are asked “how much cognitive dissonance the CEO felt during this excerpt of the conference call.” COGDIS = 1 when both of these components are present and 0 otherwise.
Very experienced and inexperienced auditors are identified using EXP, which is an indicator variable that equals 1 if the response was from a very experienced audit professional (24 years in audit on average), and 0 if the response was from an inexperienced accounting major enrolled in an auditing course. Cognitive dissonance prompt is a between subjects manipulated variable that has two parts. First, participants receiving this manipulation are given additional initial instructions that said the following: “One cue found to be useful in detecting deception in these CEO responses is cognitive dissonance. Cognitive dissonance is the negative, uncomfortable emotion a person feels when they are saying something that they know is not true. Those experiencing cognitive dissonance feel uncomfortable, uneasy, and bothered.” Second, after answering our principal dependent measure, these participants are asked “how much cognitive dissonance the CEO felt during this excerpt of the conference call.” COGDIS = 1 when both of these components are present and 0 otherwise. The dependent measure, red flag detection accuracy (RF_ACCURACY), analyzed for only the five fraud companies, is constructed in three steps. First, each participant is given a 1 (0) if they identify (do not identify) a CEO’s sentence as a red flag. This is done for each CEO sentence. Second, each CEO sentence is classified as fraudulent or not fraudulent. This subjective classification is done by a coauthor and several teams of research assistants with no knowledge of the red flag results, using hand collected data about each of the five frauds. There were 3,117 fraudulent sentences (out of a total 12,293 sentences) analyzed by subject auditors. RF_ACCURACY measures sensitivity of auditors to fraudulent sentences, and equals 1 if the auditor correctly flags a fraudulent sentence as a red flag and 0 if the auditor does not flag a fraudulent sentence.
### TABLE 1
Descriptive Statistics

<table>
<thead>
<tr>
<th>Variable</th>
<th>Inexperienced (N = 706)</th>
<th>Very Experienced (N = 121)</th>
<th>All Responses (N = 827)</th>
<th>Test of Experienced vs. Inexperienced</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean (std.)</td>
<td>Mean (std.)</td>
<td>Mean (std.)</td>
<td>Test Statistic (p-value)</td>
</tr>
<tr>
<td>EXP</td>
<td>0.00</td>
<td>0.00</td>
<td>1.00</td>
<td>1.00</td>
</tr>
<tr>
<td></td>
<td>(0.00)</td>
<td>1.00</td>
<td>(0.00)</td>
<td>1.00</td>
</tr>
<tr>
<td>AUD_EXP</td>
<td>0.02</td>
<td>(0.13)</td>
<td>23.62</td>
<td>(13.08)</td>
</tr>
<tr>
<td></td>
<td>1.00</td>
<td>0.00</td>
<td>38.00</td>
<td>1.00</td>
</tr>
<tr>
<td>AUDIO_TIME</td>
<td>7.24</td>
<td>(0.60)</td>
<td>7.26</td>
<td>(0.61)</td>
</tr>
<tr>
<td></td>
<td>8.25</td>
<td>6.30</td>
<td>8.25</td>
<td>6.30</td>
</tr>
<tr>
<td>FSCORE</td>
<td>1.81</td>
<td>(0.89)</td>
<td>1.89</td>
<td>(0.88)</td>
</tr>
<tr>
<td></td>
<td>2.97</td>
<td>0.41</td>
<td>2.97</td>
<td>0.42</td>
</tr>
<tr>
<td>DECEP.Company</td>
<td>0.51</td>
<td>(0.50)</td>
<td>0.52</td>
<td>(0.50)</td>
</tr>
<tr>
<td></td>
<td>1.00</td>
<td>0.00</td>
<td>1.00</td>
<td>0.00</td>
</tr>
<tr>
<td>DECEP.Judgment</td>
<td>0.46</td>
<td>(0.50)</td>
<td>0.40</td>
<td>(0.49)</td>
</tr>
<tr>
<td></td>
<td>1.00</td>
<td>0.00</td>
<td>1.00</td>
<td>0.00</td>
</tr>
<tr>
<td>ACCURACY</td>
<td>0.53</td>
<td>(0.50)</td>
<td>0.67</td>
<td>(0.47)</td>
</tr>
<tr>
<td></td>
<td>1.00</td>
<td>0.00</td>
<td>1.00</td>
<td>0.00</td>
</tr>
</tbody>
</table>

Very experienced and inexperienced auditors are identified using EXP, which is an indicator variable that equals 1 if the response was from a very experienced audit professional (24 years in audit on average), and 0 if the response was from an inexperienced accounting major enrolled in an auditing course. AUD_EXP is the number of years of experience as an assurance professional reported by the participant. AUDIO_TIME is the duration in minutes of the audio version of the conference call assigned to the participant. FSCORE is the fraud score of the company for the fiscal quarter being discussed on the earnings conference call. Fraud and non-fraud companies are identified using DECEP_COMPANY, which is an indicator variable that equals 1 if conference call pertains to a quarter that was eventually restated due to an irregularity (Fraud Company), and 0 otherwise (Non-Fraud Company). DECEP_JUDG is an indicator variable that equals 1 if the participant judges the conference call to be related to a company quarter that will eventually be restated due to an irregularity, and 0 otherwise. The dependent measure is ACCURACY, which is an indicator variable that identifies whether the participant correctly classifies the conference call: it equals 1 if (DECEP_COMPANY = 1 and DECEP_JUDG = 1 or DECEP_COMPANY = 0 and DECEP_JUDG = 0) and 0 otherwise.

Test statistics are $\chi^2$ or F-statistics for dichotomous or continuous variables, respectively, that control for repeated measures via random effects estimation of a generalized linear model in SAS 9.4 (PROC GLIMMIX). All p-values are two tailed. Results are excluded when the participant indicated that they were aware of financial improprieties for the company before participating in this experiment. N/A indicates no statistical test is performed as the difference in means occurs by definition.
TABLE 2
Accuracy of Very Experienced and Inexperienced Auditors for Fraud and Non-Fraud Companies

Panel A: Descriptive Statistics – ACCURACY by EXP, DECEP_COMPANY, and COGDIS

<table>
<thead>
<tr>
<th></th>
<th>Inexperienced</th>
<th></th>
<th>Very Experienced</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Non-Fraud</td>
<td>Fraud</td>
<td>Total</td>
<td>Non-Fraud</td>
</tr>
<tr>
<td></td>
<td>Company</td>
<td>Company</td>
<td></td>
<td>Company</td>
</tr>
<tr>
<td>Not Prompted</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(Std. Dev.)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>N</td>
<td>0.56 (0.50)</td>
<td>0.51 (0.50)</td>
<td>0.53 (0.50)</td>
<td>0.83 (0.38)</td>
</tr>
<tr>
<td>Prompted</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(Std. Dev.)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>N</td>
<td>0.57 (0.50)</td>
<td>0.47 (0.50)</td>
<td>0.52 (0.50)</td>
<td>0.72 (0.45)</td>
</tr>
<tr>
<td>Total</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(Std. Dev.)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>N</td>
<td>0.56 (0.50)</td>
<td>0.49 (0.50)</td>
<td>0.53 (0.50)</td>
<td>0.78 (0.42)</td>
</tr>
</tbody>
</table>

Panel B: Analysis Results

<table>
<thead>
<tr>
<th>Source</th>
<th>DF</th>
<th>$X^2$</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>EXP</td>
<td>1</td>
<td>10.13</td>
<td>&lt; 0.01</td>
</tr>
<tr>
<td>DECEP_COMPANY</td>
<td>1</td>
<td>8.34</td>
<td>&lt; 0.01</td>
</tr>
<tr>
<td>COGDIS</td>
<td>1</td>
<td>0.16</td>
<td>0.69</td>
</tr>
<tr>
<td>EXP*DECEP_COMPANY</td>
<td>1</td>
<td>2.27</td>
<td>0.13</td>
</tr>
<tr>
<td>EXP*COGDIS</td>
<td>1</td>
<td>0.51</td>
<td>0.48</td>
</tr>
<tr>
<td>DECEP_COMPANY*COGDIS</td>
<td>1</td>
<td>3.03</td>
<td>0.08</td>
</tr>
<tr>
<td>EXP* DECEP_COMPANY*COGDIS</td>
<td>1</td>
<td>4.15</td>
<td>0.04</td>
</tr>
<tr>
<td>AUDIO</td>
<td></td>
<td>3.32</td>
<td>0.07</td>
</tr>
</tbody>
</table>
Panel C: Planned Contrasts, Simple Effects, and Comparisons

<table>
<thead>
<tr>
<th>Source</th>
<th>DF</th>
<th>$X^2$ / t</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>H1: Accuracy rates will be highest for very experienced auditors assessing non-fraud companies (contrast weight=3), significantly lower for very experienced auditors assessing fraud companies (-1) and similarly low for inexperienced auditors assessing both fraud (-1) and non-fraud companies (-1).</td>
<td>1</td>
<td>10.84</td>
<td>&lt; 0.01*</td>
</tr>
<tr>
<td>EXP given Non-Fraud Company</td>
<td>1</td>
<td>3.04</td>
<td>&lt; 0.01*</td>
</tr>
<tr>
<td>DECEP_COMPANY given Very Experienced</td>
<td>1</td>
<td>2.34</td>
<td>0.01*</td>
</tr>
<tr>
<td>H2: -1.5 (Inexperienced / Non-Fraud / No Prompt), -1.5 (Inexperienced / Fraud / No Prompt), -1.5 (Inexperienced / Fraud / No Prompt), 2.5 (Experienced / Non-Fraud / No Prompt), 2.5 (Experienced / Non-Fraud / Prompt), -1.5 (Experienced / Fraud / No Prompt), 2.5 (Experienced / Fraud / Prompt)</td>
<td>1</td>
<td>17.46</td>
<td>&lt; 0.01*</td>
</tr>
<tr>
<td>H2: -1 (Inexperienced / Fraud / No Prompt), -1 (Inexperienced / Fraud / No Prompt), -1 (Inexperienced / Fraud / No Prompt), 3 (Experienced / Fraud / Prompt)</td>
<td>1</td>
<td>5.70</td>
<td>0.01*</td>
</tr>
<tr>
<td>Accuracy rate for very experienced auditor given a cognitive dissonance prompt and evaluating a fraud company, is greater than each of the other three accuracy rates evaluating a fraudulent company</td>
<td></td>
<td></td>
<td>all $p &lt; 0.03*$</td>
</tr>
</tbody>
</table>

* These p-values are for effects that occur in the expected direction suggested by our theory, and are therefore the one-tailed test of the signed t-statistic. Other reported p-values are two-tailed.

Very experienced and inexperienced auditors are identified using EXP, which is an indicator variable that equals 1 if the response was from a very experienced audit professional (24 years in audit on average), and 0 if the response was from an inexperienced accounting major enrolled in an auditing course. Fraud and non-fraud companies are identified using DECEP_COMPANY, which is an indicator variable that equals 1 if conference call pertains to a quarter that was eventually restated due to an irregularity (Fraud Company), and 0 otherwise (Non-Fraud Company). DECEP_JUDG is an indicator variable that equals 1 if the participant judges the conference call to be related to a company quarter that will eventually be restated due to an irregularity, and 0 otherwise. The dependent measure is ACCURACY, which is an indicator variable that identifies whether the participant correctly classifies the conference call: equals 1 if (DECEP_COMPANY = 1 and DECEP_JUDG = 1 or DECEP_COMPANY = 0 and DECEP_JUDG = 0) and 0 otherwise. Cognitive dissonance prompt is a between subjects manipulated variable that has two parts. First, participants receiving this manipulation are given additional initial instructions that said the following: “One cue found to be useful in detecting deception in these CEO responses is cognitive dissonance. Cognitive dissonance is the negative, uncomfortable emotion a person feels when they are saying something that they know is not true. Those experiencing cognitive dissonance feel uncomfortable, uneasy, and bothered.” Second, after answering our principal dependent measure, these participants are asked “how much cognitive dissonance the CEO felt during this excerpt of the conference call.” COGDIS = 1 when both of these components are present and 0 otherwise. AUDIO is an indicator variable that equals 1 if the conference call excerpt was provided both as a transcript and as audio, 0 if as a transcript only.
### TABLE 3
Very Experienced and Inexperienced Auditors’ Red Flagged Issue Notations for Fraud and Non-Fraud Companies

**Descriptive Statistics – Auditors’ Red Flagged Issues by EXP and DECEP_COMPANY**

<table>
<thead>
<tr>
<th>Variable</th>
<th>Non-Fraud Company</th>
<th>Fraud Company</th>
<th>t</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean (std. dev.)</td>
<td>Mean (std. dev.)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>% RF_ISSUES</td>
<td>% RF_ISSUES</td>
<td></td>
<td></td>
</tr>
<tr>
<td>N</td>
<td>348</td>
<td>358</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ISSUE_CHARS</td>
<td>240.39 (251.82)</td>
<td>243.34 (271.50)</td>
<td>0.21</td>
<td>0.83</td>
</tr>
<tr>
<td>RF_ISSUES</td>
<td>2.23 (1.81)</td>
<td>2.40 (2.12)</td>
<td>1.31</td>
<td>0.19</td>
</tr>
<tr>
<td>FIN</td>
<td>0.54 (0.78)</td>
<td>0.87 (1.06)</td>
<td>5.42</td>
<td>&lt; 0.01</td>
</tr>
<tr>
<td>SPEECH</td>
<td>0.98 (1.19)</td>
<td>0.86 (1.17)</td>
<td>1.55</td>
<td>0.12</td>
</tr>
<tr>
<td>SKNOW</td>
<td>0.71 (0.89)</td>
<td>0.67 (0.91)</td>
<td>0.55</td>
<td>0.58</td>
</tr>
</tbody>
</table>

**Very Experienced Auditors**

| N        | 58 | 63 |
| ISSUE_CHARS | 169.41 (188.18) | 310.68 (325.41) | 3.61 | < 0.01 |
| RF_ISSUES | 2.15 (2.17) | 3.91 (3.00) | 4.08 | < 0.01 |
| FIN      | 0.63 (0.98) | 1.24 (1.17) | 3.20 | < 0.01 |
| SPEECH   | 0.82 (1.26) | 1.27 (1.39) | 2.10 | 0.04 |
| SKNOW    | 0.70 (1.04) | 1.40 (1.59) | 3.13 | < 0.01 |

For each of the four companies they evaluated, we ask auditors to read (listen to) the transcript and write down the time stamp (line number from the transcript) of each red flag and as much detail about the red flag as possible. Two separate raters with knowledge of accounting but with no knowledge of the hypotheses being tested nor the experience level of the auditor providing the response, independently coded the open-ended notes auditors make when noting red flags. The coders counted the number of issues and sorted them into 14 categories. We average their responses together and then arrange the 14 categories into three broad groups: financial statement items (FIN), speech (SPEECH), and suspect knowledge (SKNOW). The FIN group contains red flagged issues pertaining to financial statements, financial results, company strategy, and topics regarding the appropriate application of judgment when applying accounting standards given the operating environment. The SPEECH group captures speech hesitations, filled pauses (ah’s and um’s), language use that avoids answering questions or deflects blame, and speech that suggests the speaker was nervous. The SKNOW group contains red flagged issues identifying instances where the executive appeared to lack knowledge of the subject matter by giving answers that seem incomplete, inaccurate, internally inconsistent, or lacking appropriate level of confidence. The sum of FIN, SPEECH and SKNOW represents the total number of red flagged issues identified (RF_ISSUES). ISSUE_CHARS is the number of characters provided in the text box from which the red flagged issues were coded. Very experienced and inexperienced auditors are identified using EXP, an indicator variable that equals 1 if the response was from a very experienced audit professional (24 years in audit on average), and 0 if the response was from an inexperienced accounting major enrolled in an auditing course. Fraud and non-fraud companies are identified using DECEP_COMPANY, an indicator variable that equals 1 if conference call pertains to a quarter that was eventually restated due to an irregularity (Fraud Company), and 0 otherwise (Non-Fraud Company). DECEP_JUDG is an indicator variable that equals 1 if the participant judges the conference call to be related to a company quarter that will eventually be restated due to an irregularity, and 0 otherwise. Test statistics and related p-values are adjusted for repeated measures.
### TABLE 4
The Effect of Audit Experience and Cognitive Dissonance Prompt on Accuracy of Red Flag Detection for Fraudulent Statements

#### Panel A: Descriptive Statistics – RF_ACCURACY by EXP and COGDIS

<table>
<thead>
<tr>
<th></th>
<th>Inexperienced</th>
<th>Experienced</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Not Prompted</td>
<td>0.17</td>
<td>0.21</td>
<td>0.18</td>
</tr>
<tr>
<td>(Std. Dev.)</td>
<td>(0.38)</td>
<td>(0.41)</td>
<td>(0.38)</td>
</tr>
<tr>
<td>N</td>
<td>1362</td>
<td>224</td>
<td>1586</td>
</tr>
<tr>
<td>Prompted</td>
<td>0.17</td>
<td>0.32</td>
<td>0.20</td>
</tr>
<tr>
<td>(Std. Dev.)</td>
<td>(0.38)</td>
<td>(0.47)</td>
<td>(0.40)</td>
</tr>
<tr>
<td>N</td>
<td>1290</td>
<td>241</td>
<td>1531</td>
</tr>
<tr>
<td>Total</td>
<td>0.17</td>
<td>0.27</td>
<td>0.19</td>
</tr>
<tr>
<td>(Std. Dev.)</td>
<td>(0.38)</td>
<td>(0.44)</td>
<td>(0.39)</td>
</tr>
<tr>
<td>N</td>
<td>2652</td>
<td>465</td>
<td>3117</td>
</tr>
</tbody>
</table>

#### Panel B: Analysis Results

<table>
<thead>
<tr>
<th>Source</th>
<th>DF</th>
<th>$X^2$</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>EXP</td>
<td>1</td>
<td>12.30</td>
<td>&lt; 0.01</td>
</tr>
<tr>
<td>COGDIS</td>
<td>1</td>
<td>3.20</td>
<td>0.08</td>
</tr>
<tr>
<td>EXP*COGDIS</td>
<td>1</td>
<td>1.99</td>
<td>0.16</td>
</tr>
</tbody>
</table>

#### Panel C: Planned Contrasts, Simple Effects, and Comparisons

<table>
<thead>
<tr>
<th>Source</th>
<th>DF</th>
<th>$X^2$ / t</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>The accuracy of very experienced auditors receiving the cognitive dissonance prompt is highest (contrast weight=3), significantly lower for very experienced auditors that are not prompted (-1), and similarly low for inexperienced auditors who are both prompted (-1) and not prompted (-1).</td>
<td>1</td>
<td>11.97</td>
<td>&lt; 0.01</td>
</tr>
<tr>
<td>COGDIS given Very Experienced</td>
<td>1</td>
<td>1.78</td>
<td>0.08</td>
</tr>
<tr>
<td>COGDIS given Inexperienced</td>
<td>1</td>
<td>0.43</td>
<td>0.67</td>
</tr>
</tbody>
</table>

Very experienced and inexperienced auditors are identified using EXP, which is an indicator variable that equals 1 if the response was from a very experienced audit professional (24 years in audit on average), and 0 if the response was from an inexperienced accounting major enrolled in an auditing course. Cognitive dissonance prompt is a between subjects manipulated variable that has two parts. First, participants receiving this manipulation are given additional initial instructions that said the following: “One cue found to be useful in detecting deception in these CEO responses is cognitive dissonance. Cognitive dissonance is the negative, uncomfortable emotion a person feels when they are saying something that they know is not true. Those experiencing cognitive dissonance feel uncomfortable, uneasy, and bothered.” Second, after answering our principal dependent measure, these participants are asked “how much cognitive dissonance the CEO felt during this excerpt of the conference call.” COGDIS = 1 when both of these components are present and 0 otherwise. The dependent measure, red flag detection accuracy (RF_ACCURACY), analyzed for only the five fraud companies, is constructed in three steps. First, each participant is given a 1 (0) if they identify (do not identify) a CEO’s sentence as a red flag. This is done for each CEO sentence. Second, each CEO sentence is classified as fraudulent or not fraudulent. This subjective classification is done by a coauthor and several teams of research assistants with no knowledge of the red flag results, using hand collected data about each of the five frauds. There were 3,117 fraudulent sentences (out of a total 12,293 sentences) analyzed by subject auditors. RF_ACCURACY measures sensitivity of auditors to fraudulent sentences, and equals 1 if the auditor correctly flags a fraudulent sentence as a red flag and 0 if the auditor does not flag a fraudulent sentence.