

NHH



WP 2017/6
ISSN: 2464-4005
www.nhh.no

WORKING PAPER

Mitigating the Dilution Effect in Auditors' Judgement Using a Frequency Response Mode

Aasmund Eilifsen, Natalia Kochetova and
William F. Messier, Jr.

Department of Accounting, Auditing and Law
Institutt for regnskap, revisjon og rettsvitenskap

NORWEGIAN SCHOOL OF ECONOMICS

Mitigating the Dilution Effect in Auditors' Judgments Using a Frequency Response Mode

Aasmund Eilifsen
NHH Norwegian School of Economics, Bergen, Norway

Natalia Kochetova*
Saint Mary's University, Halifax, Nova Scotia, Canada

William F. Messier, Jr.
University of Nevada, Las Vegas, Nevada, United States
NHH Norwegian School of Economics, Bergen, Norway

September 2017

Running head: Dilution and frequency mode

*Corresponding author:

Department of Accounting and Commercial Law
Sobey School of Business
Saint Mary's University
SB 318, 923 Robie Street
Halifax, Nova Scotia B3H 3C3
Canada
+1-902-420-5800
natalia.k@smu.ca

The authors would like to thank Chris Agoglia, Ken Dalton, Erin Hamilton, Gopal Krishnan, Marietta Peytcheva, Aaron Saiewitz, Sandra Shelton, Jason Smith, Scott Vandervelde, and the participants of the research workshops at University of Nevada, Las Vegas, University of Massachusetts-Amherst, Saint Mary's University, 2014 SDA Bocconi 5th Workshop on Audit Quality, and 2015 Midyear Auditing Meeting of American Accounting Association for their helpful comments. Professor Kochetova-Kozloski appreciates Saint Mary's University Faculty of Graduate Studies and Research Internal Grant Program for providing financial support for this research. Professor Messier received financial support for this research from the Kenneth and Tracy Knauss Endowed Chair in Accounting at UNLV and his Adjunct Professor position at NHH.

Mitigating the Dilution Effect in Auditors' Judgments Using a Frequency Response Mode

ABSTRACT

This paper investigates the potential of using a frequency response mode to reduce the dilution effect of non-diagnostic evidence on auditors' fraud risk judgments. In two experiments, we test one hypothesis and examine a research question related to the dilution effect where response mode (frequency versus probability) and type of non-diagnostic or irrelevant information are manipulated between participants. Results of the hypothesis tests show that auditors' fraud risk judgments demonstrate a significantly lower dilution effect when they evaluate diagnostic and non-diagnostic or irrelevant, evidence using a frequency response mode, as compared to the probability response mode; this effect is most pronounced when auditors are provided with favorable non-diagnostic or irrelevant evidence.

Keywords: Dilution; Fraud risk; Frequency response mode; Non-diagnostic information; Irrelevant cues.

JEL Classifications: M4, M40, M420.

Data Availability: Summary data are available from the authors upon request.

Mitigating the Dilution Effect in Auditors' Judgments Using a Frequency Response Mode

INTRODUCTION

During an audit, auditors search for, review, and evaluate substantial amounts of evidence. Some of this evidence may be diagnostic (relevant) to the judgment task while some of it may be non-diagnostic (or irrelevant).¹ Numerous studies in psychology and auditing have demonstrated that people's probability judgments are systematically influenced by irrelevant or non-diagnostic evidence (e.g., Fanning, Agoglia, and Piercey 2015; Glover 1997; Hackenbrack 1992; Hoffman and Patton 1997; Nisbett, Zukier, and Lemley 1981; Shanteau 1975; Zukier 1982). The basic findings of this research indicate that the presence of non-diagnostic evidence leads to a *dilution effect*; that is, individuals make less extreme (more regressive) judgments than those in the presence of diagnostic evidence only.² While most of this audit research was conducted in the 1990s, discussions with practitioners suggest that the amount of information (diagnostic, non-diagnostic, and irrelevant) that comes to an auditor has increased in recent years, mainly due to advances in technology (i.e., various general news sources, financial news sites, and social media) and changes in auditing standards, such as business risk auditing (Bell, Marrs, Solomon, and Thomas 1997; Bell, Peecher, and Solomon 2005). Brown-Liburd, Issa, and Lombardi (2016) argue that attention to irrelevant information has the potential to significantly limit the value that can be obtained from incorporating Big Data into the audit process.

Prior research in auditing by Hackenbrack (1992) shows that auditors are prone to the dilution effect. In Hackenbrack's study, auditors assessed how much a company's exposure to

¹ In our discussion of the background for Experiment 2, we make an argument for considering audit evidence on a continuum from diagnostic to non-diagnostic to irrelevant.

² The dilution effect has also been shown to occur in a legal context (Smith, Stasson, and Hawkes 1998/1999) and in consumer behavior (Meyvis and Janiszewski 2002). Other accounting studies that have examined the dilution effect include Reneau and Blanthorne (2001), Waller and Zimbelman (2003), Seow (2009), Wood (2012), and Fanning et al. (2015). Waller and Zimbelman (2003) provide a review of the dilution literature in accounting.

fraudulent reporting changed when presented with a mixture of diagnostic and non-diagnostic evidence. He found that the auditors' fraud risk judgments became less extreme in the presence of non-diagnostic evidence. Subsequent work in auditing by Hoffman and Patton (1997) and Glover (1997) examined whether accountability and time pressure eliminated or mitigated the dilution effect. Hoffman and Patton (1997, 228) showed that, “auditors' judgments exhibited the dilution effect both when they were held accountable and when they were not”. Glover (1997) also found that accountability had no influence on the dilution effect, but he observed that time pressure reduced it. Shelton (1999) noted that audit managers and partners are less susceptible to the dilution effect than senior auditors. Hoffman and Patton (1997) proposed several steps to mitigate the dilution effect, such as making auditors aware of the issue and developing decision aids. Lastly, Fanning et al. (2015) found that directional goals can ameliorate the dilution effect when investors are presented with post-decisional information in the context of litigation risk assessments. In the current study, we propose an alternative approach (i.e., use of a frequency response mode) to mitigate the dilution effect in auditors' fraud risk judgments.

Kochetova-Kozloski, Messier, and Eilifsen (hereafter KME) (2011) demonstrate that auditors' judgments about fraud can be improved when information is presented as natural frequencies instead of probabilities.³ By using a frequency response mode instead of a probability response mode, KME find that statistical reasoning within a Bayesian framework can be improved, especially in low base rate events (i.e., fraud). Their results show that the auditors' fraud likelihood judgments using a frequency response mode, as compared to a probability response mode, are closer to the Bayesian benchmark. Following the frequency approach appears to assist auditors in better use of information about the base rate of fraud in the population of

³ “Natural frequencies are absolute frequencies as encoded through direct experience and have not been normalized with respect to the base rates” of the event or non-event (Hoffrage and Gigerenzer 2004, 251).

clients with observed fraud risk. The current study examines whether using a frequency response mode can mitigate the dilution effect documented by earlier auditing studies. We use an approach conventionally used in psychology (LaBella and Koehler 2004) to test for dilution whereby information cues are introduced in a step-by-step fashion: first, diagnostic (or relevant) information cues are provided to participants; this is followed by an information processing task leading to an outcome judgment; next, non-diagnostic (or irrelevant) information cues are provided to participants, and they are asked to repeat the information processing task in light of new cues, resulting in an outcome judgment. The non-zero difference between outcome judgments based on each group of cues indicates dilution. This contrasts with an alternative approach to testing dilution where diagnostic (relevant) and non-diagnostic (irrelevant) cue are combined into bundles and dilution is then measured by examining the averaging effect of such cue bundles on outcome judgments (e.g., Fanning et al. 2015; Lambert and Peytcheva 2017).⁴

We conduct two experiments to test one hypothesis and investigate a research question related to the mitigation of the dilution effect in fraud risk assessments. In the first experiment, we manipulate response mode (frequency versus probability) and type of non-diagnostic information (neutral versus favorable versus unfavorable) in a 2 x 3 between-participants design. Our results show that auditors demonstrate a lower dilution effect when they use information in frequency response mode, as compared to probability response mode. We find that this effect is most pronounced when auditors are provided with favorable non-diagnostic evidence; an important result, since fraudsters are likely to present favorable information to mislead the auditor when fraud is present. We also find that when using a frequency response mode, the residual dilution effect is different between conditions with favorable and unfavorable non-

⁴ From an audit perspective, both approaches are consistent with the way auditors search and evaluate evidence (see Ashton and Ashton 1988; Knechel and Messier 1990; Tubbs, Messier, and Knechel 1990).

diagnostic cues, and between neutral and unfavorable cues. Based upon the results in Experiment 1, we conduct a second experiment to clarify issues related to the nature of non-diagnostic information presented to the participants. In the Experiment 2, based on prior work by Hoffman and Patton (1997), we use information cues that are, taken in isolation, irrelevant to fraud risk, as compared to cues rated as non-diagnostic by experts used in Experiment 1. Consistent with findings in Experiment 1, results of Experiment 2 show lower dilution effect using a frequency response mode in the presence of favorable irrelevant evidence. We also find differences between the residual dilutive effect of irrelevant evidence in the frequency response mode between conditions with favorable and unfavorable cues. Overall, the significance of these findings is that a simple manipulation that presents base rate information in a frequency format appears to mitigate, although does not eliminate, the dilution effect of non-diagnostic (or irrelevant) evidence. A more detailed analysis shows that this finding is driven by the responses to favorable non-diagnostic or irrelevant evidence. Reduction of the dilution effect in fraud risk judgment when using frequency response mode has the potential to prevent auditors from underestimating such risk when it exists, despite the positive information about the client that is not directly relevant to the likelihood of fraud.

The remainder of the paper is organized as follows. The next section provides background on the dilution effect and the development of the research hypothesis and research question. This is followed by a presentation of the two experiments. The last section presents a discussion of the results, limitations and areas for future research.

BACKGROUND AND HYPOTHESIS DEVELOPMENT

Background - Dilution Effect

The dilution effect is a judgment bias in which individuals underutilize relevant or diagnostic information when non-diagnostic or irrelevant information is also present, leading to outcome judgments that are less extreme or regressive. How might seemingly non-diagnostic or irrelevant information affect judgments in this way? The psychology literature proposes two competing accounts of the psychological mechanisms that underlie the dilution effect; namely a *perceptual* approach and a *conversational* approach (see Kemmelmeier 2004 for a review). The perceptual approach finds its basis in the early work of Nisbett et al. (1981) and Zukier (1982) who argue that such judgments are made based on the perceived similarity between a target and one's conception of the outcome group's characteristics. In short, the decision maker relies on the representativeness heuristic (Kahneman and Tversky 1973) in which a target is considered to be a member of an outcome group if it appears "representative" of the group. For example, an auditor is likely to provide a higher fraud risk judgment when presented with evidence that is diagnostic of (relevant to) management fraud than when presented with the same evidence combined with evidence that is not diagnostic of (irrelevant to) management fraud. Thus, the company's management will appear to be less representative of the target (fraud) group in the presence of non-diagnostic or irrelevant evidence, resulting in lower ("diluted") fraud risk judgments.

The conversational account of the dilution effect is quite different and assumes that dilution is an experimental artifact resulting from participants' mistaken reliance on conversational norms (Hilton 1995; Schwarz 1996; Tetlock and Boettger 1989). Following this line of reasoning, participants in such studies expect the experimenter to only communicate information relevant to the issue at hand. Thus, when presented with information by the experimenter, the participants infer that the experimenter considers all pieces of information to

be relevant and subsequently use diagnostic and non-diagnostic (or irrelevant) information in their judgment even though they themselves may or may not consider the information diagnostic (or relevant).⁵

In our hypothesis development and testing, we rely on the perceptual approach for three reasons. First, we believe that the type of audit evidence and the way it is presented to audit professionals is not subject to conversational norms that have been examined in psychology. Auditors are faced with a combination of diagnostic and non-diagnostic (or irrelevant) evidence daily in their professional work by virtue of their interactions with clients and the audit evidence-gathering process whereby they continually must discern relevant information from irrelevant information. Second, the perceptual approach has been the basis for prior auditing studies that examined the dilution effect (e.g., Glover 1997; Hackenbrack 1992; Hoffman and Patton 1997). Third, in the post-experimental questionnaire we measure participants' perceptions of whether information cues provided to them were diagnostic or non-diagnostic of fraud, and we use observations with only correctly classified diagnostic cues and progressively exclude observations with misclassified non-diagnostic (irrelevant) cues, thus incorporating an empirical basis into perceptual approach into our analyses.

Hypothesis and Research Question

In prior auditing research on the dilution effect, researchers required their participants to make probabilistic judgments about fraud (Hackenbrack 1992; Hoffman and Patton 1997; Wood 2012; Lambert and Peytcheva 2017), risk disclosure thresholds (Fanning et al. 2015), going-concern (Shelton 1999; Lambert and Peytcheva 2017), and to assess the risk of material misstatements in accounts receivable (Glover 1997; Waller and Zimbelman 2003). KME show

⁵ See Igou (2007) and Kemmelmeier (2007a, b) for an exchange on which approach has the most empirical support. Bonner (2008, 175) states that “theories explaining the dilution effect appear to be still under construction; accountants could help advance theory development.”

that using a frequency response mode can bring auditors' fraud risk judgments closer to the Bayesian solution, especially in the case that is most similar to fraud in the real world — a low base rate of occurrence (one percent). KME base their approach on research by Gigerenzer and his colleagues (e.g., Gigerenzer, Hoffrage, and Kleinbölting 1991; Gigerenzer and Hoffrage 1995) and others (Cosmides and Tooby 1994, 1996) who maintained that if people are asked to estimate the probability of a single event, the question does not connect to probability theory in their minds, whereas the frequency of such an event does (Gigerenzer and Goldstein 1996; Gigerenzer 2004). KME (2011, 839-840) argue that this is due to two reasons. First, Bayesian computations are cognitively simpler when information is encoded in a frequency format rather than in a probability format. Second, the estimation of the likelihood of a single event and the judgment of frequency are cognitively different processes (Cosmides and Tooby 1994, 1996; Gigerenzer et al. 1991).

A recently evolved theory in experimental psychology “hybrid load” theory explains why frequency format may ameliorate the dilution effect (Lavie, Hirst, de Fockert, and Viding 2004; Wilson, Muroi, and MacLeod 2011). This theory and associated empirical evidence suggest that the dilution effect due to irrelevant cues occurs because of limited-capacity processing in both perceptual and cognitive control systems. For tasks with high cognitive load, such as auditor's multi-cue risk judgment or evidence evaluation, the addition of irrelevant, or distractor, items increases working memory load, which in turn, “reduces the capacity available for active control and consequently reduces its ability to prevent distractors from affecting behavior” (Wilson et al. 2011, 333). In such situations, a decision-maker has a reduced capacity to segregate non-diagnostic (irrelevant, distractor) cues from diagnostic cues and classify them as information that is to be ignored (LaBella and Koehler 2004, 1085), thus allowing for the non-diagnostic cues to

influence judgment, along with diagnostic cues. If frequency response mode, as argued and demonstrated by KME (2011) based on Gigerenzer et al.'s studies, reduces working load demands via a simplified cognitive algorithm, then such "freed up" cognitive capacity may become available for improved classification of non-diagnostic cues, thus reducing their interference with an outcome judgment, and lowering the dilution effect (but not eliminating it). Therefore, we propose the following hypothesis:

H1: Auditors demonstrate a lower dilution effect when they receive case information and make required judgments in a frequency response mode as compared to a probability response mode.

In Experiment 1, we follow Hackenbrack (1992) and distinguish between three types of non-diagnostic evidence: favorable, unfavorable, and neutral. In the fraud-risk setting, *favorable* non-diagnostic evidence would be positive client information that does not relate directly to possible fraud. *Unfavorable* non-diagnostic evidence describes negative client information that does not directly relate to the presence of client fraud. *Neutral* non-diagnostic evidence includes client information that is neither positive nor negative and evaluated as unrelated to the presence of client fraud by the auditor. Hackenbrack (1992) hypothesizes that non-neutral (favorable and unfavorable) non-diagnostic evidence has a higher dilutive capacity than neutral non-diagnostic evidence. The logic behind this prediction is that non-neutral, non-diagnostic evidence is more salient and auditors will devote more attention to such evidence (e.g., Tversky 1977). Also, the literature in psychology indicates that neutral non-diagnostic evidence is more likely to be ignored (e.g., LaBella and Koehler 2004). Hackenbrack (1992) finds mixed results across the two versions of his task (increasing versus decreasing fraud risk). In the increasing fraud risk version, the non-neutral non-diagnostic evidence leads to more regressive fraud risk judgments (i.e., lower fraud risk judgments) than the neutral non-diagnostic evidence, thus indicating higher dilution effect. In the decreasing fraud risk version, he finds no difference between the regressive

effect of non-neutral, non-diagnostic evidence and such effect of neutral, non-diagnostic evidence. Hoffman and Paton (1997) likewise distinguish between favorable and unfavorable non-diagnostic information but find no differences in their dilutive effect. Based on prior research that has mixed findings for the diverse types of non-diagnostic evidence, we do not find a basis to predict whether the use of a frequency response mode leads to a differential residual dilution effect across the different types of non-diagnostic evidence. Therefore, we pose the following research question:

RQ: When using a frequency response mode, do auditors exhibit a dilution effect differentially across the types of non-diagnostic evidence?

EXPERIMENT 1⁶

Design

We used a 2 (*Response Mode*) \times 3 (*Type of Non-Diagnostic Evidence*) \times 2 (*Order*) between-participants design. *Response Mode (RM)* was manipulated at two levels: frequency response mode or probability response mode; the *Type of Non-Diagnostic Evidence (TYPE-EV)* was manipulated at three levels: neutral, favorable, or unfavorable; and *Order (ORDER)* of the non-diagnostic evidence cues was manipulated at two levels.⁷

Expert Measures of Evidence Diagnosticity

To develop a list of diagnostic and non-diagnostic evidence for use in the management fraud case, we identified forty-one fraud-related factors from auditing sources (AICPA 2002: AU 316; IAASB 2013: ISA 240).⁸ We tested this list of fraud-related factors for their diagnosticity

⁶ Both experiments were approved by an IRB/REB process in the US, Canada, and Norway.

⁷ There were three diagnostic and three non-diagnostic evidence cues in each instrument (see Appendices 2 and 3); we used two orders of non-diagnostic cues. *ORDER* was not significant at conventional levels ($p > 0.05$), and was therefore not included in the analyses.

⁸ Fraud-related factors also may either increase or decrease assessed fraud risk; fraud-risk factors (“red flags”) increase likelihood of fraud. In our list (see Appendix 1) we include both fraud risk-increasing and fraud-risk decreasing factors.

with a group of seven experienced managers employed by major public accounting firms in Norway (see Appendix 1). We selected three diagnostic fraud-related factors and nine non-diagnostic fraud-related factors (three each for neutral, favorable, and unfavorable) in accordance with their ratings. Appendix 2 presents the fraud-related factors included in the experiment by category of diagnosticity (diagnostic or non-diagnostic of fraud) and type of non-diagnostic evidence (unfavorable, neutral, favorable). It is evident from Appendix 1 that the experts do not always agree on the classification of the pieces of evidence into the four categories. Thus, while we attempted to rely on the experts' categorization, we had to use some degree of judgment in selecting the pieces of evidence for inclusion in the instrument. Five of the factors were selected as diagnostic evidence by all the experts (9, 12, 23, 35, and 40). We selected three of these diagnostic factors (9, 35, and 40) for inclusion in the experimental materials. For the non-diagnostic evidence, in general, we selected the type of evidence that was categorized most often as neutral, favorable, or unfavorable - but not diagnostic of fraud risk. For example, item 8 "Client management attempts to justify improper accounting on the basis of materiality" was tied with 14 "Non-financial client management participates in determination of significant estimates" in being listed as non-diagnostic - unfavorable. However, item 8 was listed as diagnostic by 4 of the managers but has a low rating of likelihood of fraud (mean = 2). Additionally, some of the more frequently rated non-diagnostic types of evidence were omitted because we were uncertain if they were relevant for our participant pool.

Case Materials and Procedure

The basic management fraud case was adapted from KME and is presented in Appendix 3. The first page of the instrument informed the participants that the study involved estimating the likelihood of management fraud in a client company and that all the participating individuals

and their respective firms would remain anonymous. The second page contained the informed consent form. Participants were then presented with the management fraud case and three diagnostic fraud-related factors. Next, the participants assessed the level of fraud using either a frequency or probability response mode. On the following page, the participants were told, “your assistant indicates that the clients selected from the database also possess the following characteristics IN ADDITION to the ones indicated on the previous page.” The participants were then presented with three non-diagnostic fraud-related factors that were all neutral, favorable, or unfavorable; and asked to reassess the likelihood of management fraud using the same response mode. The presentation of the diagnostic cues followed by the non-diagnostic cues is consistent with the belief revision procedure followed in prior psychology research (LaBella and Koehler 2004; Troutman and Shanteau 1977).⁹

Next, the participants were asked to rate the diagnostic and non-diagnostic cues in the same manner as the expert managers. That is, we measured participants’ perceptions of whether information cues provided to them were, in the participants’ opinion, diagnostic or non-diagnostic of fraud, and if they were favorable, unfavorable, or neutral.¹⁰ Lastly, the participants were asked a series of demographic questions. The two of the researchers administered the experiment during a regular class session for two annual groups of graduate auditing students, many with audit experience.

Dependent Variables

We developed two measures of the dependent variables. First, we used the absolute value of the deviation of the participants’ response (assessed likelihood of fraud) in the presence of

⁹ Auditors continually receive evidence during an audit, and it would not be unrealistic to receive diagnostic evidence followed by non-diagnostic evidence.

¹⁰ A strength of the current study is that, unlike prior studies in auditing that examine the dilution effect, we use an expert panel to classify the cues *ex ante*, and then measure, *ex post*, whether the participants evaluate the evidence as consistent with its manipulation.

only the diagnostic evidence from the respective Bayesian response: $F-DEV = |\text{Participant's Fraud Likelihood Response} - \text{Bayesian Fraud Likelihood Response}|$.¹¹ This dependent variable is used to replicate KME's result in their low base rate condition (one percent). We believe it is necessary to replicate KME's result before proceeding to test the hypothesis. Second, to test H1 and examine the research question, we used the signed differences between the participants' fraud likelihood judgment after evaluating both diagnostic and the additional non-diagnostic evidence and the earlier judgment following only the diagnostic evidence ($F-REV = \text{Participants' Fraud Likelihood Response after receiving both diagnostic and non-diagnostic evidence} - \text{Participants' Fraud Likelihood Response after receiving only diagnostic evidence}$) for *each* response mode. This dependent variable accounts for both direction and the magnitude of revisions in participants' fraud likelihood judgments.

Participants

The participants in the experiment were auditors and students in the Master's in Accounting and Auditing (MRR) program at the Norwegian School of Economics (NHH) in Bergen, Norway. Completion of the MRR program, passing a rigorous examination, and three years of practice experience allow a candidate to be eligible for state authorization (the highest level of certification). This is the same population of auditors as used previously by KME, but a different sample.

One hundred and eighty-one (181) questionnaires were returned, of which one hundred seventy-four (174) were complete. We then eliminated sixty-six (66) observations by applying the following screens: (1) we first eliminated fifty-eight (58) participants who classified any of the diagnostic cues as not indicative of fraud, and (2) we eliminated eight (8) participants who

¹¹ To have a uniform scale measure for the dependent variables for statistical tests, we converted percentages and frequencies to proportions in the probability and frequency conditions, respectively.

classified all three non-diagnostic cues as diagnostic. Thus, the final sample included one hundred and eight (108) graduate students, where seventy (70) were practicing auditors.

Participants' mean age was 27.33 years; 65 participants (60.2 percent) were male. Thirty-four participants (32 percent) had a prior master's degree (other than in auditing), while all others had a bachelor's degree; 24 participants (18 percent) held the lower level of certification while 84 participants (78 percent) were in the process of obtaining their first professional designation. The practicing auditor participants had an average of 30.96 months of audit experience.¹² Of the 70 auditor participants, 35 (50 percent) were senior auditors, 33 (47 percent) were staff or associates, and 2 participants (3 percent) were managers. Most of the auditor participants (57, or 81 percent of auditor subsample) worked for a Big 4 firm at the time of the experiment. Based on their education and experience mix, participants in our sample should be familiar with the simplified fraud risk judgment task used in this study.

Replication of KME

Prior to testing the main hypothesis, we replicate KME's results.¹³ The case used in this study is similar to KME's low base rate scenario (one percent). The one major difference is the inclusion of the three pieces of diagnostic evidence. Since the same data are included in both response modes, the results of our test should show that the use of a frequency response mode leads to responses that are closer to the Bayesian benchmark.

We compute the absolute difference for each participant's response from the Bayesian benchmark for both the frequency and probability response modes in the presence of only

¹² Audit experience is not significant in any of our analyses; therefore, we do not include it as covariate in reporting our results. Analyses performed on a sub-sample of practicing auditors do not produce qualitatively different findings from those reported below.

¹³ This replication is not pure because KME's (2011) experimental instrument contains no individuating information about the client, whereas the instrument used here does. However, the purpose of this replication is nonetheless achieved as the goal is to demonstrate the bias-ameliorating impact of frequency response mode on judgments which have been shown by prior research to be subject to base-rate neglect when no non-diagnostic (irrelevant) cues are supplied to the participants.

diagnostic evidence. The Bayesian benchmarks for the frequency and probability response modes are 0.0776 and 0.0767, respectively (KME 2011, 846). Our tests show that for the low base rate (one percent), the absolute deviations from the Bayesian benchmark are smaller in the frequency response mode (marginal mean =0.293)¹⁴ than in the probability response mode (marginal mean =0.536) ($F = 18.069, p = 0.000$, one-tailed, not tabled). This result replicates KME and allows us to use the current dataset for testing the study's main hypothesis and examining the research question.

Tests of the Hypothesis and Research Question

We test our hypothesis using the signed differences between the participants' two fraud likelihood judgments. Table 1 presents the descriptive data on the dependent variables. Table 2 presents the ANOVA using the signed revisions *F-REV*. We test H1 by examining whether participants using a frequency response mode demonstrate a lower dilution effect, as compared to participants using a probability response mode. As shown in Table 2 (Panels A and B), there is a significant main effect for response mode (*RM*) ($F=3.231, p=.038$, one-tailed), thus supporting H1.¹⁵ The mean signed revision for the frequency response mode (-0.038) is significantly lower than the mean signed revision for the probability response mode (-0.102). The main effect for type of evidence is also significant (*TYPE-EV*) ($F=9.285, p=.000$) while the interaction $RM \times TYPE-EV$ ($F=2.778, p=.076$, two-tailed) is marginally significant. The simple main effects show that *RM* is significant when type of non-diagnostic evidence is favorable ($F=6.310, p=.009$, one-tailed; marginal mean -0.254 in probability response mode versus -0.083 in frequency response mode), but it is not significant when *TYPE-EV* is neutral or unfavorable ($p=.247$ and $p=.250$, both one-tailed, respectively). Thus, our results show that the use of a frequency response mode

¹⁴ Note that this mean is significantly different from zero ($t=6.808, p=.000$, two-tailed) indicating that the participants still exhibited a significant base rate neglect. This result is also consistent with KME (2011, 853).

¹⁵ As our predictions are directional, we use one-tailed p -values where appropriate.

only reduced the dilution effect in the presence of favorable non-diagnostic evidence. However, this finding is significant since clients are more likely to present auditors with evidence that is favorable for the client but not diagnostic of an underlying problem. For example, Messier, Simon, and Smith (2013, 167) note that when investigating differences identified by analytical procedures “managers may unintentionally or intentionally provide non-misstatement business explanations for a significant difference that has resulted from an error or fraud.” Such explanations are similar to favorable, non-diagnostic evidence.

Insert Tables 1 and 2 about here

The research question asked whether the magnitude of the dilution effect differs across types of non-diagnostic evidence in the frequency response mode. The results of the 1×3 ANOVA (not tabled) indicate that *TYPE-EV* is significant ($F=3.905$, $p=.026$, two-tailed). Follow-up tests indicate that the magnitude of dilution effect does not differ between favorable and neutral non-diagnostic evidence ($p=.998$, two-tailed; marginal means -0.083 and -0.089 , respectively); however, there is a significantly different regressive (dilutive) effect of non-diagnostic evidence between conditions with favorable and unfavorable cues ($F= 6.046$, $p=.020$, two-tailed, marginal means $-.083$ and $.051$, respectively), and between neutral and unfavorable cues ($F= 7.151$, $p=.011$, two-tailed, marginal means -0.089 and 0.051 , respectively). We conclude that in frequency response mode auditors exhibit dilution effect when presented with favorable or neutral non-diagnostic cues (as indicated by negative marginal means), which is statistically significantly different from their revisions of fraud risk assessments upwards in the presence of unfavorable diagnostic cues (i.e., reverse to dilution effect).

Sensitivity Analyses

We test robustness of our results by performing several sensitivity analyses. First, we employ an alternate specification of the dependent variable, an absolute revision of the fraud likelihood judgments after the introduction of non-diagnostic cues: $F\text{-}ABSREV = |\text{Participants' Fraud Likelihood Response after receiving both diagnostic and non-diagnostic evidence} - \text{Participants' Fraud Likelihood Response after receiving only diagnostic evidence}|$. This is a more conservative version of the dependent variable as it does not take into account direction of revision, only the magnitude, thereby truncating the distribution. We reperform the tests of H1 and research question. The results are reported in Table 3 (Panels A and B). *RM* remains significant ($F=6.918, p=.005$, one-tailed), with the marginal mean in frequency response mode lower than in probability response mode (0.072 versus 0.158, respectively). *RM* x *TYPE-EV* interaction is also significant ($F=3.762, p=.027$, two-tailed). Simple main effects tests show that *RM* is significant when non-diagnostic evidence is favorable (not tabled, $F=7.190, p=.006$, one-tailed; marginal means are 0.083 and 0.260 in frequency and probability modes, respectively), which is consistent with our findings using *F-REV* as a dependent variable. Also, consistent with results reported earlier, *RM* is not significant in condition with neutral non-diagnostic cues ($F=.504, p=.241$, one-tailed). However, we observe a statistically significant difference between frequency (marginal mean 0.051) and probability (marginal mean 0.162) response modes in the cells with unfavorable non-diagnostic cues (not tabled, $F=3.289, p=.040$, one-tailed). Thus, when we use *F-ABSREV* as a dependent variable, H1 is supported in conditions with favorable and unfavorable, but not neutral, non-diagnostic cues.

Insert Table 3 about here

Our results for testing the research question (not tabled) indicate that *TYPE-EV* is not significant ($F=.234, p=.792$, two-tailed; marginal means are 0.083, 0.084, and 0.051 in favorable,

neutral, and unfavorable conditions, respectively). Therefore, results regarding the research question are more sensitive to specification of the dependent variable, indicating that they are affected by the direction of revision.

We reperform our tests of H1 and research question using a reduced sample that includes only observations with all correctly classified diagnostic and non-diagnostic cues. That is, we apply the strictest screen by excluding all participants who rated *any* of the three non-diagnostic cues as diagnostic. This procedure results in a sample with 63 observations.¹⁶ Tables 4 and 5 present the results. When we use *F-REV* as a dependent variable, *RM* is not significant (Table 4, Panel A: $F=.962$, $p=.166$, one-tailed), *TYPE-EV* is significant ($F=2.387$, $p=.051$, one-tailed), and the interaction is not significant ($p=.118$, two-tailed). We do not find statistically significant differences between signed measures of revisions across the three types of non-diagnostic cues in frequency response mode ($F=.688$, $p=.511$, two-tailed, not tabled). When we use an absolute measure of the dependent variable, *F-ABSREV*, our results are consistent with previously reported tests of H1: *RM* is significant (Table 5, Panels A and B: $F=3.812$, $p=.028$, one-tailed; marginal means are 0.072 and 0.182 in frequency and probability response modes, respectively). We find a marginally significant interaction $RM \times TYPE-EV$ (Table 5, Panel A: $F=2.883$, $p=.064$, two-tailed). Post hoc test indicate that *RM* is significant ($F=6.596$, $p=.008$, one-tailed, not tabled) when non-diagnostic evidence is favorable, such that dilution effect is lower in frequency response mode (marginal mean for *F-ABSREV* is .088) than in probability response mode (marginal mean for *F-ABSREV* is .227). We do not find an effect for *RM* when non-diagnostic cues are neutral ($F=.637$, $p=.217$, one-tailed, not tabled) or unfavorable ($F=1.808$, $p=.108$, one-tailed, not tabled). Therefore, we find support for H1 when non-diagnostic evidence

¹⁶ We also conducted an analyses by excluding observations with 2 or more incorrectly classified non-diagnostic cues from the sample (with $n=94$), our results are identical to those reported earlier for the main sample ($n=108$).

is favorable on the reduced sample using *F-ABSREV* as a dependent variable. When we test whether the mean dilution is different between different types of non-diagnostic evidence in frequency response mode, we do not find significant differences (all p 's > .05).

Insert Tables 4 and 5 about here

On balance, our sensitivity tests indicate that our results are not sensitive to alternative specifications of the dependent variable (*F-REV* vs *F-ABSREV*) for the main sample as we continue to find support for H1 in the cell with favorable cues (both dependent variables), as well as in the cell with unfavorable cues (*F-ABSREV*). Our findings regarding the research question also appear to be robust to alternative specifications of the dependent variable. However, when we exclude all observations from the sample in which we detect any disagreement with experts in terms of classification of both diagnostic and non-diagnostic cues, our support for H1 is limited to absolute (unsigned) specification of the dependent variable in presence of favorable cues.

EXPERIMENT 2

Background

Based on Experiment 1 it is apparent that both our experts and participants disagreed on the diagnosticity of the evidence cues, especially evidence that was expected to be non-diagnostic of fraud risk. In rethinking the logic behind the selection of diagnostic versus non-diagnostic cues, we reason that, in an audit setting, there is likely a continuum for the diagnosticity of evidence.¹⁷ Hackenbrack (1992, 127) relies on psychology research (Nisbett et al. 1981; Zukier 1982; and Tetlock and Boettger 1989) to explain the dilution effect in terms of auditors' reliance on a similarity-based inference process. In his view, “auditors judge the

¹⁷ Both Hackenbrack (1992) and Hoffman and Patton (1997) allude to the diagnosticity of evidence running along a continuum.

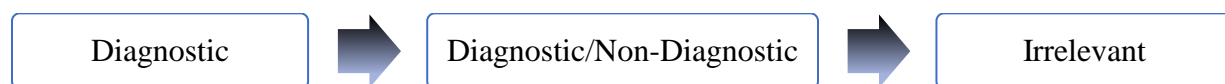
likelihood of some event by comparing their knowledge of the client (client-specific knowledge) with their knowledge of the conditions that produce the event (event-specific knowledge). The greater the perceived similarity of the two sources of knowledge, the greater the assessed likelihood of the event.”¹⁸ When performing a specific task, such as evaluating the possibility of fraud, the auditor may obtain client-specific knowledge. This client-specific knowledge may be diagnostic or non-diagnostic of the presence of fraud. However, as pointed out in the literature (e.g., Nisbett et al. 1981, 273), we cannot exclude the possibility that any information is perceived as diagnostic or non-diagnostic but only argue what is plausible. This issue is relevant in an auditing setting, especially since auditing guidance generally draws attention to possible symptoms of events (e.g., fraud, going concern uncertainty, unreasonable accounting estimates) that normally turn out to be false positive (Type 1 error) (AICPA 2002: AU 316; IAASB 2013: ISA 240).

Auditing standards and audit firm guidance list fraud risk factors, or “red flags,” that cover a broad range of situations believed to be potential indicators of a fraudulent act. However, prior research has documented that the presence of many of these risk factors is not necessarily indicative of fraud (Hogan, Rezaee, Riley, and Velury 2008, 7). For example, Bell and Carcello (2000) find that some of the fraud risk factors cited in auditing standards and cited by auditors (Loebbecke and Willingham 1988) are not very predictive of fraud (e.g., compensation arrangements based on recorded performance, high management turnover, and misstatements detected in prior period’s audit). The predictive ability of red flags remains debatable. Their value has not been documented as especially effective in research or practice; rather, they seem to corroborate malfeasance once discovered (Trompeter, Carpenter, Jones, and Riley 2014, 792).

¹⁸ In a similar vein, Hoffman and Patton (1997, 228) state, “auditors will be performing other audit tasks as they think about fraud; in addition to encountering information relevant to the fraud risk assessment, auditors will simultaneously encounter information relevant to other audit tasks but irrelevant to fraud.”

In the design of Experiment 1, we started out with forty-one (41) fraud-related factors drawn from existing auditing standards (AICPA 2002: AU 316; IAASB 2013: ISA 240) and the research and professional literature. We then examined the experts panel's ratings of those factors that were perceived as diagnostic or non-diagnostic for the presence of fraud.

Nonetheless, it may be possible that auditors view some non-diagnostic factors as diagnostic; experts' and participants' ratings of the evidence cues confirmed this in Experiment 1. It is also plausible that an auditor could receive client-specific information that is clearly irrelevant to the fraud risk judgment. These arguments suggest that the "diagnosticity" of information may fall along the follow continuum:



The diagnostic end of the continuum would contain information that is clearly relevant to the specific fraud event; i.e., it is a robust "red flag" indicating increased likelihood of fraud. We can think of the fraud risk factors identified by Bell and Carcello (2000) (and those clearly rated by our experts) as falling at this end of the continuum. The middle part of the continuum would contain information that some auditors might view as diagnostic and others - as non-diagnostic - given the client circumstances. There are many fraud-related factors identified in auditing standards that auditors believe to be diagnostic - but which are not (e.g., see Hogan et al. 2008; Trompeter et al. 2014); indeed, Bell and Carcello (2000) and our experts' and participants' ratings confirm that this is possible. Finally, at the other end of the continuum, would be information that has no relevance to the event (fraud risk) being judged. In Experiment 2, we use information cues from the diagnostic and irrelevant ends of the continuum to test our hypothesis and investigate our research question.

Design

We used a 2 (*Response Mode*) \times 2 (*Type of Irrelevant Evidence*) between-participants design. *Response Mode* (*RM*) was manipulated at two levels: frequency response mode or probability response mode while the *Type of Irrelevant Evidence* (*TYPE-EV*) was manipulated at two levels: favorable or unfavorable. The orders of presentation of the diagnostic evidence cues and of the irrelevant evidence cues were randomized.

Case Materials and Procedure

We used the same basic case as in Experiment 1, except that we replaced the nine non-diagnostic cues with three favorable and three unfavorable *irrelevant* cues. These six cues were selected from the four favorable and four unfavorable irrelevant cues used by Hoffman and Patton (1997).¹⁹ Appendix 2 contains the six irrelevant cues. We used the same format of presentation of the diagnostic risk factors, followed by the irrelevant factors, as in Experiment 1. The participants made fraud likelihood judgments using either a frequency or probability response mode. Next, they were asked to rate the diagnostic evidence and irrelevant cues in the same manner as Experiment 1. That is, we measured participants' perceptions of whether information cues provided to them were, in the participants' opinion, relevant or irrelevant to the likelihood of fraud, and if they were favorable or unfavorable. Lastly, the participants were asked a series of demographic questions.

The experiment was conducted online using *Qualtrics* and was administrated by the director of MRR program. The first group of participants consisted of graduate students in their last semester of the MRR program. The announcement was made about voluntary participation in the fraud risk judgment study during class. Next, a questionnaire in *Qualtrics* was distributed to the students by email and completed outside the class setting. The second group of participants

¹⁹ We thank Vicky Hoffman for providing access to her experimental information. One of her irrelevant cues was not considered because it related to IT and was outdated. The six cues used in Experiment 2 were considered appropriate for the experiment.

consisted of practicing auditors in each of the Big Four firm in Bergen, Norway. A contact person in each office distributed the questionnaire by email to the firm's auditors. The email included a cover letter like the one given to student participants. The firms were asked to send the email to auditors with more than one year of audit experience.

Dependent Variables

Similar to Experiment 1, we developed two measures of the dependent variable. First, we used the absolute value of the deviation of the participants' response (assessed likelihood of fraud) in the presence of only the diagnostic evidence from the respective Bayesian response: $F-DEV = |\text{Participant's Fraud Likelihood Response} - \text{Bayesian Fraud Likelihood Response}|$. This dependent variable is used to replicate KME's result in their low base rate condition (one percent). Second, we used the signed differences between the participants' fraud likelihood judgment after evaluating both diagnostic and the additional irrelevant evidence and the earlier judgment following only the relevant evidence ($F-REV = \text{Participants' Fraud Likelihood Response after receiving both relevant and irrelevant evidence} - \text{Participants' Fraud Likelihood Response after receiving only relevant evidence}$) to test the hypothesis and examine the research question.

Participants

Ninety-five (95) participants came from the class of students in the MRR program at NHH; after deleting observations where participants withdrew from the experiment (3) or provided incomplete responses (34), we obtained fifty-eight (58) usable responses. These responses came from twenty-six (26) participants who had audit experience and thirty-two (32) participants without audit experience. A second group of participants was obtained from Norwegian Big 4 audit firms. One hundred and nine (109) auditors participated in the

experiment, providing sixty-one (61) usable responses. We eliminated four (4) who chose not to participate and forty-three (43) unfinished responses. Thus, our total sample in Experiment 2 consists of one hundred nineteen (119) participants, of which eighty-seven (87) had audit experience and thirty-two (32) did not. Consistent with the approach used in Experiment 1, we eliminated participants who classified all three irrelevant cues as diagnostic (nine).²⁰ This resulted in a final sample of one hundred and ten observations (110), of which twenty-eight (28) were students with no audit experience and eighty-two (82) were practicing auditors.

Participants' mean age was 32.39 years; 78 participants (71 percent) were male. Forty-nine participants (45 percent) had a prior master's degree, while all others had a bachelor's degree; 54 participants (49 percent) held a certification while 56 participants (51 percent) were in the process of obtaining a professional designation. The practicing auditor participants had an average of 8.38 years of audit experience.²¹ Of the 82 auditor participants, 36 (44 percent) were senior auditors, 14 (7 percent) were managers, 12 (15 percent) were senior managers, 11 (13 percent) were partners, and 5 (1 percent) were staff auditors.²² Most of the auditor participants (79, or 96 percent of the subsample) worked for a Big 4 firm at the time of the experiment. Thus, our sample in Experiment 2 has a greater level of audit experience than in Experiment 1.

Replication of KME

We again replicate KME's results by computing the absolute difference for each participant's response from the Bayesian benchmark for both the frequency and probability response modes in the presence of only diagnostic evidence (KME 2011, 846). Our tests show that for the low base rate (one percent), the absolute deviations from the Bayesian benchmark are

²⁰ None of the participants classified diagnostic cues as irrelevant to the likelihood of fraud.

²¹ Similarly, to Experiment 1, when we control for audit experience, it is not significant in any of our analyses; therefore, we do not include it as covariate in reporting our results. Analyses performed on a sub-sample of practicing auditors do not produce qualitatively different findings from those reported below.

²² Four participants marked their position as "other."

smaller in the frequency response mode (marginal mean =0.468)²³ than in the probability response mode (marginal mean =0.592) ($F = 3.325, p = 0.036$, one-tailed, not tabled). This result replicates KME and allows us to use the current dataset for testing the study's main hypothesis.

Tests of the Hypothesis and Research Question

We test our hypothesis using signed differences between the participants' two fraud risk judgments. Table 6 presents the descriptive data on the dependent variables. Table 7 presents the ANOVA using the signed revisions (*F-REV*). We test H1 by examining whether participants using a frequency response mode demonstrate a lower dilution effect, as compared to participants using a probability response mode. As shown in Table 7 (Panels A and B), the main effect for response mode is not significant (*RM*) ($F=0.176, p=.338$, one-tailed). The mean signed revision for the frequency response mode is lower, but not significantly so, than the mean signed revision for the probability response mode (marginal means 0.009 versus 0.011). The main effect for type of evidence is significant (*TYPE-EV*) ($F=7.922, p=.006$, two-tailed), as is the interaction $RM \times TYPE-EV$ ($F=4.676, p=.033$, two-tailed). The simple main effects show that *RM* is significant when type of irrelevant evidence is favorable ($F=2.611, p=.006$, one-tailed, marginal mean 0.009 in probability response mode vs. -0.005 in frequency). When *TYPE-EV* is unfavorable, *RM* is marginally significant but in the "wrong" direction; i.e., the marginal mean in probability mode is lower than in the frequency mode ($F=2.030, p=.080$, one-tailed marginal mean 0.013 in probability response mode vs. 0.022 in frequency). Thus, our results in testing H1 using signed revision as a dependent show that the use of a frequency response mode only reduced the dilution effect in the presence of favorable irrelevant evidence; in the presence of unfavorable evidence we observed the opposite effect. However, the finding in cells with

²³ Note that this mean is significantly different from zero ($t=9.378, p=.000$, two-tailed) indicating that the participants still exhibited a significant base rate neglect. This result is also consistent with KME (2011, 853) and with Experiment 1.

favorable irrelevant evidence is consistent with the results of Experiment 1. We continue to believe it is noteworthy because it addressed the situations where clients present auditors with evidence that is favorable for the client but irrelevant to specific instance of fraud.

Insert Tables 6 and 7 about here

To examine the research question, we compare the magnitude of dilution effect between cells with favorable and unfavorable irrelevant evidence in the frequency response mode. Results (not tabled) show that *TYPE-EV* is significant ($F=12.776$, $p=.001$, two-tailed), indicating that dilutive effect of irrelevant evidence is larger for cell with unfavorable cues (marginal mean 0.022) than for the cell with favorable cues (marginal mean -0.005). The finding of this difference is consistent with results of Experiment 1; that is, frequency response mode with irrelevant unfavorable cues appears to increase fraud risk assessments and produce opposite-to-dilution effect.

Sensitivity Analyses

We test robustness of these results using the same approach as for Experiment 1.²⁴ We reperform our tests of H1 and research question using a reduced sample excluding observations with incorrectly classified irrelevant cues. Similar to the approach used in Experiment 1, we first eliminated *all* observations where participants rated *any* irrelevant cues as relevant; this procedure results in a sample of twenty-three (23) observations. Since twenty-three observations are not sufficient to perform statistical tests of H1 required for a 2×2 ANOVA, we applied a

²⁴ When we employ an alternate specification for the dependent variable based on absolute, as compared to signed, revision of the fraud likelihood judgments after the introduction of irrelevant cues, our statistical tests do not show a significant effect for *RM*, *TYPE-EV*, or their interaction (not tabled, all p 's $>.05$). Likewise, we do not find a significant difference between average absolute revisions in the presence of favorable (marginal mean .014) versus unfavorable (marginal mean 0.022) irrelevant cues in frequency response mode ($F=1.344$, $p=.251$, two-tailed, not tabled). Therefore, we conclude that our results for Experiment 2 are sensitive to the specification of the dependent variable; when the dependent variable does not take into account direction of the revision but only considers the magnitude (as in *F-ABSREV*) we are not able to find support for results we demonstrated with the dependent variable which takes into account *both* the direction and the magnitude of the revision (as in *F-REV*).

slightly less conservative screen and eliminated all observations where two or more irrelevant cues were rated as relevant to fraud risk; with this screen we obtained a reduced sample of seventy-two (72) observations. Table 8 presents the results. When we use *F-REV* as a dependent variable, *RM* is not significant (Table 8, Panel A: $F=.138$, $p=.356$, one-tailed), *TYPE-EV* is significant ($F=4.185$, $p=.045$, two-tailed), and the interaction is also significant ($F=4.103$, $p=.047$, two-tailed). Post-hoc tests indicate that *RM* is marginally significant ($F=2.565$, $p=.059$, one-tailed, not tabled) when irrelevant evidence is favorable, such that the dilution effect is lower in frequency response mode (marginal mean for *F-REV* is -0.008) than in probability response mode (marginal mean for *F-REV* is 0.009). This difference is not significant at conventional levels when irrelevant evidence cues are unfavorable ($F=1.780$, $p=.096$, one-tailed, not tabled; marginal means 0.009 and 0.020 in probability and frequency response mode, respectively). These results are consistent with our Experiment 2 findings reported earlier on the main sample; that is, we find support for H1 only when irrelevant evidence is favorable.

Insert Table 8 about here

When we examine research question on the reduced sample of 72, we find a statistically significant difference between signed measures of revisions between the cells with favorable and unfavorable cues ($F=6.614$, $p=.014$, two-tailed, not tabled). The marginal mean of *F-REV* in the cell with favorable cues (.007) is significantly lower than in the cell with unfavorable cues (.020). Again, this result is consistent with Experiment 2 findings reported earlier using the main sample.²⁵

DISCUSSION, LIMITATIONS, AND FUTURE RESEARCH

²⁵ Finally, when we use an absolute measure of the dependent variable, *F-ABSREV*, we do not find statistically significant results for *RM*, *TYPE-EV*, or their interaction (not tabled, all p 's $>.05$). Likewise, we do not find statistically significant difference between cells with favorable and unfavorable irrelevant cues in frequency response mode (not tabled, $p>.05$). This is consistent with our earlier results using the main sample for Experiment 2.

Discussion

In this paper, we set out to examine whether the dilution effect identified in prior research can be mitigated by a frequency response mode. If successful, this simple approach to representing probability information to auditors may mitigate a bias in risk judgment that has been shown to be extremely robust to various settings (Hackenbrack 1992; Hoffman and Patton 1997; Glover 1997). Our results indicate that while the use of a frequency response mode reduced the dilution effect, this finding is driven by the responses to cases where non-diagnostic (or irrelevant) evidence is favorable. Using signed revisions of their sequential risk assessments as dependent variables, we found that participants' judgments were less influenced by favorable non-diagnostic (Experiment 1) or favorable irrelevant (Experiment 2) evidence in a frequency than in a probability response mode. This is an important finding since clients who are committing fraud are likely to present favorable (non-diagnostic, irrelevant) explanations or evidence to an auditor's inquiry about fraud.

We also find that in frequency response mode auditors react differentially to the types of non-diagnostic evidence when we focus on signed revisions of risk assessments. Specifically, we find that auditors demonstrate an increase in fraud risk assessments, opposite-to dilution-effect, in response to unfavorable non-diagnostic cues, rather than reducing or regressing, their risk assessments. This change is statistically different from their reaction to favorable or neutral non-diagnostic cues (in Experiment 1). One potential explanation of this finding is that contemporary auditors are more sensitized to negative information due to a greater emphasis on professional skepticism in recent years, and thus may have developed greater affective reaction to unfavorable cues (Bhattacharjee, Moreno, and Riley 2012). We believe that this is the first study to document such a result and it warrants further investigation. Another potential explanation of this

result is that our participants followed a conversational approach to non-diagnostic (irrelevant) cues (Hilton 1995; Schwarz 1996; Tetlock and Boettger 1989); in other words, they assumed them to be diagnostic notwithstanding cue ratings provided in the debriefing questionnaire. Overall, our results indicate the potential of the frequency response mode to mitigate, but not eliminate, the dilution effect of the non-diagnostic or irrelevant favorable evidence.

Limitations

This study is subject to several limitations. First, we used a stylized case adapted from prior literature to measure risk assessments. Real-world fraud risk assessment scenarios include a multitude of diagnostic and non-diagnostic cues. Second, we relied on individual auditor's judgments to measure fraud risk assessments and to construct dependent variables. In the field, such judgments are made by the audit team, often with input from experts, such as forensic specialists. Third, due to the necessity of maintaining experimental control, we did not incorporate a fraud brainstorming session or an opportunity to perform additional research about the client into the fraud judgment task. Fourth, in testing dilution effect, we used a belief revision approach from psychology research (LaBella and Koehler 2004; Troutman and Shanteau 1977) where diagnostic (relevant) cues are followed by the non-diagnostic (irrelevant) cues. Results may have been different if used various types of aggregation of diagnostic (relevant) and non-diagnostic (irrelevant) cues within cells and compared such cells between participants (e.g., see Fanning et al. 2015; Lambert and Peytcheva 2017). Fifth, we conducted a joint test of frequency format of information and frequency format of questions (Von Sydow 2011).

Future Research

Overall, our results indicate that a frequency response mode reduces Type I error in that it ameliorates the under-assessment of fraud risk in light of positive, but non-diagnostic or

irrelevant, information cues about the client. Although our response mode manipulation did not eliminate dilution effect, it ameliorated it for cases with favorable non-diagnostic or irrelevant evidence. Future research should investigate *why* the dilution effect appeared to be unaffected by response mode when non-diagnostic or irrelevant cues were neutral or unfavorable. Also, it is not clear, based on our research and on prior literature, why auditors appeared to increase their fraud risk assessments in response to unfavorable cues in frequency response mode. Further, one could argue that the approach we have chosen, albeit routed in relevant literature in psychology, yields limited degree of generalization to the audit setting because auditors may encounter cues simultaneously over the course of their audit work, and that it may be more natural to test whether risk assessments would differ in a setting with a mixture of diagnostic and non-diagnostic cues versus a baseline setting where only diagnostic cues are provided. Testing robustness of our findings in such a setting presents another venue for future research. Researchers may extend our work by considering, separately, the impact of frequency information and frequency questions. Additional possible research venue is examining the role of various types and length of experience and expertise in exhibiting and mitigating dilution. Finally, one could examine the incremental effect of time pressure and other contextual features of the audit process on reduction in dilution effect in frequency response mode. Overall, ways of reducing dilution effect in situations where multiple types of cues and multiple features of the audit environment are at play simultaneously, as in the current Big Data environment, is a fruitful area for future research.

APPENDIX 1 **Experiment 1: Expert Panel's Ratings of the Fraud-Related Factors (n=7)**

Information Cues ^a	Response ^b					
	(a)		(b)	(c)		
	Yes <i>n</i>	No <i>n</i>	Mean <i>n</i> =7 ^c	Unfavorable <i>n</i>	Neutral <i>n</i>	Favorable <i>n</i>
1. Client operates in a highly saturated market, accompanied by declining profit margins.	5	2	2.00	1	1	-
2. The client management is interested in increasing the entity's stock price.	6	1	2.17	1	-	-
3. The client company conducts significant operations across international borders, in countries with differing business and cultural environments.	5	2	2.20	1	1	-
4. Client management receives bonuses based on achievement of aggressive targets for operating results.	6	1	3.00	1	-	-
5. When dealing with auditors, management attempts to influence the scope of the audit work.	6	1	2.33	-	1	-
6. The client company has an internal audit department. The head of the department reports to the Audit Committee of the Board of Directors.	-	7	-	1	1	5
7. The client company is subject to new accounting requirements.	2	5	2.00	-	5	-
8. Client management attempts to justify improper accounting on the basis of materiality.	4	3	2.00	3	-	-
9. The client is involved in related-party transactions not in the ordinary course of business.	7	-	3.00	-	-	-
10. The client exhibits a marginal ability to meet debt covenant requirements.	6	1	3.00	1	-	-
11. Most of the client employees have adequate skills and knowledge to perform their jobs.	1	6	1.00	1	3	2
12. The client organization has a very complex structure. It is difficult to determine what individuals have control of the entity.	7	-	2.86	-	-	-
13. The client has a history of disputes with predecessor auditor on accounting matters.	6	1	3.00	1	-	-
14. Non-financial client management participates in determination of significant estimates.	3	4	2.67	3	1	-
15. Client information technology staff has high turnover rate.	1	6	2.00	4	2	-
16. Management has clear objectives in terms of budgets, profits, and other financial and operating goals.	1	6	4.00	-	3	3
17. The client needs to obtain additional debt or equity financing to maintain the current level of research and development programs.	5	2	2.40	-	2	-
18. The client company operates in an industry undergoing rapid changes in technology and product obsolescence.	5	2	2.60	2	-	-
19. There is excessive pressure on client's operating personnel to meet financial targets						-

set up by management, including sales incentive goals.	6	1	2.83	1	-	
20. Client management investigates and documents deviations from established controls.	1	6	2.00	1	1	4
21. Client management has a practice of committing to analysts to achieve aggressive forecasts.	5	2	3.20	1	1	-
22. Management has an effective means of communicating ethical values to employees.	-	7	-	1	2	5
23. Client has significant bank accounts in tax-heavens jurisdictions for which there appears to be no clear business justification.	7	-	3.71	-	-	-
24. Client company uses electronic data interchange (EDI) that allows for transmittal of transactions with suppliers over telecommunications network.	-	7	-	-	5	2
25. Client management places time constraints on the auditor regarding the issuance of the auditor's report.	3	4	1.33	1	3	-
26. Several client executives have heavy concentrations of their personal net worth in the entity.	4	3	3.25	1	2	-
27. Client management has communicated very optimistic earnings trend expectations in a few recent press releases.	4	3	3.50	1	2	-
28. Client management has been responsive to the prior recommendations from its auditors.	-	7	-	-	2	5
29. There is a class action lawsuit filed under securities law pending against the client.	6	1	3.17	1	-	-
30. Revenue recognition from several long-term projects is based on estimates that are difficult to corroborate.	5	2	2.60	2	-	-
31. Client management team is dominated by a single person who appears to be a strong leader.	4	3	1.75	1	2	-
32. The client company has been reporting negative cash flows from operations while reporting positive, growing earnings.	6	1	3.17	1	-	-
33. The major assets of the client company are inventory; property, plant, and equipment, and its customer mailing list.	2	5	2.50	-	5	-
34. The client industry is undergoing significant decline in customer demand and increasing business failures.	4	3	3.25	1	2	-
35. Close to year-end, the client was involved in several highly complex transactions that pose difficult "substance over form" questions.	7	-	3.14	-	-	-
36. In the past, client management failed to correct reportable conditions on a timely basis.	3	4	3.33	3	1	-
37. All of the client's accounting functions and data processing are centralized.	1	6	1.00	-	4	2
38. If the client company reports poor financial results during the upcoming year, it may lose a large governmental contract.	6	1	3.00	1	-	-
39. The client company has demonstrated rapid growth during the last year, especially when compared to that of the other	5	2	3.00	-	2	-

companies in the same industry.						
40. The client accounting personnel has attempted to place informal restrictions on auditors that limit their access to information.	7	-	3.71	-	-	-
41. The client entity has a performance measurement system that includes financial and non-financial indicators.	4	3	1.75	-	2	1

^a Diagnostic cues included were 9, 35 and 40. Non-diagnostic cues included were favorable (6, 20, and 22), neutral (7, 11, and 37), and unfavorable (8, 14, and 15).

^b Managers were asked the following questions about each information cue:

(a) Does this characteristic point to the possibility of financial statement fraud in a client company?

☐ Yes ☐ No

(b) If you answered “yes” to (a), how likely is the client to be involved in financial statement fraud?

(Use a scale below to check a number corresponding to your answer.)

Not Likely			Somewhat		Very
At All			Likely		Likely
1	2	3	4	5	
<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

(c) If you answered “no” to (a), is this client characteristic: favorable, neutral, or unfavorable in your typical assessment of the client?

☐ Unfavorable ☐ Neutral ☐ Favorable

^c The mean is the likelihood of fraud on a scale from 1 to 5 for those experts answering “Yes” to the possibility of fraud.

APPENDIX 2

Evidence Cues Used in Experiments 1 and 2

Diagnostic cues (fraud-related factors) used in Experiments 1 and 2:

- The client's accounting personnel have attempted to place informal restrictions on auditors that limit their access to information.
 - The client is involved in related party transactions that are not in the ordinary course of business.
 - Close to year-end, the client was involved in several highly complex transactions that pose difficult "substance over form" questions.
-

Non-diagnostic cues used in Experiment 1:

Non-diagnostic unfavorable cues (non-diagnostic of fraud, but unfavorable in nature):

- Non-financial management at these clients participates in determination of significant estimates.
- Client managements attempted to justify improper accounting on the basis of materiality.
- Clients' information technology staff has high turnover rate.

Non-diagnostic neutral cues (non-diagnostic of fraud, but neutral in nature):

- All of the client's accounting functions and data processing are centralized.
- The client company is subject to new accounting requirements.
- Most of the client employees have adequate skills and knowledge to perform their job.

Non-diagnostic favorable cues (non-diagnostic of fraud, but favorable in nature):

- The client company has an internal audit department. The head of the department reports to the Audit Committee of the Board of Directors.
 - Management has an effective means of communicating ethical values to employees.
 - Client management investigates and documents deviations from established controls.
-

Irrelevant cues used in Experiment 2:

Irrelevant unfavorable cues (irrelevant to fraud, but unfavorable in nature):

- Due to dissatisfaction with the quality and level of service that it was receiving, the client switched advertising agencies.
- Management and labor representatives indicate that there is a possibility of a strike in the coming year.
- The client's patent on a unique product feature has expired.

Irrelevant favorable cues (irrelevant to fraud, but favorable in nature):

- A change in the local tax rate structure caused a decrease in the client's property taxes.
- The benefit plans made available to client employees are more generous than the industry average.

- The client has devoted resources to developing methods of recycling by-products of its production process.

APPENDIX 3

Experimental Materials: Management Fraud Case

Frequency Response Mode

A public accounting firm has assembled a large database of management descriptions as a part of a study intended to determine whether management descriptions are helpful in detecting management fraud. Upon completion of the database, your firm determined that the rate of material management fraud in the overall client base is 10 out of 1000. In addition, the firm estimated that if a client were involved in material management fraud, the description of the client management would match the “fraudulent profile” from a database in eight (8) of ten (10) cases. However, the firm estimates that in ninety-five (95) out of 990 cases, the descriptions in the database will identify the client management as fraudulent when it is, in fact, honest.

100 clients were selected at random from the firm’s client database. The database indicates the descriptions of all of these client’s managements matches the “fraudulent profile.” In addition, it contains the following specifically noted characteristics about these clients:

Probability Response Mode

A public accounting firm has assembled a large database of management descriptions as a part of a study intended to determine whether management descriptions are helpful in detecting management fraud. Upon completion of the database, the firm determined that the rate of material management fraud in the overall client base is 1%. In addition, the accounting firm estimated that if a client were involved in material management fraud, the description of the client management would match the “fraudulent profile” from a database in 79% of cases. However, the firm estimates that in 9.6% of cases the descriptions in the database will identify the client management as fraudulent when it is, in fact, honest.

A client was selected at random from the firm’s client database. The database indicates the description of this client’s management matches the “fraudulent profile.” In addition, it contains the following specifically noted characteristics about this client:

(Three diagnostic (relevant) fraud-related factors presented here; they were the same in Experiments 1 and 2.)

Question:

Frequency Response Mode: How many clients out of these 100 do you expect to be engaged in material management fraud?
_____ out of 100.

Probability Response Mode: What is the probability that the management of this client is engaged in material fraud? _____%

Presented on the next page of the instrument:

After further research, your assistant indicates that the clients selected from the database also possess the following characteristics IN ADDITION to the ones indicated on the previous page:

(Three neutral, favorable, or unfavorable non-diagnostic cues/three favorable or unfavorable irrelevant cues presented here.)²⁶

Question:

Frequency Response Mode: Given this additional information, how many clients out of these 100 do you expect to be engaged in material management fraud? _____ out of 100.

Probability Response Mode: Given this additional information, what is the probability that the management of this client is engaged in material fraud? _____%

²⁶ Experiment 1 materials contained neutral, favorable, or unfavorable non-diagnostic cues; Experiment 2 materials contained favorable or unfavorable irrelevant cues.

REFERENCES

- American Institute of Certified Public Accountants (AICPA). 2002. *AU Section 316: Consideration of Fraud in a Financial Statement Audit*. Available at <http://pcaobus.org/Standards/Auditing/Pages/AU316.aspx>
- Bell, T., F. Marrs, I. Solomon, and H. Thomas. 1997. *Auditing Organizations through a Strategic Systems Lens: The KPMG Business Measurement Process*. New York, NY: KPMG Peat Marwick LLP.
- Bell, T. B., M. E. Peecher, and I. Solomon. 2005. *The 21st Century Public Company Audit: Conceptual Elements of KPMG's Global Audit Methodology*. New York, NY: KPMG International.
- Bhattacharjee, S., K. K. Moreno, and T. Riley. 2012. The interplay of interpersonal affect and source reliability on auditors' inventory judgments. *Contemporary Accounting Research* 29 (4): 1087-1108.
- Bonner, S. E. 2008. *Judgment and Decision Making in Accounting*. Upper Saddle River, NJ: Pearson Prentice Hall.
- Brown-Liburd, H., H. Issa, and D. Lombardi. 2015. Behavioral Implications of Big Data's Impact on Audit Judgment and Decision Making and Future Research Directions. *Accounting Horizons* 29 (2): 451-468.
- Cosmides, L., and J. Tooby. 1994. Origins of domain specificity: The evolution of functional organization. In *Mapping the Mind: Domain-specificity in Cognition and Culture*, eds. L.
- Cosmides, L., and J. Tooby. 1996. Are humans good intuitive statisticians after all? Rethinking some conclusions from the literature on judgment under uncertainty. *Cognition* 58 (1): 1-73.
- Fanning, K., C. P. Agoglia, and M. D. Piercey. 2015. Unintended consequences of lowering disclosure thresholds. *The Accounting Review* 90(1): 301-320.
- Gigerenzer, G. 2004. Fast and frugal heuristics: The tools of bounded rationality. In *Blackwell Handbook of Judgment and Decision Making*, eds. D. Koehler and N. Harvey: 62-88. Oxford, UK: Blackwell.
- Gigerenzer, G., and D. G. Goldstein. 1996. Reasoning the fast and frugal way: Models of bounded rationality. *Psychological Review* 103 (4): 650-69.
- Gigerenzer, G., and U. Hoffrage. 1995. How to improve Bayesian reasoning without instruction: Frequency formats. *Psychological Review* 102 (4): 684-704.

- Gigerenzer, G., U. Hoffrage, and H. Kleinbölting. 1991. Probabilistic mental models: A Brunswikian theory of confidence. *Psychological Review* 98 (4): 506-28.
- Glover, S. M. 1997. The Influence of time pressure and accountability on auditors' processing of nondiagnostic information. *Journal of Accounting Research* 35 (2): 213-26.
- Hackenbrack, K. 1992. Implications of seemingly irrelevant audit evidence in judgment. *Journal of Accounting Research* 30 (1): 126-36.
- Hilton, D. J. 1995. The social context of reasoning: Conversational inference and rational judgment. *Psychological Bulletin* 118: 248-71.
- Hoffman, V. B., and J. M. Patton. 1997. Accountability, the dilution effect, and conservatism in auditors' fraud judgments. *Journal of Accounting Research* 35 (2): 227-37.
- Hoffrage, U., and G. Gigerenzer. 2004. How to improve the diagnostic inferences of medical experts. In *Experts in Science and Society*, eds. E. Kurz-Milcke and G. Gigerenzer: 249-68. New York: Kluwer/Plenum.
- Hogan, C., Z. Rezaee, R. Riley, and U. Velury. 2008. Financial statement fraud: Insights from the academic literature. *Auditing: A Journal of Practice & Theory* 27 (2): 231-252.
- International Auditing and Assurance Standards Board (IAASB). 2013. *ISA 240, The Auditor's Responsibilities Relating to Fraud in and Audit of the Financial Statements*. Available at <http://www.ifac.org/publications-resources/2013-handbook-international-quality-control-auditing-review-other-assurance-a>
- Igou, E. R. 2007. Additional thoughts on conversational and motivational sources of the dilution effect. *Journal of Language and Social Psychology* 26 (1): 61-68.
- Kahneman D. and A. Tversky. 1973. On the psychology of prediction. *Psychological Review* 80 (4): 237-251.
- Kemmelmeier, M. 2004. Separating the wheat from the chaff: Does discriminating between diagnostic and nondiagnostic information eliminate the dilution effect? *Journal of Behavioral Decision Making* 17: 231-243.
- Kemmelmeier, M. 2007a. Does the dilution effect have a conversational basis? *Journal of Language and Social Psychology* 26 (1): 48-60.
- Kemmelmeier, M. 2007b. Is diagnostic evidence on the dilution effect weakened when nondiagnostic objections are added? A response to Igou. *Journal of Language and Social Psychology* 26 (1): 69-74.

- Kochetova-Kozloski, N., W. F. Messier, Jr., and A. Eilifsen. 2011. Improving auditors' fraud judgments using a frequency response mode. *Contemporary Accounting Research* 28 (3): 837-58.
- LaBella C., and D. J. Koehler. 2004. Dilution and confirmation of probability judgments based on nondiagnostic evidence. *Memory & Cognition* 32 (7): 1076-89.
- Lambert, T.A., and M. Peytcheva. 2017. When does a stronger information set result in weaker audit judgments? An exploration of averaging in audit evidence integration. Working paper, Lehigh University.
- Lavie, N., A. Hirst, J. W. de Fockert, and E. Viding. 2004. Load theory of selective attention and cognitive control. *Journal of Experimental Psychology: General* 133 (3): 339-54.
- Messier, W. F., Jr., C. A. Simon, and J. L. Smith. 2013. Two decades of behavioral research on analytical procedures: What have we learned? *Auditing: A Journal of Practice & Theory* 32 (1): 139-181.
- Meyvis, T., and C. Janiszewski. 2002. Consumers' beliefs about product benefits: The effect of obviously irrelevant product information. *Journal of Consumer Research* 28: 618-635.
- Nisbett, R. E., H. Zukier, and R. E. Lemley. 1981. The dilution effect: Non-diagnostic information weakens the implications of diagnostic information. *Cognitive Psychology* 13: 248-277.
- Reneau, J. H., and C. Blanthorne. 2001. Effects of information sequence and irrelevant distractor information when using a computer-based decision aid. *Decision Sciences* 32 (1): 145-163.
- Seow, J.-L. 2009. Cue usage in financial statement fraud risk assessments: effects of technical knowledge and decision aid use. *Accounting & Finance* 49: 183-205.
- Shanteau, J. 1975. Averaging versus multiplying combination rules of inference judgment. *Acta Psychologica* 39: 83-89.
- Shelton, S. W. 1999. The effect of experience on the use of irrelevant evidence in auditor judgment. *The Accounting Review* 74 (2): 217-224.
- Smith, H. D., M. F. Stasson, and W. G. Hawkes. 1998/1999. Dilution in legal decision making: Effect of non-diagnostic information in relation to amount of diagnostic evidence. *Current Psychology* 17 (4): 333-345.
- Tetlock, P. E., and R. Boettger. 1989. Accountability: A social magnifier of the dilution effect. *Journal of Personality and Social Psychology* 57: 388-398.

- Trompeter, G. M., T. D. Carpenter, K. L. Jones, and R. A. Riley, Jr. 2014. Insights for Research and Practice: What We Learn about Fraud from Other Disciplines. *Accounting Horizons* 28 (4): 769–804.
- Troutman, C., and J. Shanteau. 1977. Inferences based on nondiagnostic information. *Organizational Behavior and Human Performance* 19: 43-55.
- Tversky, A. 1977. Features of Similarity. *Psychological Review* 84 (4): 327-352.
- Von Sydow, M. 2011. The Bayesian logic of frequency-based conjunction fallacies. *Journal of Mathematical Psychology* 55: 119-39.
- Waller, W. S., and M. F. Zimbelman. 2003. A cognitive footprint in archival data: Generalizing the dilution effect from laboratory to field settings. *Organizational Behavior and Human Decision Processes* 91: 254-268.
- Wilson, D. E., M. Muroi, and C. M. MacLeod. 2011. Dilution, not load, affects distractor processing. *Journal of Experimental Psychology: Human Perception and Performance* 37 (2): 319-355.
- Wood, L. I. 2012. The impact of decision aid use on the dilution effect when assessing fraud. *Journal of Finance and Accountancy* (March): 23-42.
- Zukier, H. 1982. The dilution effect: The role of the correlation and the dispersion of predictor variables in the use of nondiagnostic information. *Journal of Personality and Social Psychology* (December): 1163-1174.

TABLE 1
Experiment 1: Means of the Dependent Variables by Experimental Condition (Standard Deviation)^a

Response Format	Type of Non-diagnostic Cues	<i>Fraud Risk Assessment</i>		<i>Revision of Fraud Risk Assessment</i>	
		Initial <i>FRI</i>	Revised <i>FR2</i>	Signed <i>F-REV</i>	Absolute <i>F-ABSREV</i>
<i>Frequency</i>	Favorable <i>n</i> =15	.2565 (.3446)	.1738 (.2611)	-.0827 (.1758)	.0827 (.1758)
	Neutral <i>n</i> =25	.2918 (.3595)	.2090 (.3209)	-.0828 (.1732)	.0836 (.1728)
	Unfavorable <i>n</i> =16	.4244 (.3858)	.4750 (.4185)	.0506 (.1231)	.0506 (.1231)
	Overall Mean <i>n</i> =56	.3202 (.3631)	.2756 (.3556)	-.0447 (.1697)	.0739 (.1589)
<i>Probability</i>	Favorable <i>n</i> =18	.6966 (.1446)	.4430 (.2658)	-.2536 (.2088)	.2602 (.1999)
	Neutral <i>n</i> =17	.4771 (.3683)	.4247 (.3658)	-.0524 (.0649)	.0524 (.0649)
	Unfavorable <i>n</i> =17	.6240 (.2890)	.6238 (.3191)	-.0002 (.2722)	.1621 (.2148)
	Overall Mean <i>n</i> =52	.6011 (.2890)	.4961 (.3250)	-.1050 (.2268)	.1602 (.1911)
	Column Mean <i>n</i> =108	.4554 (.3570)	.3818 (.3572)	-.0737 (.2006)	.1155 (.1796)

^a The values in each cell are the mean (standard deviation) for each of the following variables: *Fraud Risk Assessment* = the raw scores of participants' assessment of the likelihood of fraud under each response mode; initial, in presence of diagnostic cues only (*FRI*); and revised, in the presence of both diagnostic and non-diagnostic cues (*FR2*). *Revision of Fraud Risk Assessment* = the signed (*F-REV*) and absolute (*F-ABSREV*) difference between the participants' raw scores of revised fraud risk assessment (*FR2*) and the participants' raw scores of initial fraud risk assessment (*FRI*).

TABLE 2
Experiment 1: Tests of H1: Signed Revision (*F-REV* as a Dependent Variable)

Panel A: Analysis of variance
(*n*=108)

Source	SS	df	<i>F</i>	<i>p</i> -value ^a
Intercept	.518	1	15.654	.000
<i>RM</i>	.107	1	3.231	.075
<i>TYPE-EV</i>	.614	2	9.285	.000
<i>RM</i> x <i>TYPE-EV</i>	.184	2	2.778	.076
Error	3.373	102		

RM = *RESPONSE MODE* = (probability vs. frequency)

TYPE-EV = *TYPE OF NON-DIAGNOSTIC EVIDENCE* = (favorable vs. neutral vs. unfavorable)

^a all *p*-values are two-tailed

Panel B: Estimated marginal means by *RM* and *TYPE-EV* (standard error)

		<i>TYPE-EV</i>			
		Favorable <i>n</i> =33	Neutral <i>n</i> =42	Unfavorable <i>n</i> =33	Row Mean <i>n</i> =108
<i>RM</i>	Frequency <i>n</i> = 56	-.083 (.047)	-.083 (.036)	.051 (.045)	-.038 (.025)
	Probability <i>n</i> = 52	-.254 (.043)	-.052 (.044)	.000 (.044)	-.102 (.025)
	Column Mean <i>n</i> = 108	-.168 (.032)	-.068 (.029)	.025 (.032)	-.070 (.018)

TABLE 3
Experiment 1: Sensitivity Analysis for Tests of H1: Absolute Revision (*F-ABSREV* as a Dependent Variable)

Panel A: Analysis of variance
(*n*=108)

Source	SS	df	<i>F</i>	<i>p</i> -value ^a
Intercept	1.397	1	49.800	.000
<i>RM</i>	.194	1	6.918	.010
<i>TYPE-EV</i>	.195	2	3.479	.035
<i>RM</i> x <i>TYPE-EV</i>	.211	2	3.762	.027
Error	2.861	102		

RM = *RESPONSE MODE* = (probability vs. frequency)

TYPE-EV = *TYPE OF NON-DIAGNOSTIC EVIDENCE* = (favorable vs. neutral vs. unfavorable)

^a all *p*-values are two-tailed

Panel B: Estimated marginal means by *RM* and *TYPE-EV* (standard error)

		<i>TYPE-EV</i>			
		Favorable	Neutral	Unfavorable	Row Mean
		<i>n</i> =33	<i>n</i> =42	<i>n</i> =33	<i>n</i> =108
<i>RM</i>	Frequency	.083	.084	.051	.072
	<i>n</i> = 56	(.043)	(.033)	(.042)	(.023)
	Probability	.260	.052	.162	.158
	<i>n</i> = 52	(.039)	(.041)	(.041)	(.023)
Column Mean		.171	.068	.106	.115
<i>n</i> = 108		(.029)	(.026)	(.029)	(.017)

TABLE 4
Experiment 1: Sensitivity Analysis for Tests of H1 on Reduced Sample: Signed Revision (*F-REV* as a Dependent Variable)

Panel A: Analysis of variance (<i>n</i> =63)				
Source	SS	df	<i>F</i>	<i>p</i> -value ^a
Intercept	.487	1	11.267	.001
<i>RM</i>	.042	1	.962	.331
<i>TYPE-EV</i>	.206	2	2.387	.101
<i>RM</i> x <i>TYPE-EV</i>	.192	2	2.222	.118
Error	2.462	57		

RM = *RESPONSE MODE* = (probability vs. frequency)

TYPE-EV = *TYPE OF NON-DIAGNOSTIC EVIDENCE* = (favorable vs. neutral vs. unfavorable)

^a all *p*-values are two-tailed

Panel B: Estimated marginal means by *RM* and *TYPE-EV* (standard error)

		<i>TYPE-EV</i>			
		Favorable <i>n</i> =30	Neutral <i>n</i> =23	Unfavorable <i>n</i> =10	Row Mean <i>n</i> =63
<i>RM</i>	Frequency <i>n</i> = 31	-.088 (.058)	-.126 (.056)	.002 (.104)	-.071 (.044)
	Probability <i>n</i> = 32	-.272 (.050)	-.064 (.069)	-.050 (.085)	-.129 (.040)
	Column Mean <i>n</i> = 63	-.180 (.038)	-.095 (.044)	-.024 (.067)	-.100 (.030)

TABLE 5
Experiment 1: Sensitivity Analysis for Tests of H1 on Reduced Sample: Absolute Revision
(*F-ABSREV* as a Dependent Variable)

Panel A: Analysis of variance
(*n*=63)

Source	SS	df	<i>F</i>	<i>p</i> -value ^a
Intercept	.792	1	20.411	.000
<i>RM</i>	.148	1	3.812	.056
<i>TYPE-EV</i>	.101	2	1.301	.280
<i>RM</i> x <i>TYPE-EV</i>	.224	2	2.883	.064
Error	2.213	57		

RM = *RESPONSE MODE* = (probability vs. frequency)

TYPE-EV = *TYPE OF NON-DIAGNOSTIC EVIDENCE* = (favorable vs. neutral vs. unfavorable)

^a all *p*-values are two-tailed

Panel B: Estimated marginal means by *RM* and *TYPE-EV* (standard error)

		<i>TYPE-EV</i>			
		Favorable <i>n</i> =30	Neutral <i>n</i> =23	Unfavorable <i>n</i> =10	Row Mean <i>n</i> =63
<i>RM</i>	Frequency <i>n</i> = 31	.088 (.055)	.126 (.053)	.002 (.000)	.072 (.041)
	Probability <i>n</i> = 32	.272 (.048)	.064 (.066)	.210 (.080)	.182 (.038)
	Column Mean <i>n</i> = 63	.180 (.036)	.095 (.042)	.106 (.064)	.127 (.028)

TABLE 6
Experiment 2: Means of the Dependent Variables by Experimental Condition (Standard Deviation)^a

Response Format	Type of Non-diagnostic Cues	<i>Fraud Risk Assessment</i>		<i>Revision of Fraud Risk Assessment</i>	
		Initial <i>FRI</i>	Revised <i>FR2</i>	Signed <i>F-REV</i>	Absolute <i>F-ABSREV</i>
<i>Frequency</i>	Favorable <i>n</i> = 32	.4575 (.4179)	.4525 (.4088)	-.0050 (.2999)	.0138 (.0270)
	Unfavorable <i>n</i> =29	.5990 (.4022)	.6210 (.4088)	.0221 (.0291)	.0221 (.0291)
	Overall Mean <i>n</i> =61	.5248 (.4133)	.5326 (.4141)	.0079 (.0323)	.0177 (.0281)
<i>Probability</i>	Favorable <i>n</i> =22	.6845 (.3108)	.6936 (.3061)	.0091 (.0334)	.0146 (.0313)
	Unfavorable <i>n</i> =27	.6328 (.3488)	.6454 (.3507)	.0126 (.0193)	.0126 (.0193)
	Overall Mean <i>n</i> =49	.6560 (.3299)	.6670 (.3289)	.0110 (.0263)	.0135 (.0251)
	Column Mean <i>n</i> =110	.5832 (.3824)	.5925 (.3824)	.0093 (.0297)	.0158 (.0268)

^a The values in each cell are the mean (standard deviation) for each of the following variables: *Fraud Risk Assessment* = the raw scores of participants' assessment of the likelihood of fraud under each response mode; initial, in presence of relevant cues only (*FRI*); and revised, in the presence of both relevant and irrelevant cues (*FR2*).
Revision of Fraud Risk Assessment = the signed (*F-REV*) and absolute (*F-ABSREV*) difference between the participants' raw scores of revised fraud risk assessment (*FR2*) and the participants' raw scores of initial fraud risk assessment (*FRI*).

TABLE 7
Experiment 2: Tests of H1: Signed Revision (*F-REV* as a Dependent Variable)

Panel A: Analysis of variance
(*n*=110)

Source	SS	df	<i>F</i>	<i>p</i> -value ^a
Intercept	.010	1	12.66	.001
		2		
<i>RM</i>	.000	1	.176	.675
<i>TYPE-EV</i>	.006	1	7.922	.006
<i>RM</i> x <i>TYPE-EV</i>	.004	1	4.676	.033
Error	.085	106		

RM = *RESPONSE MODE* = (probability vs. frequency)

TYPE-EV = *TYPE OF IRRELEVANT EVIDENCE* = (favorable vs. unfavorable)

^a all *p*-values are two-tailed

Panel B: Estimated marginal means by *RM* and *TYPE-EV* (standard error)

		<i>TYPE-EV</i>		
		Favorable	Unfavorable	Row Mean
		<i>n</i> =54	<i>n</i> =56	<i>n</i> =110
<i>RM</i>	Frequency	-.005	.022	.009
	<i>n</i> = 61	(.005)	(.005)	(.004)
	Probability	.009	.013	.011
	<i>n</i> = 49	(.006)	(.005)	(.004)
Column Mean		.002	.017	.010
<i>n</i> = 110		(.004)	(.004)	(.003)

TABLE 8
Experiment 2: Sensitivity Analysis for Tests of H1 on Reduced Sample: Signed Revision (*F-REV* s a Dependent Variable)

Panel A: Analysis of variance (<i>n</i> =72)				
Source	SS	df	<i>F</i>	<i>p</i> -value ^a
Intercept	.004	1	4.935	.030
<i>RM</i>	.000	1	.138	.712
<i>TYPE-EV</i>	.003	1	4.185	.045
<i>RM</i> x <i>TYPE-EV</i>	.003	1	4.103	.047
Error	.054	68		

RM = *RESPONSE MODE* = (probability vs. frequency)

TYPE-EV = *TYPE OF IRRELEVANT EVIDENCE* = (favorable vs. unfavorable)

^a all *p*-values are two-tailed

Panel B: Estimated marginal means by *RM* and *TYPE-EV* (standard error)

		<i>TYPE-EV</i>		
		Favorable <i>n</i> =40	Unfavorable <i>n</i> =32	Row Mean <i>n</i> =72
<i>RM</i>	Frequency <i>n</i> = 39	-.007 (.006)	.020 (.007)	.006 (.005)
	Probability <i>n</i> = 33	.009 (.007)	.009 (.007)	.009 (.005)
	Column Mean <i>n</i> = 72	.001 (.005)	.014 (.005)	.008 (.003)

The Norwegian School of Economics (NHH) is one of the leading business schools in Scandinavia, and is students' number one choice for a business education in Norway. The School's **Department of Accounting, Auditing and Law** is an ambitious and thriving environment for research and learning. Our faculty has three main research and teaching areas: Financial Accounting and Auditing, Management Accounting and Control, and Economics, Ethics and Law. For details, see the School's website: www.nhh.no.