Experts' Collective Judgments and Learning in Analytical Review

Minwoo Lee
Western Kentucky University

Stacy Bibelhauser
Western Kentucky University

Yining Chen
Western Kentucky University

Accurate estimation of financial information has become one of the most critical issues in today's auditing environment. This study extends the literature by investigating how auditors' collective probability and state judgments in the estimation process are affected by the saliency of the prior probability of relevant events. Consistent with current audit practice, this study performed investigations based on collective judgments by expert groups. The results indicate that expert audit teams make more accurate judgments regarding account default and learn more significantly from feedback in the more salient prior probability condition. We also found that while audit teams are somewhat biased in probability judgments, their state judgments are highly accurate. Overall, they make judgments more normatively than individual experts.

Keywords: analytical review, Bayesian probability, probability and state judgments, learning

INTRODUCTION AND LITERATURE REVIEW

Motivation and Purpose

With the legislation of the Sarbanes-Oxley Act (SOX) and the increased financial reporting regulations by the Public Company Accounting Oversight Board (PCAOB), high-quality financial reporting to protect investors has become one of the most critical issues in the financial world. SOX and PCAOB have also obliged auditors to face greater accountability in identifying accounts susceptible to misstatement in a company's financial statements. As PCAOB (2018) requires, auditors play an important role in the financial reporting process by assuring accounting estimates presented in financial statements. Certain provisions under SOX stipulate increased accountability of auditors for material reporting errors.

Despite the heavy regulation on financial reporting and greater auditor accountability, significant reporting errors have been documented in the post-SOX and PCAOB era. Stuber and Hogan (2021) provided evidence that PCAOB inspections did not improve the accuracy of accounting estimates. Boyle, Lewis-Western, and Seidel (2021) found that "difference in error" between quarterly and annual financial statements (which is expected to remain the same) has become greater compared to that of the pre-SOX era. Conducting audits on the bases of Generally Accepted Accounting Principles (GAAP) and Generally Accepted Auditing Standards (GAAS) provides some assurance to the users of financial statements.

However, since accounting and auditing involve estimating various events, material misstatements often result from inaccurate estimations (Christensen, Glover, & Wood, 2012). The documented results regarding reporting errors and the increased accountability of auditors following SOX and PCAOB warrant research investigating the accuracy of estimations made by auditors in the financial reporting process.

Unlike other audit areas such as internal control testing, estimating events in audits requires correct auditor judgments, not necessarily increased audit efforts (Griffith, Hammersley, Kadous, & Young, 2015). In estimating financial amounts, auditors rely on information that is not always precise in revealing the underlying events. Therefore, in an audit, it is crucial to identify and correctly assess potential risks associated with judgments made based on imperfect information. The auditor's ability to correctly utilize information during an audit is critical for users to make decisions, when relying on audited financial statements.

In making the judgments, auditors use analytical review, which is an important tool to estimate events and assess the likelihood of account misstatements. The American Institute of Certified Public Accountants (AICPA) addressed the importance and usefulness of analytical review in its pronouncements and required its use when planning an audit and determining the extent of audit testing (AICPA, 1988; 2002). Prior academic studies also confirmed the usefulness and popularity of analytical review. They reported evidence that analytical review is an effective and efficient procedure in detecting misrepresentation in accounts (Pinho, 2014) and is widely used among auditors in practice (Lin & Fraser, 2003). In examining extant audit research, Appelbaum, Kogan, and Vasarhelyi (2018) discussed a variety of uses and different aspects of analytical review used in public audits.

This paper performs an experimental investigation on auditors' probability and state judgments made in estimating accounting events as part of analytical review. Two things are worth noting. First, since the information that auditors receive is not a perfect signal of future events, their judgments may be biased. The saliency level of information is known as one of the major factors affecting judgment biases (Moser, 1989; Smith, Taylor, & Prawitt, 2016; Matin, 2019). Second, audits are mostly performed in a group. In addition to the collective endeavor, the current audit practice heavily relies on various specialists (Jenkins, Negangard, & Oler, 2018; Hux, 2017; PCAOB, 2015). As audits use a combined effort of a group of experts (i.e., auditors and non-auditor specialists), desirably audit judgment studies should focus on *group* decision-making by experts rather than novices such as students or individual experts.

The collective ability of an expert audit team to make correct decisions with the appropriate use of relevant information is crucial in the audit process. In view of that, this paper intends to address collective judgments by expert groups. Specifically, the purpose of this study is to investigate how differing information conditions, determined by the saliency of relevant information, affect experts' collective probability and state of nature judgments in the estimation process.

Judgments and Biases in Analytical Review

Many audit situations require auditors to determine the probabilities of relevant events (e.g., customer account default). When evaluating the likelihoods, the auditor receives signals generated by the analytical review for which the auditor knows the accuracy rate. To correctly judge the probability of an event in question, the auditors should also consider the other source of information, the normal chance (i.e., prior probability) of the event. The accurate probability assessment is the *Bayesian* probability, which incorporates the normal chance in determining the likelihood of the event.

Unfortunately, auditors often rely heavily on the signal and its accuracy rate rather than properly incorporating the normal chance (i.e., prior probability) of the event (Lee, Ross, & Little Jr., 2012; Lee, Little Jr., & Hunt, 2017). The reported evidence from this research is in line with the PCAOB concern about auditors' tendency toward ignoring relevant information appropriately in the judgment process (Fay & Montague, 2015; PCAOB, 2018).

Auditors make judgments of the probability of events in a situation similar to the "cab problem" setting, which was introduced by a judgment bias study of Treversky & Kahneman (1974). In this type of situation, individuals often make judgment errors due to base rate fallacy (BRF). BRF is the tendency of decision-makers to ignore or underweight the base rate (i.e., normal chance) of an event and focus heavily on new information in determining the likelihood of the event. Therefore, to the extent that auditors fail to incorporate the normal chance, their probability assessments will generate errors. The next section describes the "cab problem" setting that this paper explores in an audit context and Table 1 shows both accurate and biased probability assessments in this setting.

Prior studies show that individuals are subject to bias in making judgments. In the non-accounting literature, Martins (2006) claimed that heuristics used by individuals create biases in human probabilistic reasoning. In the accounting and auditing literature, several studies showed that auditors are biased or display partially irrational behavior (Shanteau, 1989; Heiman, 1990; Heiman-Hoffman & Moser, 1995) when making judgments. In evaluating information for their decisions, individuals are found to be biased. They are shown to be affected more by the information which is *perceived* more valuable (Moser, 1989) and by the information they choose to obtain rather than the information which is readily available (Smith, Taylor, & Prawitt, 2016). Normatively, the perceived value of information or how information is obtained should not affect the judgments.

An incorrect probability assessment may cause auditing to be either inefficient or risky (Kinney, 1987; Gimbar & Mercer, 2021). Consequently, the manner that auditors use information when making probability judgments is of interest and importance. Because the quality of judgments by decision-makers is not ideal, as Bonner (1999) asserted, research seeking to understand the judgment behavior of auditors benefits not only researchers but all stakeholders in the financial market. Also, better understanding of the judgment behavior will enable auditors to improve their decision-making skills in auditing.

Information Conditions and Judgments: From Novices to Individual Experts

Limited research is available on the effect of the informational environment on judgment bias of decision-makers. In an investment decision-making setting, Moser (1989) demonstrated that the perceived value of given information affects how it is incorporated when creating judgments. He found significant differences in predictions of firm performance with differing information conditions. In the auditing context, Ng and Tan (2007) found that enhancing the saliency of a qualitative materiality factor makes auditors increasingly incorporate that information in financial statements.

These results suggest that, in the cab problem setting, decision-makers perceiving more salient prior probability of an event as more valuable will incorporate the prior probability more heavily in estimating how likely the event will occur. Consequently, the decision-makers should generate less judgment bias. However, Lee et al. (2012) reported conflicting evidence. They found that, in an analytical review setting, the probability judgment errors do not significantly differ between two conditions having different perceived values of the prior probability information. Noteworthy is that their study asked students, who typically do not have required expertise, to perform an unfamiliar task. The use of non-expert subjects in the Lee et al. study may account for the inconsistent results.[1]lastly

Considerable research investigates the issues of using novice and expert subjects. Bonner (1994) showed that audit performance is affected by not only audit skill but task complexity. These factors were shown to exert a positive and negative influence, respectively, on audit performance. Similarly, Mohd-Sanusi and Mohd-Iskander (2007) found that the effect of audit effort on performance is weaker if the audit task is complex. As to experience, which may determine skill, Shanteau (1989) and Smith and Kida (1991) confirmed the positive effect of experience on judgment performances. The reason for more precise judgments was proposed by Hoffman, Joe, and Moser (2003). They demonstrated that compared to inexperienced auditors, experienced auditors make judgments differently to better utilize information (e.g., the prior probability of an event in question) that leads them to attend to evidence. The Association of Chartered Certified Accountants (2009) described analytical review decisions as complex tasks that demand expertise and professional judgments.

Given these findings and the nature of analytical review, research on the effect of differing information conditions (i.e., prior probability perceived as more valuable versus less valuable) in analytical review is considered a joint test of multiple effects. It simultaneously tests the effect of skill or experience as mentioned above and that of information conditions. To differentiate the effect of the latter, researchers should control for the former. To control for the effect skill and experience, Lee et al. (2017) used subjects with professional experience or auditing knowledge in their experiment analyzing the effects of different information conditions on audit judgments in analytical review. Unlike novice subjects, the expert subjects in their experimental study incorporated the information in a more normative manner. As a result, even though the expert subjects in their study generated the error, the magnitude was significantly smaller when the prior probability of the event in question was perceived as more valuable (i.e., more salient). This result collaborates the argument and findings in the previous research that the information perceived as more valuable is incorporated to a greater extent in the judgment process (Moser, 1987; Clor-Procell & Warfield, 2014; Ng & Tan, 2007).

Unexplored Area: Experts' Collective Judgments in Analytical Review

Numerous studies found that individual decision-makers are subject to judgment biases due to their failure to reflect value-relevant information in their judgment process. However, the general finding in the audit judgment literature suggests that groups perform better than individuals (see a survey studies in this literature by Trotman, Bauer, & Humphreys, 2015). Camerer (1987), in the non-audit context, confirmed the cancellation hypothesis suggesting that since individual biases are random, they cancel against others, and the resulting aggregated outcomes could be rational.

Auditing, in general, is a collaborative effort by a group. Although several auditing studies dealt with group decisions, they primarily focused on judgments in the audit review process rather than judgments that our study intends to investigate (i.e., probability and state judgments in analytical review). Moreover, most literature on judgment and decision-making in auditing pertained to individual judgments despite the fact that performing an audit requires collective decisions. Auditors' probability and state judgments as a group in analytical review remain unexplored in previous literature. Therefore, it is relevant and valuable to explore such judgments and find whether the performances of groups are superior to those of individuals. With investigations on group performances and further analyses, this study extends the Lee et al. (2017) study that used individual experts. The paucity of research on experts' collective judgments underscores the importance of our research which incorporates the collaborative nature of audits and the impact of different information conditions on probability and state judgments in analytical review.

Specific Research Issues

As stated earlier, this study aims to investigate how differing information conditions affect experts' collective probability and state of nature judgments in the analytical review process of an audit. We examine experts' collective judgments on how likely customers' accounts will default (i.e., probability judgments) and whether the customers will eventually default on the accounts or pay (i.e., state judgments). As the experts make judgments in the "cab problem" setting, their judgments are investigated in two information conditions that are different in saliency as to the normal chance of default. Specifically, the following three issues are addressed in this study. Refer to Figure 1 that depicts those issues.

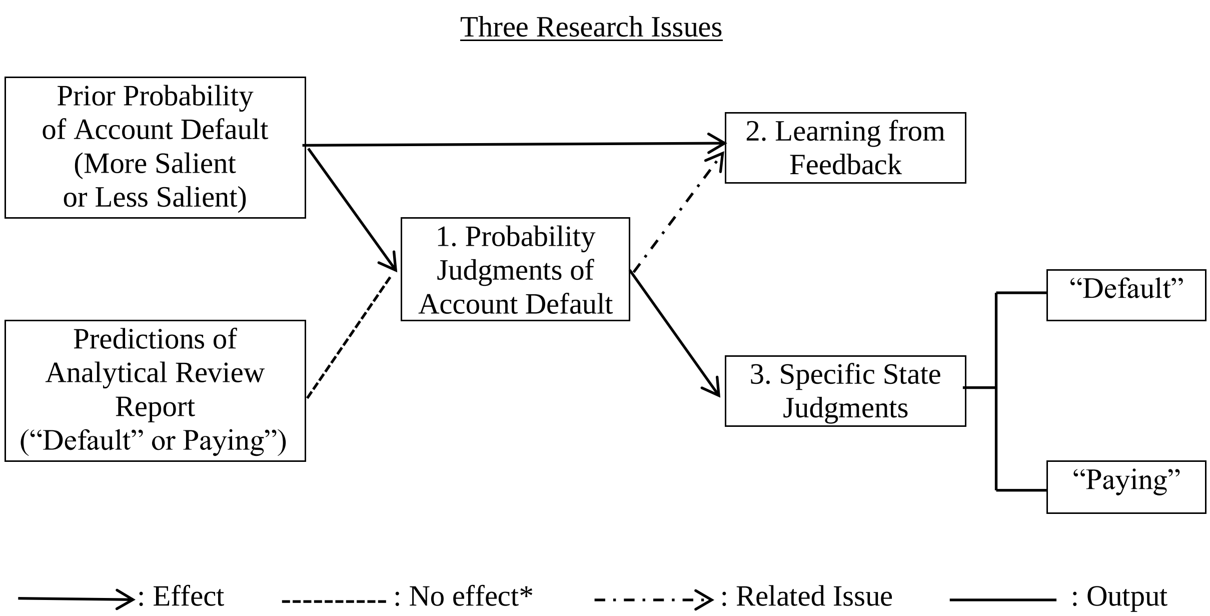
First, we investigate the effect of the saliency level of the prior probability of an event (i.e., account default) on the probability assessment of that event. We use expert groups as our subjects. As previous literature found that the prior probability information is under-weighted to create judgment errors, this study examines how prior probability saliency affects experts' collective probability judgment error and whether the effect differs from that reported for experts making judgments on an individual basis.

Next, we delve into the learning effect of feedback on the experts' judgment and decision-making. Bonner and Walker (1994) purported discounting evidence of judgment bias when a study denies the participants an opportunity to learn from previous experience. In an investment setting, Ganguly, Kagel, and Moser (1994, 2000) provided experimental evidence that participants' probability judgments become more accurate as they learn from previous experience. In our study, the subjects make the same probability assessments over multiple periods with feedback opportunities provided. In an audit planning situation not requiring probability judgments, Earley (2003) found that the lack of feedback information keeps auditors' reasoning low resulting in poor audit performance. In a performance evaluation context, Krumwiede, Swain, Thornock, and Eggett (2013) documented evidence of the feedback effect on long-term learning that evaluators increasingly rely on relevant information over multiple periods as they obtain outcome feedback. However, Leung and Trotman (2005, 2008) found that outcome feedback information does not function well to facilitate learning in complex audit situations.

Given the mixed results on the effectiveness of feedback in learning in the *general* auditing context and the paucity of research investigating expert group learning from feedback in the *specific* context of probability assessment, our investigation is of interest and importance. Our study explores differential long-term effects of learning from feedback on the expert subjects' collective probability judgments in different saliency conditions of prior probability.

Lastly, this paper investigates collective decisions involving specific state of nature judgments (i.e., whether a customer will pay the account or default on it). The state judgments are made based on probability judgments of the associated events. As Lee et al. (2017) assert, since the financial statements reflect state judgments rather than stating probabilities of relevant events, state judgments are final actions and therefore more relevant from the users' perspective than probability judgments.[2] As demonstrated in past literature, decision makers' probability judgments are biased. However, actions taken based on probability judgments can become less biased (Eger & Dickhaut, 1982). There has been little research in accounting/auditing literature dealing with how auditors' probability judgment bias affects the accuracy of their state judgments. Some studies in the law literature (e.g., Hunt & Mostyn, 2020; Meier, 2014) investigated how the probability judgments of events affect court sentencings (i.e., state judgments). Drawing on this literature, our study examines the accuracy rate of state judgments by experts making collective judgments.

FIGURE 1
RESEARCH ISSUES IN THE STUDY



[*] Analytical review report is also used in making probability judgments, but different predictions are not expected to differentially affect the accuracy (or error) of judgments.

Regulators and users of financial information increasingly emphasize the significance of audit practitioners' professional judgments in performing audits. Given the limited existing research on the issues needing consideration, this study adds to behavioral accounting research with novel evidence, employing expert subjects making group decisions on how differing information conditions affect their judgments. Potentially, our study is of value in the absence of research exploring group probability and state judgments by experts in analytical review.

To the auditing profession, understanding the causes of judgment errors and the opportunities to learn from feedback are essential in performing quality audits in the complex process of financial estimation. The results of our study can provide implications for the practical training of audit practitioners. The results may suggest an effective means of priming the auditing profession with the way they should think to improve their judgments.

The remainder of our paper is comprised of the following sections: The next section introduces the experimental methodology employed in the study. The following two sections report the results of probability assessments and state predictions, respectively. Finally, the last section summarizes the results and discusses the limitations and future research directions.

EXPERIMENTAL METHODOLOGY

Judgment Situation

Our experiment involved auditors assessing the probability that a client's customer account will default. The participants received a specific judgment scenario whereby they evaluated the account's collectability. The judgment situation is as follows:

> As an audit group, you and your coworkers are auditing Anderson Company, a client company that has not received payment from one of the client's (i.e., Anderson's) customers.[3] The customer's account carries a large balance and is over-due for a long time. It is possible that the customer is just late in making payment, in which case Anderson will receive money from the account. However, it may be true also that the account will default. Normally, in the client's industry, customer accounts that are as over-due as and as large as this specific customer's account have an *x*% probability that they will default. As part of analytical review, you have received a report from an independent credit analyst. As to the eventual status of this customer account, the report makes one of two possible predictions, "Default" or "Paying." The report is accurate 60% of the time in making its predictions. What is your group's assessment of the probability that this customer's account will default?

This study examines the impact of saliency of the prior probability (i.e., normal chance of default) on the subject's probability and state judgment performances. For that purpose, we manipulated the information condition with two levels of the prior probability of default. Depending on the condition, the value of *x* in the above scenario is either 25% or 38%, representing different degrees of saliency. Probability saliency refers to the extent to which the probability is specific in estimating a future event. Note that a 50% probability is the least specific or definite (i.e., most obscure) one in predicting whether an event in question will occur or not. If a probability is more distant from 50%, it is more definite. Therefore, the 25% probability is considered more salient (i.e., definite) than the 38% one.

FIGURE 2
EXPERIMENTAL CONDITIONS

Condition 1	Condition 2
<u>Normal Chance of Account Default</u> Prior Probability - More Salient (25%) <u>Independent Analytical Review Report</u> Accuracy (60%)	<u>Normal Chance of Account Default</u> Prior Probability - Less Salient (38%) <u>Independent Analytical Review Report</u> Accuracy (60%)

We randomly assigned the subject groups to one of the two saliency conditions manipulated by the experiment. Figure 2 shows these experimental conditions. Table 1 presents both Bayesian (accurate) and total BRF probability and state assessments for different information conditions.

When creating the above experimental scenario, two issues were taken into consideration. First, given that account misstatements are a primary concern in analytical review, understanding the decision-making situations is essential to probability assessments for account misstatements. However, prior studies on probability judgments, such as Lee et al. (2017) and Lee et al. (2012), presented participants with vague situations and asked the subjects to assess the probability of misstatement but failed to provide reasons that accounts may be misstated. Unclear situations might influence the probability assessment since task familiarity can affect judgment bias (Hoffman et al., 2003; Mohd-Sanusi & Mohd-Iskander, 2007). To avoid this possible confounding effect, we presented the subjects with a scenario that distinctly described the judgment situation by providing the reasons for potential account misstatements. It should be noted that when the probability of default on receivables is substantially high, these accounts will be overstated if not adjusted for the likelihood of default. Therefore, assessing the probability of default is equivalent to judging the probability of account misstatements.

TABLE 1
CORRECT (BAYESIAN) AND TOTAL BRF JUDGMENTS OF ACCOUNT DEFAULT [a]

Information Condition	Analytical Review Report Accuracy	Report Prediction	Correct Bayesian Probability [a,d]	Total BRF Probability [a,b]	Diff [c]
More Salient Prior Probability (25%)	60%	Default	33% (Paying) [d]	60% (Default) [d]	27% points
		Paying	18% (Paying) [d]	40% (Paying) [d]	22% points
Less Salient Prior Probability (38%)	60%	Default	48% (Paying) [d]	60% (Default) [d]	12% points
		Paying	29% (Paying) [d]	40% (Paying) [d]	11% points

Note: [a] All probabilities in the table are the assessed likelihood that the account will default
[b] Probability assessment that solely depends on the analytical review report.
[c] Absolute deviation of Total BRF probability assessment from Bayesian one in percentage points
[d] State judgments are made based on assessed probabilities. See the state judgment results later in the paper. If the assessed probability of account default is 50% or higher, the subjects will predict the actual state of the account to be "Default." Otherwise, the predicted state will be "Paying (no default)."

Secondly, the scenario should be realistic. We kept the normal chance (i.e., prior probability) of account default less than 50% to keep the subjects from facing an unfamiliar situation. In case the prior probability of account default is set too high to be realistic, that may force the subjects to make probability judgments in an unusual situation and possibly cause them to generate a greater amount of bias in their judgments.

Subjects and Groups

A number of prior studies used college students as subjects for assessing various auditing judgments. For example, in the analytical review setting, Lee et al. (2012) used student subjects for the task of probability assessments. A critical disadvantage of using students is that the subjects have to perform a task that is not familiar to them. Since this effect was not controlled in those studies, it is difficult to determine whether their results were driven by the use of novice subjects. The Lee et al. (2017) study, which employed expert subjects, showed that experienced subjects make better judgments in analytical reviews than non-expert subjects. Our study also uses experts because audit judgments require professional expertise and experience. Because the participants understand the process and purpose of analytical review, they qualify as expert subjects performing familiar tasks.

Although the use of expert subjects is a significant improvement from prior studies, there remains another issue. Previous studies on auditors' judgments in analytical review focused on individual performance (Lee et al., 2017; Lee et al., 2012). Because audits require a collective work of a group of practitioners, ideally audit judgment studies should focus on group decision-making. Our study examines the collective probability judgments of groups using experts.

We recruited a total of 54 experts to participate in the experiment as subjects. They consisted of certified public accountants (CPAs), other practitioners, and academicians with relevant experience in their respective fields. Table 2 presents the makeup of the subjects. We created 18 groups having three members each.[4]

In forming the groups, assigned participants of *each occupation subcategory* of practitioners and academicians to the groups as evenly as possible. Given that there were 18 CPAs in the sample, we randomly assigned one CPA to each of the 18 groups. We also assigned non-CPA practitioners and academicians by random drawings in such a way that no group had more than 1 participant from any *occupation subcategory*. As a result, most groups had a desirable mix: a CPA, a non-CPA practitioner, and an academician. We randomly assigned the 18 groups to one of the prior probability saliency conditions (25% and 38%), with nine groups to each condition.

TABLE 2
EXPERIMENTAL SAMPLE

Field Category	Occupation Subcategory	Number of Participants
Practitioner	Certified public accountants	18
	Accounting staff	4
	Financial analysts	9
	Subtotal	31
Academician	Accounting professors	8
	Business professors	7
	Accounting graduate students	8
	Subtotal	23
Total		54

Our sample adequately reflected recent findings in the specialist literature regarding the composition of audit teams. In today's audits, the vast majority of audit teams include specialists such as valuation, tax, forensic experts, etc., as well as auditors (PCAOB, 2015). Prior studies generally confirmed the benefits of relying on specialists in making judgments and measuring fair values (Canon & Bedard, 2017; Jenkins et al., 2018). Hux (2017) synthesized the specialist literature by integrating the research on the use of experts in various aspects. For our experimental study, we created groups consisting of auditors and other experts. The group composition is consistent with the currently popular practice of performing collective audits with auditors and other specialists.

Experimental Procedures

In performing our experiment, we generally followed the procedures and methods employed by Lee et al. (2017). We provided the groups with several pieces of information at the beginning of the experiment. First, the experiment instructions disclosed the prior probability (25% or 38% depending on the condition) of account default.

We informed the participants that the experiment would consist of 20 periods and that they would analyze one customer account of the client (i.e., Anderson) per period. The instructions indicated that, in each period, each group receives, as part of the analytical review, a report from an independent credit analyst (report, hereafter) predicting a state of either "Default" or "Paying," depending on the analyst's beliefs about the customer's actions. The instructions also revealed that the report correctly predicts the states 60% of the time. Based on this information, the groups were required to judge the likelihood of the customer account default.

In their investment-related experimental study, Ganguly et al. (1994) suggested that the participants may believe that the accuracy rate of specific case evidence (e.g., report predictions in our study) is higher when tested with a larger sample. Therefore, subjects may be apt to believe the accuracy rate of the report varies depending on the sample composition used in determining the accuracy rate.[5] To prevent this assumption, we told the subjects that the accuracy rate of the report was determined with a large sample of companies which is equally divided between "Default" and "Paying" predictions.

The experiment materials provided further information about the makeup of the customer accounts of the client (i.e., Anderson). We informed the subjects that over the 20 periods, they would assess 20 randomly selected customer accounts: half of which *the report predicted* "Default" and half of which *the report predicted* "Paying". These customer accounts were randomly assigned to the experimental periods, one per period.

In each of the 20 rounds, the subject groups assessed the probability of the account default in a context-specific setting known to be subject to BRF. In each period, the following procedures occurred: First, subject groups received the prediction of the report regarding a customer account ("Default" or "Paying"). Second, given the prediction of the report, the groups made a collective judgment as to the probability of default for the account presented in the period. Lastly, the groups observed the *actual* state, either "Default" or "Paying."

By revealing the actual state after the probability judgment concluded each period, the experiment provided the groups with a learning opportunity. With the actual state information, the subjects could adjust the assessed probabilities in subsequent periods. The multi-period format of this experiment enabled us to examine the long-term effect of learning from feedback.

At the conclusion of the experiment, we revealed the correct answer for each period, which is the Bayesian probability. As Table 1 shows, the Bayesian probability varies, depending on the prior probability and the analytical review report prediction. The table also presents total BRF probabilities, which completely ignore the prior probability of account default and rely solely on the prediction of the report. Also, we paid the groups based on their performances.

Other Arrangements

Prior to running the experiment, we created the predictions of the analytical review report and actual states to be used in the experiment. We modified the method used in Lee et al. (2017) and generated 20 pairs of report prediction and actual states (one per period) consistent with the prior probability and the accuracy of the report for each of the prior probability saliency conditions (25% and 38%) by using random drawings.

The monetary compensation typically awarded in this type of study is not sufficient to motivate expert subjects (Lee et al., 2017). Their study suggested that experts are keenly interested in knowing how they performed. In a relatively complex judgment situation, outcome feedback, as well as financial incentives, can be a good performance motivator (Buchhelt, Dalton, Downen, & Pippin, 2012). To incentivize our subjects to participate and thoughtfully make decisions, after the experiment, we revealed the groups' performance data and relative ranks to the participants.

We used a point system to measure group performance. Each group received 100 points per period. These points decreased when the group's probability assessment differed from the correct Bayesian probability. For each percentage difference in the absolute value between a group's probability assessment and the correct Bayesian probability, the group score decreased by one point. The overall performance measure of a group is the sum of all points over the 20 periods.

Noteworthy is that while most multi-period experimental studies paid the subjects based on their performances for only one randomly selected period, we determined the compensation to our subjects based on their overall performances over 20 periods. Our point system and compensation method performance measure made it possible for the subjects to stay serious in making judgments throughout the experiment. In addition to informing the subjects of their performance, we gave monetary prizes as a token of appreciation and added incentive. We awarded 5,000 additional points each to the three best groups in each experimental condition. The groups received cash at a rate of $1 per 100 points.

EXPERIMENTAL RESULTS: PROBABILITY JUDGMENTS

Group Measure

This study requires that subject groups make judgments regarding potential account defaults. Since group judgment and decision-making processes take many different forms, how to determine a group's opinion is an issue. Several studies in the non-accounting literature are relevant to the judgments made in our experiment. An experimental study by Ambrus, Greiner, and Pathak (2015) found that bargaining and persuasion are used in creating collective opinions, and the median view plays an influential role in that process. Nish and Masuda (2013) proposed a mathematical model for forming group opinions among individuals in a Bayesian fashion. Also, Miller (2008) presented a generalized model of aggregating different reasonings of individuals in subjective decision situations. Some surveys of collective judgment theory furnish insights to our study as to probability aggregation (List, 2012) and the effect of individual expertise on forming collective opinions (Martini & Sprenger, 2015).

In the accounting/auditing context, group brainstorming, aggregating the opinions of individual members and stimulating new ideas, has emerged as a primary method of forming collective opinions (Trotman et al., 2015). Thus, the average of the individual judgments in the group cannot be a good proxy for the group's collective opinion. Furthermore, the studies in the specialist literature suggest that audit teams heavily count on specialists on their audits (Jenkins et al., 2018; PCAOB, 2015). In this auditing environment, there is no predetermined manner in which collective opinions are formed. Also, as Dennis and Johnstone (2018) found, the outcomes of an audit team can be jointly determined by leadership and subordinate knowledge. Because group decision-making philosophy, processes, and circumstances vary across audit firms, we allowed the groups in this study the same latitude.

We acknowledge the selection and relative performances of various group measures as important topics for future research, they are not of direct relevance to this study. However, it should be reminded that the focus of our paper is on the impact of varying information conditions on auditors' collective judgments, not the choice or impact of different group measures per se.

Probability Judgment Error

This study examines the accuracy of probability judgments on account default and compares them across different information conditions. Again, the Bayesian (i.e., correct) and total BRF probability assessments can be found in Table 1 for the four situations: 2 prior probability saliency levels x 2 predictions of the report. The correct probabilities vary depending on the information condition and analytical review report prediction. Therefore, we measured the accuracy of the judgment based on probability judgment error, which is the *absolute value*[6] of the deviation of the group's assessed probability from the correct probability. This judgment error represents how distant the assessed probability is from the correct Bayesian probability.

Table 3 reports the groups' overall mean judgment error. The overall mean error indicates that the subject groups' average probability assessment was different from the correct probability by 11.031% points. The result suggests the expert groups' probability assessments are somewhat biased. This result is consistent with findings by a number of prior studies (e.g., Christensen et al., 2012; Boyle et al., 2021; PCAOB, 2018; Stuber & Hogan, 2021) that auditors generate errors at least to a certain extent in performing audits.

Lee et al. (2017) conducted a similar study with individual experts rather than expert groups and reported a higher mean judgment error for individual auditors. While their study had the same information manipulations to examine probability assessment errors in analytical review, the judgment situations differ from the scenarios present in the current study. Note that the different judgment situations between the two studies prevent direct comparison of performances between groups and individuals. However, our result may corroborate the documented advantages of groups in performing audits (Trotman et al., 2015). Our study focuses on the impacts of different information conditions on the probability judgment errors of expert groups (rather than the difference in performances between groups and individuals) and aims to discover if the impacts are different between expert groups and individual experts.

TABLE 3
DIFFERENT INFORMATION CONDITIONS AND MEAN PROBABILITY JUDGMENT ERRORS BY SUBJECT GROUPS

Probability Judgment Error[a]				
	Analytical Review Report Prediction			
Prior Probability Saliency	Default		Paying	Overall
More Salient (25%)	9.242		9.569	9.406
Less Salient (38%)	11.433		13.878	12.656
Overall	10.338		11.724	11.031
Analysis of variance				
	Statistics			
Factor	Df	Mean Square	*F*-value	Significance
Prior Probability Saliency	1	950.625	13.228	<.001
Report Prediction	1	175.003	2.435	.120
Prior Prob. Saliency x Report Prediction	1	99.225	1.381	.241

Note: [a] Each error represents mean absolute deviations from Bayesian probability. All error amounts are expressed in terms of % points.

Effects of Different Information Conditions

Given that we found that expert groups are biased in making probability judgments, we investigated the effects of information conditions on the magnitude of the error in assessing the probability. Table 3 (top panel) also reports mean probability judgment errors for all possible combinations of two saliency levels of prior probability and two predictions of the analytical review report.

Consistent with the expectation from past literature (i.e., more accurate judgment with more salient information), on average, the groups in the more salient (25%) prior probability condition generated smaller errors than the groups in the less salient (38%) condition. The mean error of the 25% condition (for all analytical review predictions together) was 9.406% points, while that of the 38% condition was 12.656% points. The analysis of variance reported in the bottom panel of Table 3 confirms that the difference is statistically significant (p-value < .001). These results confirm that the expert subject groups incorporated the prior probability in a more Bayesian manner when it was more salient. Thus, their judgments were less biased when the prior probability saliency level was higher. The result is consistent with findings in the related literature that decision-makers incorporate more conspicuous information to a greater extent in their investment decisions (Moser, 1989; Clor-Procell et al., 2014) and in auditing financial statements (Ng & Tan, 2007 because information saliency positively affects the value or reliability of the information.

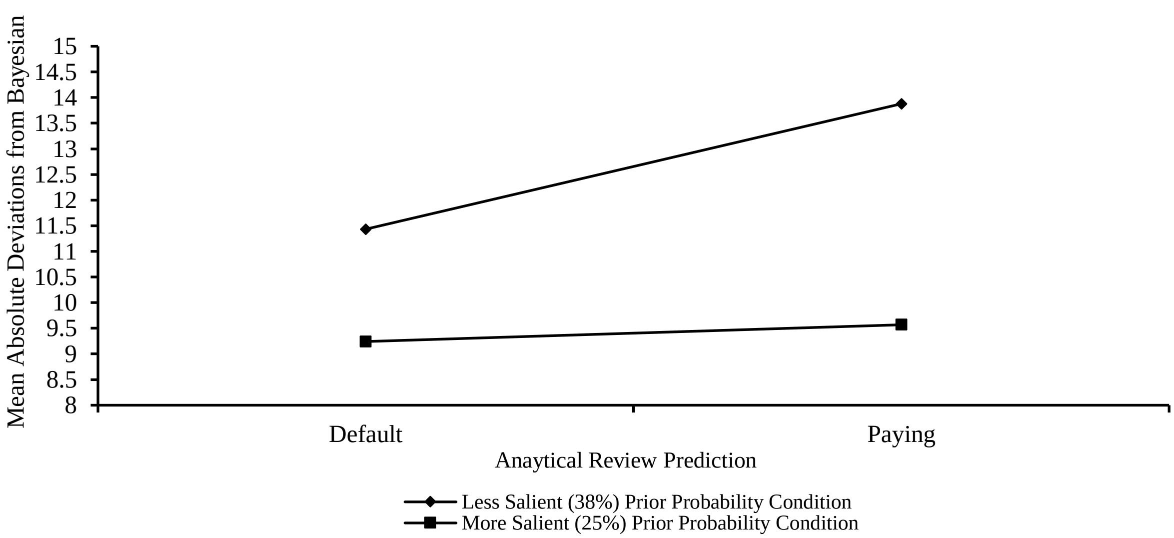
Lee et al. (2017) found a similar effect of prior probability saliency on probability judgment error with individual experts as well. This result, however, does not hold with non-expert subjects. Lee et al. (2012) reported that prior probability saliency does not affect the probability judgment error by student subjects. In the latter study, the subjects showed heavy reliance on case specific information (i.e., predictions from the analytical review) without considering the prior probability. Their result, that the accuracy rate of the analytical review predictions is the primary factor in their probability assessments, typifies BRF. The difference of results between student and expert subjects can be explained by the findings that experience enables subjects to make better judgments (Bonner, 1994; Mohd-Sansi & Mohd-Iskander, 2007; Shanteau, 1989; Smith & Kida,1991).

We also analyzed the effect of the predictions of the analytical review report on judgment errors. The groups had smaller mean errors (for all prior probability saliency conditions together) when the prediction was "Default" rather than "Paying" (10.338% points and 11.724% points, respectively). However, the bottom panel of Table 3 shows that the two errors are not significantly different (*p*-value = .120). Normatively, there should be no difference in probability judgment performance whether the prediction of the report is "Default" or "Paying." Our result is consistent with this expectation.

The same bottom panel 3 also shows that there was weak interaction of prior probability saliency with prediction of the report. In both "Default" and "Paying" report predictions, the magnitude of the probability judgment error increased as the saliency condition changed from 25% (higher saliency) to 38% (lower saliency). Between the two saliency conditions, the increase in probability judgment error is greater for the "Paying" prediction than for the "Default" prediction. Figure 3 provides a graphical representation of this interaction.

There is no a priori reason for expecting that the interaction effect exists. The increase in errors between the prior probability conditions should be identical regardless of the prediction. One could argue that the greater increase in error with the "Paying" prediction may be due to the fact that while the prediction was "Paying," the subjects were asked to provide the probability assessments of "Default." The "Paying" prediction could make probability assessments of default an unfamiliar task, which in turn could reduce the subjects' performances (Bonner, 1994). However, the interaction effect is not sufficiently significant (*p*-value = .241) to support this argument related to task familiarity.

FIGURE 3
WEAK INTERACTION EFFECT ON PROBABILITY JUDGMENT ERRORS



Furthermore, there is evidence that task familiarity does not affect auditor performance when the auditors have "high knowledge" and high accountability (Tan, Ng, & Mak, 2002). Given the fact that our subject groups included both auditor and specialists and the reported evidence that specialists improve audit quality, we consider our expert subject groups to have "high knowledge." Also, as discussed earlier, in light of the findings by Lennox and Li (2020) that asset overstatements have a relatively high litigation risk to auditors, the judgment situation in our experiment is associated with high accountability.

We compared our result of the interaction effect between prior probability saliency and analytical review prediction with that reported by Lee et al. (2017) with individual expert. They showed an interaction effect which is much more significant than ours. The interaction effect is due to the bias created in probability judgments. The task familiarity argument discussed above may hold with individual subjects in their study. Overall, the results of our study support the suggestion that groups perform more rationally than individuals. Our results are consistent with the findings in the specialist literature and also by Tan et al. (2002).

Learning From Feedback

The effect of learning on probability judgment error is the second research issue examined in this paper. Bonner and Walker (1994) suggest that evidence of biased probability judgment is less conclusive if participants do not receive a learning opportunity by obtaining feedback on their performance. Given that, we investigated how the magnitude of the error changes through experimental periods. In each of our 20 experimental periods, the actual state ("Default" or "Paying") was revealed as feedback after the subjects' judgment. To the extent that the subjects ignored the prior probability of account default, the occurrences of actual states became inconsistent with the probability assessed by the subjects. If the subjects noticed this inconsistency, they could revise their probability assessments in later periods. Given the reported evidence that decision-makers incorporate more salient information in their decisions to a greater extent (Moser, 1989; Clor-Procell et al., 2014; Ng & Tan, 2007), we expect that subjects learn more easily when they face more salient prior probability.

To test whether subject judgments become more accurate (i.e., generating less error) with learning, our study employs the following regression used by Ganguly et al., 2000):

$$GE_i = \alpha + \beta\, EPN_i + \varepsilon_i \qquad (1)$$

where: GE_i = error in group's probability assessment for group i
EPN_i = experimental period (numbered 1, 2, …,10) for group i
ε_i = residual error for group i

The experiment had ten rounds where the report predicted "Default" and another ten rounds where the report predicted "Paying." In order to examine these situations separately, we assigned EPN values for each set of predictions independently. An EPN value of 1 corresponds to the first occurrence of a "Default" prediction, and an EPN value of 10 corresponds to the last "Default" prediction. EPN values for the "Paying" predictions were coded in the same manner.

To analyze the effects of differing information conditions on learning, we ran the regression separately for each of the four combinations listed in Table 4. If learning occurs over time, the subject groups' probability judgments should become more accurate (i.e., the error should decrease) as the periods continue. Then, β, the estimated coefficient of EPN in Equation 1, should be negative.

The results of the regressions are reported in Table 4. For all combinations, the β coefficients were negative, suggesting that learning from feedback occurred to the subject groups. In the more salient (25% prior) condition, the regressions resulted in β coefficients of -0.876 (*p*-value = .004) and -0.928 (*p*-value = .012) for "Default" and "Paying" prediction cases, respectively. Consequently, we confirmed a significant learning from feedback occurring in the more salient condition. As to the less salient (38% prior) condition, the β coefficients are -0.487 (*p*-value = .091) and -0.475 (*p*-value = .088). Both coefficients have a level that is less significant than a conventional test level of 0.05.

The β coefficient with the more salient (25%) prior probability condition is more negative than that with the less salient (38%) prior probability condition for each of the "Default" and "Paying" cases. However, given that βs in the less salient prior probability condition are only weakly significant, a definite conclusion cannot be made as to whether the subjects reduced the error by a greater amount in the more salient prior probability condition.

In summary, our study demonstrates significant learning effects with more salient prior probability. This evidence contradicts the finding of Lee et al. (2017) who investigated individual experts' probability assessment performances. Their study reported that no significant learning occurred to the individual expert subjects in the same condition. The results of our study support the expectations of the literature on the information saliency that more salient information is more easily incorporated in judgments, and, therefore, learning can occur more easily. The cancellation hypothesis by Camerer (1987) may explain the discrepancy in results between the current study and Lee et al. (2017). The hypothesis suggests that group decisions contain less bias than individual decisions because the random errors made by individuals offset each other in the process of combining individual opinions. Overall, the results of our study suggest that expert groups make the judgments in a more rational manner than individuals do.

TABLE 4
REUSLTS OF REGRESSION (LEARNING FROM FEEDBACK)

Situation: Saliency of Prior Probability	Situation: Analytical Review Report Prediction	α (*t*-stat)	β (*t*-stat)	Signif. of β	Adj.R^2
More Salient (25%)	Default	14.052 (7.716)	-0.876 (-2.985)	.004	.082
	Paying	14.681 (6.534)	-0.928 (-2.562)	.012	.059
Less Salient (38%)	Default	14.111 (7.982)	-0.487 (-1.709)	.091	.021
	Paying	16.633 (9.406)	-0.475 (-1.724)	.088	.022

The significant learning effect of feedback found in the more salient information condition of our experiment is also generally consistent with evidence reported from the feedback literature in the non-probability assessment setting. Without outcome feedback information, the auditors' reasoning level is low (Earley, 2003) or auditors cannot obtain the knowledge required to perform audits successfully (Bonner & Walker, 1994). See a review of the feedback literature by Andiola, 2014 on the effect of feedback on learning and performance.

Some research in the feedback literature, not directly pertaining to probability judgments, may explain the insignificant learning effect of feedback found in the less salient prior probability condition. Understandably, if the prior probability saliency level is low, the task of probability judgments becomes more complex because the prior probability information is more obscure for the subject to use. At the same time, the task becomes more configural. A configural task is the one for which simultaneous consideration of multiple cues is required (Leung & Trotman, 2005).

In the less salient prior probability condition, decision-makers may attempt to search and use additional cues in place of the obscure prior probability information cue. Given that, the probability judgments in the more and less salient prior probability conditions are comparable to the examples of non-configural and configural task, respectively, in Leung and Trotman (2005). In addition, Bonner and Walker found that the effectiveness of outcome feedback is potentially diminished for relatively complex tasks. Also, Leung and Trotman (2005) showed that outcome feedback is more effective in performing non-configural tasks than configural ones. Extending their earlier study, Leung and Trotman (2008) found evidence that outcome feedback has no impact on information processing ability for configural tasks. These findings shed light on the difference in the significance of learning effect between the more and less salient prior probability conditions.

EXPERIMENTAL RESULTS: SPECIFIC STATE JUDGMENTS

Auditors' probability judgments influence their predictions about a specific state nature. Specific state judgments are the next course of action subsequent to probability judgments in providing accurate financial information.[7] Therefore, making correct state judgments is the final and critical action. In this experiment, participants completed a probability judgment of account default each period. The probability assessments determine which specific states of nature the group believes to occur, “Default” or “Paying.”

There has been little research in accounting/auditing literature dealing with how auditors’ probability judgment bias affects the accuracy of their state judgments. Some research in the law literature addressed a related issue. Hunt and Mostyn (2020) found that probabilistic reasoning is valid and useful in fact-finding. However, other studies suggested that the court actions or sentencings (i.e., final actions) are determined in a manner that is inconsistent with probabilities of the relevant events (Lindley, 1991; Meier, 2014). Drawing on the law literature, we have identified a relevant research issue for analytical review. Our findings suggest that expert groups generate biases when making probability judgments. Eger and Dickhaut (1982) purported that actions contain less bias than the probability judgments leading to the final actions. Given that, investigating group performances in making judgments regarding specific states is informative and interesting.

Precision in Specific State Judgments

In our experiment, if the group’s probability judgment for the “Default” state is 50% or higher, its state judgment is considered “Default.” In case the probability judgment is below 50%, the group’s predicted state is considered “Paying.”

Table 5 provides the proportion of precise state predictions made by the subjects. The overall accuracy rate is 0.847 (84.7%). The table also reports the accuracy rate in specific state predictions for each situation (subsample). Accuracy rates range from 0.756 to 0.939 (75.6% to 93.9%) depending on the situation, suggesting that the expert groups are highly accurate in predicting specific states. Our results that the subjects perform well in making state judgments, despite committing errors when making probability judgments, substantiate the argument by Eger and Dickhaut (1982).

TABLE 5
ACCURACY RATES OF GROUPS’ SPECIFIC STATE JUDGMENTS

Sample Partition	Mean Probability judgment Error[a]	State Judgments: Accuracy Rate	z-value[b]	Signif.
Entire sample	11.031	.847	13.168	< .001
Subsample: More Salient Prior Probability (25%) Only	9.406	.939	11.780	< .001
Subsample: Less Salient Prior Probability (38%) Only	12.656	.756	6.869	< .001
Subsample: Default Analytical Review Prediction Only	10.338	.800	8.050	< .001
Subsample: Paying Analytical Review Prediction Only	11.724	.894	10.572	< .001

Note: [a] From Table 3
[b] Based on one-sided binomial test regarding the sample’s mean accuracy rate being as large as the observed value, if the population’s accuracy rate is 0.5

For each situation, we performed a test to determine the significance level of the reported accuracy rate. If one makes the judgments by random guessing or exclusively using the analytical review report prediction and its accuracy rate (i.e., total BRF), the accuracy rate of state predictions would be 0.5 (50%).[8] The binomial test performed for each situation confirms that the reported accuracy rate is significantly greater than 0.5 (50%) with a p-value < .001. These results demonstrate that the subjects did not randomly guess or suffer from total BRF. Also, note that Bayesian probabilities are correct probabilities, which make their state predictions 100% accurate. Since the subjects’ state judgments are not perfectly accurate, their performances are also different from what Bayesian judgments suggest. Overall, our results indicate that the subjects made judgments in a manner different from random, total BRF, or Bayesian judgment process.

While there was a small difference in the accuracy rate between the “Default” and “Paying” prediction cases, a sizeable difference occurred between the more and less salient prior probability conditions. The mean accuracy rate of the state predictions for the higher prior probability saliency conditions is 0.939 (93.9%) and that for the lower prior saliency condition is 0.756 (75.6%). Table 3 shows that the subjects generated much smaller biases in making the probability judgments in the higher saliency condition (on the other hand, the magnitudes of the judgment error were not significantly different between the “Default” and “Paying” prediction cases). The higher accuracy rate in state judgments in this condition can be explained by the significantly smaller probability judgment error in the same condition. This result supports the argument that the higher saliency of prior probability facilitates more accurate judgments.

CONCLUSION

Summary of Findings

We performed an experiment to investigate how auditors’ group judgments are affected by different information conditions in an analytical review setting. Employing expert subjects, this study examined auditors’ *collective* probability and state judgments regarding an account default in varying information conditions. Our results show that the subjects generated errors in making the probability judgments. Still, the subjects are shown to create a smaller magnitude of error in probability assessments when the prior probability of default is more salient. However, different predictions from the analytical review report were not shown to differentially affect the magnitude of the error committed by the subjects.

We also investigated the learning effect of feedback in a multi-period probability judgment setting. The experiment provided the subjects with opportunities to learn from feedback in order to reduce the judgment errors as they repeat making the same judgment. The subjects’ learning was significant only when making the judgments in the situation having more salient prior probability.

Finally, while the subjects were found to generate errors in making probability assessments, they showed highly accurate performances in state predictions. Our results also confirm that the accuracy rate in state predictions was significantly higher when the subjects had the more salient prior probability. Overall, the reported results suggest that the expert group subjects make the state judgments in a manner which does not appear to follow either random, Bayesian or total BRF judgment processes.

There are discrepancies in results between our study and the Lee et al. (2017) study which investigated individual experts’ judgments, rather than expert’s collective judgments. Their study reported a somewhat significant interaction, which is irrational, between the prior probability saliency and analytical review report prediction in affecting the magnitude of bias in probability judgments. Also, there was no significant learning effect of feedback to individual experts in the more salient prior probability condition, where such learning is expected to occur. Thus, we found that expert groups make judgments in a more rational and normative manner than individual experts in performing analytical review. Our results are generally consistent with reported evidence in related research. The information saliency literature (e.g., Moser, 1989; Ng & Tan, 2007; Clor-Procell et al., 2014) asserts that more salient information is incorporated in the decision-makers’ judgment process to a greater extent. Research in group decision-making and judgments (e.g., Trotman et al., 2015) generally confirmed the superiority of group judgments to individual ones. Also, a number of studies in the specialist literature (e.g., Canon & Bedard, 2017; Jenkins et al., 2018) reported evidence of improved judgments of audit groups including non-auditor specialists. Such groups were used as subjects in our experiment.

Unresolved Issues and Future Research Directions

When presenting financial information to the stakeholders, one cannot overemphasize the importance of auditors’ professional judgments regarding specific states, such as account misstatement. As the audit judgment process is complex, there are several issues remaining unaddressed in our study. Accordingly, we note the limitations of our study and propose future research areas for enhancing our understanding of the behavioral aspects of auditors' judgments in analytical review.

Experimental Manipulation

With the accuracy rate of the analytical review report fixed at 60%, this experiment created four combinations of informational situations formed by the two levels of prior probability saliency (manipulated by the experiment) and two different predictions from the analytical review report. Consequently, in each of the combinations, the Bayesian probability of account default was below 50% and the resulting correct state judgment was "Paying." This uniformity may have created unintended complications in the subjects' judgments. This issue was previously noted by Lee et al. (2017), but it has not been resolved.

After a number of trial and error, we concluded that changing the prior probabilities could not resolve the issue. Although it would be a challenge especially for expert group studies, a larger sample size with the recruitment of more subjects can ease this limitation. Note that, with the numbers of participants and the groups fixed, adding one manipulation will decrease the number of subjects in a group by half. Since our research investigates group performances, the reduction of the group size below to a certain number is not desirable. [9]

Because of the limited number of participants in the experiment, we had to perform the experiment with one manipulation by holding the accuracy rate of the analytical review report constant (60%) while manipulating the prior probability saliency. If significantly more subjects participate in a future experiment, one additional manipulation on the accuracy level of the report (e.g., high and low accuracy rates) can be added. The resulting 2 x 2 manipulations on prior probability saliency and the analytical review report accuracy could resolve the issue with the Bayesian probabilities.

The Bayesian probabilities, computed consistently below 50% across all informational situations, were also a concern in interpreting the results. They limited our ability to analyze the robustness of the results as to the accuracy of the subjects' state judgments. The ideal mixture would be that the Bayesian probabilities are above 50% for two combinations and below 50% for the remaining two combinations, and the correct and total BRF state predictions differ from each other in each combination. In our experiment, only two of the four combinations satisfy this desirable scenario. In this ideal state, the effect of probability judgments on state judgments could be better analyzed. If future research addresses this concern appropriately with more manipulations with a larger sample, the judgment situation will become better for the subjects, which in turn, will facilitate the interpretation of the experimental results.

Threshold Probability Level

This study used 50% as the threshold probability for judging whether the customer account will default. An adjustment to this threshold level could be made based on litigation risk that auditors face. Given the conservative nature of audits, auditors could use a lower threshold probability to judge that an account will default for the purpose of reducing audit risk. Lennox and Li (2020) found that the likelihood of auditor being sued varies across audit contexts, with alleged asset overstatement lawsuits having a relatively high probability of litigation against auditors.

As discussed earlier, the judgment situations in our experiment pertain to potential asset overstatement. Bigus (2015) found that, consistent with prospect theory, auditors choose to exert caution to a greater extent in the stricter liability regime. A lower threshold probability level of account default, implied by the Lennox and Li and Bigus studies, will inevitably result in greater audit inefficiency. In an experiment, if the perceived audit inefficiency is greater than what is actually observed in practice, unwantedly the state judgments in an experiment will become unrealistic.

Unlike those two studies, Gimbar and Mercer (2021) found that auditors' judgments are more conservative than the desirable level inferred from actual court decisions for alleged misstatement trials. A higher threshold level is hinted by the study. However, raising the threshold level will increase the audit risk, possibly above the desirable level for experiments. Future research could employ various threshold probability levels and compare the results on the subjects' state judgment accuracies across different threshold levels.

Group Probability Judgment Measures

As audits are mostly performed by audit groups, brainstorming is a dominant method of forming group opinions. As discussed earlier, that is even more the case with audit teams that are comprised of auditors and other specialists. Given the popularity of brainstorming and the use of specialists in today's audit engagements, our experiment allowed expert subject groups to form their own collective judgments. Yet, future research could explore an interesting area such as relative performances of various group measures in making probability and state judgments. A good starting point for developing group measures would be the implications found in List (2012) and Martini and Sprenger (2015).

ENDNOTES

1. Moser (1989), using expert subjects, extended an earlier, but similar, study by Hoch (1984). The latter study based on student subjects (i.e., non-experts) showed no significant difference in the subjects' performances in dissimilar information conditions.
2. Financial statements present financial amounts which are the best estimates, given the likelihoods of relevant events. For example, the item of accounts receivable on the balance sheet represents the *final* estimate of the amount to be collected based on the probabilities regarding their collectability. However, the probabilities themselves are not presented to users on the balance sheet.
3. To be clear, in this experiment, subjects make probability assessments performed for a client company, Anderson. The client company has many customers whose accounts may default. The experimental task was to determine the probability that each of Anderson's customer accounts defaults, not the likelihood of Anderson defaulting. Manipulation checks revealed that subjects could successfully distinguish the client (Anderson) from its customers.
4. Given the difficulty in recruiting experts, the number of groups was limited. Based on interviews with auditors, we determined that a group should have three members at a minimum.
5. For instance, if the sample used to test the accuracy rate of the report includes more "Default" predictions than "Paying" predictions, the subjects may think that the accuracy rate of the report is higher when the prediction is "Default" rather than "Paying."
6. If we allowed an error to have either a positive (i.e., assessed probability is higher than correct probability) or negative (i.e., assessed probability is lower than correct probability) sign, a measurement issue would be created. Positive and negative errors could be offset, at least partially, over the 20 experimental periods to make the resulting mean error misleading. Therefore, we measured the probability judgment error by the *absolute value* of the deviation.
7. If the probability of default is high for an account, the resulting state judgments will be "Default." Then, appropriate adjustments will be made to lower the net value of customer accounts.
8. As Table 1 shows, there are four informational situations. In two of the situations, the state judgment based on total BRF probability is the same as that based on the correct probability. Note that each information situation has the same number of judgments made because of the experimental setup (e.g., equal numbers of subject groups in each prior probability saliency condition, equal numbers of "Default" and "Paying" predictions within each of the condition, etc.). Therefore, even without considering the prior probability at all (i.e., total BRF), one can make correct state judgments in 50% of the total judgments made over the 20 experimental periods.
9. Each manipulation requires a division of the sample. Manipulation on two dimensions creates four treatment conditions, and the number of the participants assigned to each condition will be reduced by half. With the number of participants fixed, we would need to lower either the number subjects assigned to each condition or the number of subjects assigned to each group by half, both of which would not be desirable. Given the nature of our experiment, it is especially critical that all groups have at least 3 members (see Endnote 4).

REFERENCES

Ambrus, A., Greiner, B., & Pathak, P.A. (2015). How individual preferences are aggregated in groups: An experimental study. *Journal of Public Economics*, *129*, 1–13.

American Institute of Certified Public Accountants. (1988). *Statement on Auditing Standards No. 56 analytical procedures*. Retrieved from https://www.aicpa.org/Research/Standards/AuditAttest/DownloadableDocuments/AU-00329.pdf

American Institute of Certified Public Accountants. (2002). *Statement on Auditing Standards No. 99 consideration of fraud in a financial statement audit.* Retrieved from https://www.aicpa.org/Research/Standards/AuditAttest/DownloadableDocuments/AU-00316.pdf

Andiola, L.M. (2014). Performance feedback in the audit environment: A review and synthesis of research on the behavioral effects. *Journal of Accounting Literature*, *33*(1–2), 1–36.

Appelbaum, D.A., Kogan, A., & Vasarhelyi, M.A. (2018). Analytical procedures in external auditing: A comprehensive literature survey and framework for external audit analytics. *Journal of Accounting Literature*, *40(1)*, 83–101. https://doi.org/10.1016/j.acclit.2018.01.001

Association of Chartered Certified Accountants. (2009). *International Standard on Auditing No. 620 using the work of auditor's expert.* Retrieved from http://www.ifac.org/system/files/downloads/a035-2010-iaasb-handbook-isa-620.pdf

Bigus, J. (2015). Loss aversion, audit risk judgments, and auditor liability. *European Accounting Review*, *24*(3), 581–606. https://doi.org/10.1080/09638180.2014.899920

Bonner, S.E. (1994). A model of the effects of audit task complexity. *Accounting Organizations and Society*, *19*(3), 213–234.

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

Bonner, S.E., & Walker, P.W. (1994). The effects of instruction and experience on the acquisition of auditing knowledge. *The Accounting Review*, *69*(1), 157–178.

Boyle, E.S., Lewis-Western, M.F., & Seidel, T.A. (2021). Do quarterly and annual financial statements reflect similar financial statement error in the post-SOX era? *Journal of Financial Reporting*, *6*(1), 1–31. https://doi.org/10.2308/JFR-2020-003

Buchheit, S., Dalton, D., Downen, T., & Pippin, S. (2012). Outcome Feedback, Incentives, and Performance: Evidence from a Relatively Complex Forecasting Task. *Behavioral Research in Accounting*, *24*(2), 1–20.

Camerer, C.F. (1987). Do biases in probability judgment matter in markets? Experimental evidence. *American Economic Review*, *77*(5), 981–997.

Cannon, N.H., & Bedard, J.C. (2017). Auditing challenging fair value measurements: Evidence from the field. *The Accounting Review*, *92*(4), 81–114. https://doi.org/10.2308/accr-51569

Christensen, B.E., Glover, S.M., & Wood, D.A. (2012). Extreme estimation uncertainty in fair value estimates: Implications for audit assurance. *Auditing: A Journal of Practice & Theory*, *31*(1), 127–146.

Clor-Proell, S.M., Proell, C.A., & Warfield, T.D. (2014). The effects of presentation salience and measurement subjectivity on nonprofessional investors' fair value judgments. *Contemporary Accounting Research*, *31*(1), 45–66. https://doi.org/10.1111/1911-3846.12041

Dennis, S.A., & Johnstone, M.K. (2018). A natural field experiment examining the joint role of audit partner leadership and subordinates' knowledge in fraud brainstorming. *Accounting, Organizations and Society*, *66*, 14–28. https://doi.org/10.1016/j.aos.2018.02.001

Earley, C.E. (2003). A note on self-explanation as a training tool for novice auditors: The effects of outcome feedback timing and level of reasoning on performance. *Behavioral Research in Accounting*, *15*(1), 111–124.

Eger, C., & Dickhaut, J. (1982). An examination of conservative information processing bias in an accounting framework. *Journal of Accounting Research*, *20*(2 – Pt. II), 711–723.

Fay, R., & Montague, N.R. (2015). I'm not biased, am I? *Journal of Accountancy*. Retrieved from https://www.journalofaccountancy.com/issues/2015/feb/auditing-judgment-bias.html
Ganguly, A.R., Kagel, J., & Moser, D.V. (1994). The effects of biases in probability judgments on market prices. *Accounting Organizations and Society*, *19*(8), 675–700.
Ganguly, A.R., Kagel, J., & Moser, D.V. (2000). Do asset market prices reflect traders' judgment biases? *Journal of Risk and Uncertainty*, *20*(3), 219–245.
Gimbar, C., & Mercer, M. (2021). Do Auditors Accurately Predict Litigation and Reputation Consequences of Inaccurate Accounting Estimates? *Contemporary Accounting Research*, *38*(1), 276–301. https://doi.org/10.1111/1911-3846.12629
Griffith, E.F., Hammersley, J.S., Kadous, K., & Young, D. (2015). Auditor mindsets and audits of complex estimates. *Journal of Accounting Research*, *53*(1), 49–77.
Heiman, V.B. (1990). Auditors' assessments on the likelihood of error explanations in analytical review. *The Accounting Review*, *65*(4), 875–890.
Heiman-Hoffman, V.B., & Moser, D.V. (1995). The Impact of an auditor's initial hypothesis on subsequent performance at identifying actual errors. *Contemporary Accounting Research*, *11*(2), 763–779.
Hoch, S.J. (1984). Availability and interference in predictive judgment. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *10*(4), 649–662.
Hoffman, V.B., Joe, J.R., & Moser, D.V. (2003). The effect of constrained processing on auditors' judgments. *Accounting, Organizations and Society*, *28*(7–8), 699–714.
Hunt, I., & Mostyn, H.J. (2020). Probability reasoning in judicial fact-finding. *The International Journal of Evidence and Proof*, *24*(1), 75–94. https://doi.org/10.1177/1365712719875753
Hux, C.T. (2017). Use of specialists on audit engagements: A research synthesis and directions for future research. *Journal of Accounting Literature*, *39*, 23–51. https://doi.org/10.1016/j.acclit.2017.07.001
Jenkins, J.G., Negangard, E.M., & Oler, M. (2018). Getting comfortable on audits: Understanding firms' usage of forensic specialists. *Contemporary Accounting Research*, *35*(4), 1766–1797. https://doi.org/10.1111/1911-3846.12359
Kinney, W.R., Jr. (1987). Attention directing analytical review using accounting ratios. *Auditing: A Journal of Practice and Theory*, *6*(1), 59–73.
Krumwiede, K.R., Swain, M.R., Thornock, T.A., & Eggett, D.L. (2013). The effects of task outcome feedback and broad domain evaluation experience on the use of unique scorecard measures. *Advances in Accounting*, *29*(2), 205–217.
Lambert, T.A., & Peytcheva M. (2020). When is the averaging effect present in auditor judgments? *Contemporary Accounting Research*, *37*(1), 277–296. https://doi.org/10.1111/1911-3846.12512
Lee, M., Little, H.T., Jr., & Hunt, A.K. (2017). Expert judgments in an audit's analytical review. *American Journal of Management*, *17*(5), 44–59.
Lee, M., Ross, M.T., & Little, H.T., Jr. (2012). An experimental study on individual and group judgments in analytical procedures of audit. *International Journal of Business and Behavioral Sciences*, *2*(6), 31–52.
Lennox, C., & Li, B. (2020). When are audit firms sued for financial reporting failures and what are the lawsuit outcomes? *Contemporary Accounting Research*, *37*(3), 1370–1399. https://doi.org/10.1111/1911-3846.12571
Leung, P.W., & Trotman, K.T. (2005). The effects of feedback type on auditor judgment performance for configural and non-configural tasks. *Accounting, Organizations and Society*, *30*(6), 537–553.
Leung, P.W., & Trotman, K.T. (2008). Effect of different types of feedback on the level of auditors' configural information processing. *Accounting and Finance*, *48*(2), 301–318.
Lin K., Fraser, I., & Hatherly, D. (2000). An experimental study of auditor analytical review judgments. *Journal of Business Finance and Accounting*, *27*(7–8), 821–857.
Lindley, D.V. (1991). Subjective probability, decision analysis and their legal consequences. *Journal of the Royal Statistical Society*, *154*(1), 83–92.

List, C. (2012). The theory of judgment aggregation: An introductory review. *Synthese*, *187*, 179–207.
Martin, R. (2019). Examination and implications of experimental research on investor perceptions. *Journal of Accounting Literature*, *43*(1), 145–169. https://doi.org/10.1016/j.acclit.2019.11.001
Martini, C., & Sprenger, J. (2015). *Opinion aggregation and individual expertise*. Working Paper. University of Helsinki, 00014, Helsinki (Finland). Retrieved from https://iris.unito.it/retrieve/handle/2318/1662568/609203/Martini%20Sprenger%202017%20Expertise%20OUP.pdf
Martins, A.C.R. (2006). Probability biases as Bayesian inference. *Judgment and Decision Making*, *1*(2), 108–117.
Meier, L. (2014). Probability, confidence, and the constitutionality of summary judgment. *Hastings Constitutional Law Quarterly*, *42*(1), 1–72. Retrieved from https://repository.uchastings.edu/hastings_constitutional_law_quaterly/vol42/iss1/1
Miller, M.K. (2008). Judgment aggregation and subjective decision-making. *Economics & Philosophy*, *24*(2), 205–231.
Mohd-Sanusi, Z., & Mohd-Iskandar, T. (2007). Audit judgment performance: Assessing the effect of performance incentives, effort and task complexity. *Managerial Auditing Journal*, *22*(1), 34–52. https://doi.org/10.1108/02686900710715639
Moser, D.V. (1989). The effect of output interference, availability, and accounting information on investors' predictive judgments. *The Accounting Review*, *64*(3), 433–448.
Ng, T.B., & Tan, H. (2007). Effects of qualitative factor salience, expressed client concern, and qualitative materiality thresholds on auditors' audit adjustment decisions. *Contemporary Accounting Research*, *24*(4), 1171–1192.
Nishi, R., & Masuda, N. (2013). Collective opinion formation model under Bayesian updating and confirmation bias. *Physical Review E*, *87*, 062123.
Pinho, C. (2014). The usefulness of analytical procedures. *International Journal of Business and Social Research*, *4*(8), 25–33.
Public Company Accounting Oversight Board. (2015). *Fact Sheet: Staff Consultation Paper on the Auditor's Use of the Work of Specialists*. Washington, D.C.: PCAOB. Retrieved from http://pcaobus.org/News/Releases/Pages/Specialist_Fact_Sheet.aspx
Public Company Accounting Oversight Board. (2018). *Amendments to Auditing Standards for Auditor's Use of the Work of Specialist, Auditing Statement No. 1210*. *PCAOB Release No. 2018-006.* Washington, D.C.: PCAOB. Retrieved from https://pcaobus.org/oversight/standards/implementation-resources-PCAOB-standards-rules/auditors-use-work-specialists
Shanteau, J. (1989). Cognitive heuristics and biases in behavioral auditing: review, comments and observations. *Accounting, Organizations and Society*, *14*(1–2), 165–177.
Smith, S., Tayler, W.B., & Prawitt, D.F. (2016). The effect of information choice on auditors' judgments and confidence. *Accounting Horizons*, *30*(3), 393–408. https://doi.org/10.2308/ACCH-51493
Smith, V.F., & Kida, T. (1991). Expertise and task realism in auditing. *Psychological Bulletin*, *109*(3), 71–93. https://doi.org/10.1037/0033-2909.109.3.472
Stuber, S.B., & Hogan, C.E. (2021). Do PCAOB inspections improve the accuracy of accounting estimates? *Journal of Accounting Research*, *59*(1), 331–370. https://doi.org/10.1111/1475-679X.12339
Tan, H., Ng, T.B., & Mak, B.W. (2002). The effects of task complexity on auditors' performance: The impact of accountability and knowledge. *Auditing: A Journal of Practice and Theory*, *21*(2), 81–95.
Trotman, K.T., Bauer, T.D., & Humphreys, K.A. (2015). Group judgment and decision making in auditing: Past and future research. *Accounting, Organization and Society*, *47*(1), 56–72. https://doi.org/10.1016/j.aos.2015.09.004
Tversky, A., & Kahneman, D. (1974). Judgment under uncertainty: Heuristics and biases. *Science, 185*(4), 1124–1131.