Expert Judgments in an Audit's Analytical Review

Minwoo Lee
Western Kentucky University

Harold T. Little Jr.
Western Kentucky University

Allen K. Hunt
Western Kentucky University

This study used expert subjects in a multi-period experiment to investigate how different information conditions affected auditors' judgments regarding an account misstatement in an audit's analytical review process. The study found that the saliency level of prior probability of the misstatement negatively affected the probability judgment errors. The results also showed that in the situation where the prior probability was less salient, a longitudinal learning effect occurred. Despite errors in probability judgments, the subjects' accuracy rates on specific state judgments were high. It was found that the saliency of prior probability positively affected the accuracy of the state judgments.

INTRODUCTION AND LITERATURE REVIEW

Auditors' Judgments in Analytical Review

Changes in accounting and auditing environments since the passage of the Sarbanes-Oxley Act and the creation of the Public Accounting Oversight Board (PCAOB) have heightened the need for identifying accounts in a company's financial statements which are prone to intentional and unintentional misstatement. One method that meets this need is an audit of a company's financial statements. It results from a systematic process of objective acquirement and evaluation of the evidence regarding assertions about economic actions and events to ascertain the extent to which those assertions correspond to established criteria (American Accounting Association, 1973). The Financial Accounting Standards Board issues the established criteria and the Generally Accepted Accounting Principles. PCAOB also issues the Generally Accepted Auditing Standards. These standards provide a level of assurance as to the accuracy and fairness of an organization's representations of its profitability, financial position, and cash flows. An improperly performed audit could result in providing users with inaccurate financial information and possibly causing stakeholders to make incorrect decisions.

Auditors use analytical review to identify unusual fluctuations in accounting information and determine areas that might require further investigation. Pinho (2014) and Wright and Ashton (1989) found evidence that the use of analytical review makes audits more effective and efficient in detecting misrepresentations. The American Institute of Certified Public Accountants (AICPA) issued statements related to analytical review. Statement on Auditing Standards No. 56 (SAS 56) (AICPA, 1988) identifies the usefulness of analytical review in an audit. Along with SAS 56, SAS 99 (AICPA, 2002) mandates the use of analytical review in planning the audit procedures and determining the extent of the audit. Lin and Fraser (2003) found that analytical review are used extensively in practice, particularly by large firms. Given the importance of analytical review in audits, there have been several studies that shed light on the use of quantitative methods to perform the analytical review. Lee and Colbert (1997) presented possible quantitative tools to use in the analytical review. Also, Blocher, Krull, Tashman, and Yates (2002) introduced business forecasting tools which are useful when forming expectations of financial amounts while performing an analytical review.

As auditors are required to make judgments in analytical review, the focus of research has shifted from quantitative approaches to decision making. Early literature in auditor judgments presented judgment-based analytical review as an efficient tool. It was found to outperform quantitative procedures (Fraser, Hatherly, & Lin, 1997) and shown to have advantages such as ease of use and flexibility (Lin, Fraser, & Hatherly, 2000). As Solomon and Trotman (2003) and Nelson and Tan (2005) reviewed in their surveys, existing and potential research interests in the judgment and decision-making aspects of auditing encompass a number of topics. The research interests include audit tasks, auditors and their attributes, auditor and other stakeholders' interaction in task performance, and judgmental errors in the process by which auditors evaluate analytical evidence. The latter is closely related to the focus of this study.

Auditors obtain assurance from analytical review depending on how consistent the analytical review outcomes are with the expectations developed from other sources (AICPA, 1988). During the planning stage of an audit, auditors perform analytical review, which might signal an irregularity suggesting an account misstatement. Since the signal is not perfectly diagnostic regarding the underlying state, the presence or absence of the irregularity signal is not sufficient evidence as to the misstatement. Therefore, the auditors must assess the probability of the misstatement. Individuals tend to rely solely on the accuracy rate of the analytical review signal in assessing the likelihood of the misstatement. However, to accurately assess the probability of the account misstatement, information independently provided by other sources as well as the accuracy rate of the signal should be considered (Lee, Ross, & Little Jr., 2012). This independently provided information is the normal chance (i.e., prior probability) of the account misstatement.

Auditors' judgments in analytical review regarding the likelihood of an account misstatement are made in the so-called "cab problem" (Tversky & Kahneman, 1974) setting which is known to be subject to base-rate fallacy (BRF). BRF is an error in probability assessment that occurs when the base rate (i.e., normal chance) of an event is ignored or not fully considered. Suppose that normally a particular account is misstated 38% of the time. There is an analytical review mechanism that signals either "Misstatement" or "No misstatement" of the account. The accuracy rate of this mechanism is 60%. Given that the analytical review signal indicates "Misstatement," what is the correct probability that the account is misstated? If an individual completely suffers from BRF, the answer is 60%, which is solely based on the case specific information particular to the case (60%). This answer ignores the base rate information (38%). The correct answer in this situation is 48%. As prescribed by Bayes' Theorem, it is determined by the Bayesian posterior probability that incorporates both the case specific and base rate information.

If auditors are subject to BRF and fail to accurately assess the probability of a relevant event, their ability to perform an effective audit is reduced, resulting in costly over-auditing or risky under-auditing (Kinney, 1987). Therefore, it is of particular interest whether auditors make probability judgments in a correct manner. Previous studies show that systemic bias occurs when an individual makes judgments. In a related literature, Grether (1992) and Tversky and Kahneman (1974) found evidence that the prediction of the likelihood of uncertain events is subject to BRF. According to Martins (2006), the biases in human probabilistic reasoning can be explained by heuristics employed by the individual. In accounting, Joyce and Biddle (1981) and Shanteau (1989) showed that their subjects are biased in their audit judgments. Heiman (1990), Frederick (1991), and Heiman-Hoffman, and Moser (1995) also showed that auditors are partially irrational in making audit judgments. As judgment and decision-making research seeks to understand and improve behavior and decisions of individuals, it is of interest to researchers, auditors, regulators, and users of accounting information. Since the quality of individuals' judgments and decisions is not always optimal, the research in this area is important for auditors to improve in decision-making (Bonner, 1999).

Expert Judgments and Research Issues

Although several studies have presented evidence that accountants are irrational to some extent in making judgments, there has been little research that addresses the BRF issue in the specific probability judgments process of analytical review. Lee et al. (2012) investigated how different information conditions affect the bias generated in auditors' probability judgments in analytical review. When prior probabilities become more salient, individuals are supposed to take prior probabilities into consideration to a greater extent (i.e., become less subject to BRF) and thus generate a smaller error. However, they found conflicting evidence. In their experiment with students, while the accuracy rate of analytical review signal was found to affect the probability judgments errors, the saliency of the prior probabilities was not.

Student subjects usually have little experience in the field of auditing and therefore performed an unfamiliar task in the Lee et al. experiment. As the International Standard on Auditing No. 620 (International Federation of Accountants, 2009) suggests, as complex as they are, analytical review decisions are the tasks that require expertise. Therefore, a study using expert subjects is more appropriate. Bonner (1994) showed that task complexity negatively affect audit performances while skill positively affects it. Also, as Shanteau (1989) and Smith and Kida (1991) suggested, experienced subjects are more familiar with the tasks and therefore may generate smaller judgment biases. Given that, it is of interest to investigate how experts' probability judgments are affected by different information conditions.

This study performs an experiment by utilizing subjects with professional experience and/or auditing knowledge. The purpose of this study is to investigate how different information conditions affect experts' probability and state judgments regarding an account misstatement in the analytical review process of an audit. There are specific research issues in this study. The first issue pertains to the prior probability. Moser (1989) found that in judgment creation the use of information depends on the perceived value of the information. The value of information is determined by the extent to which the information is salient (conspicuous). In the previous literature, it has been shown that prior probability is not fully considered, resulting in errors in assessing the likelihood of an event. Therefore, this study investigates if different saliency levels of the prior probability differently affect the extent to which it is considered and the amount of probability judgment errors made by experts.

Secondly, this study explores the learning effect in experts' probability assessments. Grether (1978) pointed out that the evidence of judgment bias can be discounted if the subjects are not given the opportunity to learn from experience. In an investment decision-making experiment, Ganguly, Kagel, and Moser (1994 & 2000) showed that subjects probability judgments improved over time. This study performs a multi-period experiment to find whether the subjects' probability judgment errors decrease over time as they obtain feedback and whether the saliency level of prior probability affects their ability to learn from the feedback experience.

Finally, this study addresses the issue of specific state judgments. Based on their assessed probability of the misstatement, auditors make judgments or predictions regarding what specific state, misstatement or no misstatement, will actually occur and make the next move assuming a certain state will occur. As the state judgments are a more ultimate action, making correct state judgments would be more important than correct probability judgments. Eger and Dickhaut (1982) showed that when subjects' actions are determined based on the probability judgment, the actions are less biased than probability judgments. This study examines the accuracy of state judgments by experts and also identifies the factors that affect their accuracy.

There has been little research investigating the effect of different informational conditions on auditors' judgments in analytical review. Given that, this study contributes to the behavioral accounting literature by providing the new evidence with expert subjects. Understanding the magnitude and causes of probability judgment errors, the learning effect through experience, and the state judgment accuracy is important in practice also. This understanding will assist the profession in developing training methods which will improve the quality of decisions made by professional accountants and auditors during the planning and final review phases of audit engagements. The paucity of research in this particular area and the relevance of this research to the practice of auditing underscore the importance of this study.

EXPERIMENT

Task and Information Conditions

This experiment employed a hypothetical situation where an account of a company may be misstated. The task of the subjects was to assess the probability that the account was misstated. To investigate whether learning took place over time, the experiment ran for 20 periods, using one company each period. The experiment created two information conditions by manipulating the saliency of the prior probability of the account misstatement.[1] The saliency is a between-subject factor that was set at two levels, 25% (more salient) and 38% (less salient).[2] Figure 1 depicts the experimental design.

FIGURE 1
EXPERIMENTAL DESIGN

Condition 1 (n = 11)	Condition 2 (n = 11)
Prior Probability - More Salient (25%) AR Accuracy (60%)	Prior Probability - Less Salient (38%) AR Accuracy (60%)

Subjects

A total of 22 expert subjects participated in this experiment. In a number of previous studies on various decision-making settings in auditing, college students were used as subjects. One of the weaknesses of these studies including Lee et al. (2012) was that the subjects were not familiar with specific details about the audit process. This study used expert subjects, who had accounting and/or audit experience. Therefore, the subjects performed a familiar task in the experiment.

TABLE 1
SAMPLE COMPOSITION

Field	Occupation	Number of Participants	
Professional	Certified public accountants	6	14
	Accounting staffs	5	
	Financial analysts	3	
Academic	Accounting professors	4	8
	Business professors	2	
	Accounting graduate students	2	
Total	22		

The expert subjects consisted of individuals with experience in either academia or practice. Specifically, participants in the experiment were certified public accountants, financial analysts, accounting professors, business professors, members of accounting staffs, and graduate accounting students. Table 1 presents the composition of the expert participant pool. The vast majority of participants were familiar with how analytical reviews are performed and used. The subjects from each occupation were as evenly divided as possible and randomly assigned to one of the two prior probability saliency conditions.

Procedures

At the start of the experiment, the subjects were informed of the prior probability that a particular account is misstated. The prior probability was either 25% or 38%, depending on the information condition to which the subjects were assigned. The account title was intentionally hidden from the subjects to ensure that they would not develop any prejudice in their judgments.

Analytical review (AR) predictions and their accuracy rate were provided in a context-specific setting that has been known to generate a significant bias in probability judgment. The subjects were told that in each period an analytical review (AR) prediction in the form of either a "Misstatement" or "No misstatement" would be given to them. It was known to the subjects that the accuracy rate of the AR prediction was 60%. Each period, the subjects' task was to assess the probability that the account was misstated. Also, the subjects were told that if their probability assessment was 50% or higher, their prediction regarding the actual state would be "Misstatement" of the account. Otherwise, there state judgement would be "No misstatement."

Further information was given to the subjects. First, the subjects were informed of the makeup of the companies included in the experiment. The subjects were told that they would be presented ten randomly selected companies for which the *AR predicted* "Misstatement" of the account and ten randomly selected companies for which the *AR predicted* "No misstatement." These companies were presented in a random order, one company each period. Secondly, it was revealed to the subjects that the AR accuracy rate was determined by testing the AR with a large sample of companies, half of which misstated the account and half of which did not.[3]

For each of the 20 periods, the following procedures were taken:

1. A *prediction* from AR, either "Misstatement" or "No misstatement," is presented.
2. The subjects assess the probability that the account is misstated in percentage.
3. The *actual* state, either "Misstatement" or "No misstatement," is revealed.

TABLE 2
BAYESIAN AND BASE RATE FALLACY PROBABILITY JUDGMENTS OF MISSTATEMENT

Information Condition	Prior Prob./ AR Accuracy	AR Prediction	Bayesian Probability [a]	Total BRF Probability [a,b]	Diff [c]
1	(25%/60%)	Misstatement	33%	60%	27%
		No misstatement	18%	40%	22%
2	(38%/60%)	Misstatement	48%	60%	12%
		No misstatement	29%	40%	11%

Note: [a] Assessed probability that the actual state is "Misstatement"
If a probability is assessed at 50% or higher, the associated state judgment is "Misstatement." Otherwise, the state judgment is "No misstatement."
[b] Probability assessed solely based on the accuracy rate of AR prediction with the prior probability completely ignored
[c] Difference in % point between total BRF and Bayesian probabilities

The actual state revelation is intended for providing the subjects with the opportunity to learn from experience by taking the actual state information into consideration for their probability assessments in subsequent periods. For all periods, the correct probabilities were revealed to the subjects at the end of the experiment when the payments to the subjects were calculated. Table 2 shows the correct answer, which is the *Bayesian* conditional probability, for each scenario.

Generation of Prediction-Actual State Pairs

Before the experiment was performed, AR predictions and actual states to be revealed to the subjects were generated. For this purpose, the method used by Ganguly et al. (1994) was adopted. The generation of the AR predictions and actual states was based on the probabilities known to the subjects. For example, to generate the pairs for the 38% prior probability condition, first an actual state was determined by randomly drawing a card from a deck of 50 of which 19 (38%) were marked "Misstatement" with the rest "No misstatement." Then, a second card was drawn from another 20 card deck of which 12 (60%) were marked "Correct" with the rest "Incorrect."

The combinations of the marks from the two cards formed the prediction-actual state pairs. For example, if the first card was marked "Misstatement" and the second card was marked "Incorrect", the resulting AR prediction-actual state pair would be "No misstatement-Misstatement." This procedure is repeated until a sufficient number (80) of such pairs is obtained. From these pairs, ten pairs were randomly selected from the pairs for which the AR prediction was "Misstatement" and ten pairs from the pairs for which the AR prediction was "No misstatement." These 20 pairs were randomly assigned to the 20 periods.

Payment Scheme

The subjects were paid based on the differences between the subject's probability assessments and the correct answers. As the experimental currency, franc was used. For each percent of absolute difference from the correct answer, the payment was reduced by two francs from 100 francs. The total payment to a subject was calculated by adding all payments over the 20 judgment periods.[4]

In case the total payment to a subject was below 500 francs, a participation fee of 500 francs would be paid. As an additional incentive, a 5,000 franc gift card was given to each of the top three performers in each prior probability saliency condition.[5] Francs were converted into cash at an approximate rate of $1 per 100 francs.

RESULTS ON PROBABILITY JUDGEMENTS

Probability Judgment Errors in Different Conditions

The probability judgment error is measured by the absolute value of the difference between the subject's assessed probability and the Bayesian conditional probability.[6] The top panel of Table 3 reports the probability judgment errors for different combinations of prior probability saliency levels (between-subject factor) and AR predictions (within-subject factor). In the bottom panel of Table 3, the analysis of variance (ANOVA) reveals the significances of difference in probability judgment errors between two levels of prior probability saliency and two AR predictions along with their interaction. Also, Figure 2 depicts the probability judgment errors in those conditions and the interaction effect.

Overall, the subjects' mean probability judgment error was 13.568% point. This suggests that experts were somewhat biased in probability judgment. It is expected that if the saliency level of prior probability is higher, the subjects' error is smaller. As anticipated, the mean error associated with the more salient (25%) prior probability is significantly smaller than that associated with the less salient (38%) prior probability (12.709% point vs. 14.427% point). The difference was significant (p-value = .041). This particular result suggests the more salient prior probability appeared more valuable to the subjects. Thus, the subjects incorporated the prior probability in their probability assessments to a greater extent and generated the smaller amount of error. This result is consistent with the suggestion by Brody, Coulter, and

Daneshfar (2003) in a different audit setting that probability judgments are affected by the explicitness of the description of the event in question.

TABLE 3
PROBABILITY JUDGMENT ERRORS IN DIFFERENT INFORMATION CONDITIONS

Mean Absolute Deviations from Bayesian			
	AR Prediction		
Prior Probability	Misstatement	No misstatement	Overall
25% (More salient)	11.827	13.591	12.709
38% (Less salient)	14.873	13.982	14.427
Overall	13.350	13.786	13.568

Analysis of Variance				
Factor	df	Mean Square	*F*-value	Significance
Prior Probability	1	324.736	4.209	.041
AR Prediction	1	20.945	.271	.603
Prior Prob. x AR Prediction	1	193.782	2.512	.114

Note: All error amounts are in % point.

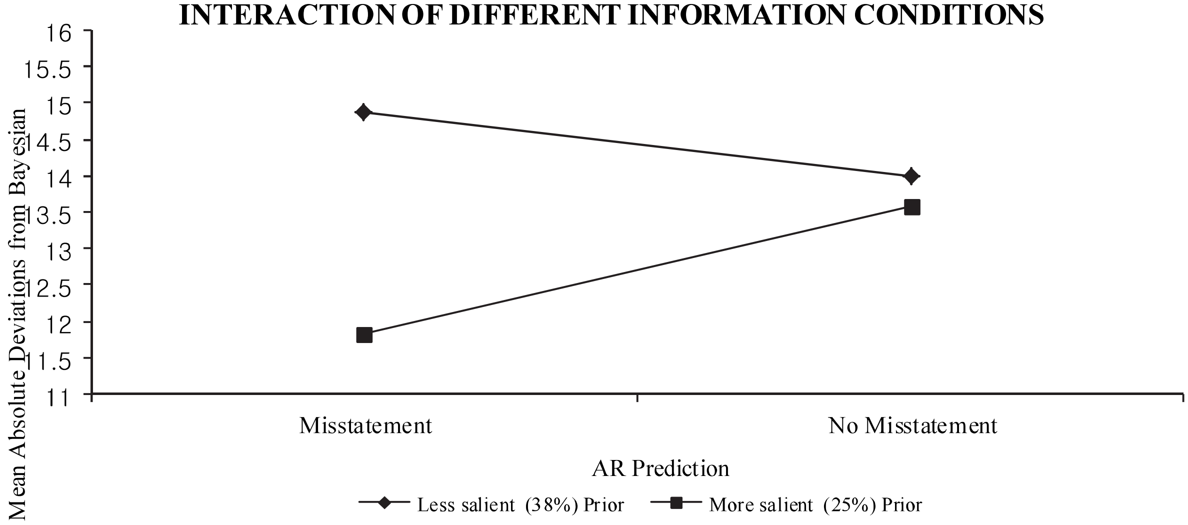
On the other hand, AR prediction was not shown to be a significant determinant of error. The mean errors by the subjects were not significantly different from each other between "Misstatement" and "No misstatement" AR prediction cases (*p*-value = .603). An interaction effect, although weakly significant (*p*-value = .114), was found between the prior saliency and AR prediction. The mean error associated with "Misstatement" AR predictions was smaller than that with "No misstatement" in the 25% prior probability condition, while the opposite was true in the 38% prior probability condition. These offsetting effects may explain the insignificant difference in probability judgment errors between the two AR prediction cases.

The results of this study using expert subjects are different from those of Lee et al. (2012), which used non-expert subjects. Contrary to this study, the Lee et al. study reported no significant difference in the judgment error of non-experts between the more and less salient prior probability conditions. Instead, it found that the AR accuracy rate, which was not analyzed in this study, negatively affected the error. They were found to rely heavily on AR predictions in making judgments. These results suggest that the non-experts subjects suffered from BRF more than the experts. On the other hand, the expert subjects in this study were found to be more rational in probability judgments by incorporating the more salient prior probability to a greater extent.

Learning from Feedback in Probability Judgments

It was shown above that the expert subjects did generate errors in making probability judgments. Given that, this study further investigated the trend of the magnitude of error across experimental periods. In each period, the subjects were presented an AR prediction and the actual state was revealed after the subjects made a probability judgment. Therefore, the subjects were given the opportunity to learn from feedback. It is of interest to test if the subjects appropriately integrated the actual states as feedback information to revise their probability assessments in subsequent periods.

FIGURE 2
INTERACTION OF DIFFERENT INFORMATION CONDITIONS



If a learning effect from feedback occurs, the probability judgment errors should decrease over time. The following regression was employed to test whether the subjects' judgments became more Bayesian (i.e., accurate) with the learning experience:

$$PJE_j = a + b\ EP_j + e_j \qquad (1)$$

where:
PJE_j = probability judgment error for subject j
EP_j = experimental period number (1, 2,...,10) for subject j
e_j = residual term for subject j

The above regression was run separately for each combination of prior probability saliency levels and AR predictions. The EP values were assigned separately for "Misstatement" and "No misstatement" AR prediction cases. For example, the first period with "Misstatement" AR prediction was assigned an EP value of 1, and the last period with the same "Misstatement" AR prediction was identified as 10. The same procedure was taken for "No misstatement" predictions.

Table 4 reports the regression results. For each situation, b, the estimated coefficient of EP, was negative indicating that the probability judgment error became smaller as the subjects learn from feedback by repeating judgments as periods progressed. For the more salient prior (25%) probability conditions, the significance level of the learning effect (represented by b) was not shown significant for either "Misstatement" or "No misstatement" AR prediction. However, the effect was shown to be significant for the condition where the prior probability was less salient (38%). The b coefficients were -0.803 for the "Misstatement" case and -0.573 for the "No misstatement case. The *p*-values of the coefficients were .002 and .018 for "Misstatement" and "No misstatement" AR predictions, respectively.

These results suggest that learning occurred when the prior probability was less salient. As reported in Table 2, the mean error of the less salient prior probability case is significantly greater than that of the more salient prior cases (14.427% point and 12.709% point, respectively). Although the reason that significant learning occurred only in this condition is not clear, it can be argued that the more significant learning occurs if the performance level is inferior, since there is more room for improvement. The results were consistent with the argument.

TABLE 4
REGRESSION OF PROBABILITY JUDGMENT ERRORS ON PERIODS

Condition	a (*t*-stat)	b (*t*-stat)	Signif. of b	Adj.R^2
More Salient (25%) Prior Probability				
Misstatement AR Predictions	14.303 (8.060)	-0.450 (-1.574)	.118	.013
No misstatement AR Predictions	16.509 (7.549)	-0.531 (-1.505)	.135	.011
Less Salient (38%) Prior Probability				
Misstatement AR Predictions	19.291 (12.059)	-0.803 (-3.116)	.002	.074
No misstatement AR Predictions	17.133 (11.542)	-0.573 (-2.395)	.018	.042

Even though the significant learning effect was observed in the 38% prior probability conditions, one could still question whether learning was substantial. To address the issue, the cumulative improvement over nine (first to tenth) periods was examined. In the "Misstatement" and "No misstatement" cases, the cumulative learning effect was -7.227% point (-0.803 coefficient x 9 periods) and -5.157 % point (-0.573 coefficient x 9 periods), respectively. Therefore, on average the subjects can reduce the probability judgment error by those amounts.

As Table 2 shows, in the 38% prior probability conditions, the difference between the total BRF and Bayesian probabilities are 12% point and 11% point for "Misstatement" and "No misstatement" cases respectively. Suppose, in the "Misstatement" AR prediction situation, an individual suffers from total BRF and starts the probability assessments from 60%. The initial probability assessment would be 12% point different from the correct answer of 48%. Over the next nine periods, the individual can improve the probability assessment by reducing the error by 7.227% point. This improvement in probability assessment can be considered of a substantial amount. If the individual suffers from BRF less, his or her probability assessment is less than 12% point away from the Bayesian probability. Then, this 7.227% point is more substantial. A similar argument can be made for the "No misstatement" AR prediction situation.

RESULTS ON STATE JUDGMENTS

Auditors' probability judgments naturally lead to specific state predictions. Based on their probability assessments, auditors make judgments regarding what actual state, either "Misstatement" or "No misstatement", will eventually occur. The state judgments are the ultimate decision that determine whether or not the next course of actions (e.g., adjusting balances of related accounts) are needed. Even though the subjects were found to make errors in probability judgments, it can be argued that the subjects' state judgments, which are a function of the probability judgments, may be less biased (Eger and Dickhaut, 1982).

This study investigated if the expert subjects' state judgment performances were consistent with this argument. Furthermore, it explored the issues of whether subjects' state judgments were systematically different from those derived from random decision-making and how experts' state judgments were affected by different information conditions.

As mentioned previously, in the experiment if a subject's assessed probability of the account misstatement was 50% or higher, a "Misstatement" state judgment was made. To examine the accuracy of subjects' state judgments, correct state judgments were determined first. The correct state judgments are made based on Bayesian probability assessments.[7]

Precision in State Judgments

Table 5 reports the subjects' mean accuracy rates in making state judgments for various scenarios. The overall mean accuracy rate was 0.809. Despite that the subjects generated a considerable amount of errors in their probability judgments, their state predictions were highly accurate. This result is consistent with the finding by Eger and Dickhaut (1982).

The significance levels of reported accuracy rates are also reported in Table 5. For each of the nine cases examined, a binomial test was performed to determine the reported accuracy rate is significantly different from that associated with random predictions. Random predictions would be correct 50% of the time. The binomial test confirmed that in each case the reported accuracy rate was significantly different from 0.5 with a p-value of less than 0.0001. Thus, the results suggest that the reported accuracy rates are significantly higher than those of random predictions.

TABLE 5
PRECISION OF STATE JUDGMENTS

Cases	Accuracy	z-value [a]	Signif.
(1) Entire sample	0.809	12.967	<.0001
(2) More salient prior	0.845	10.248	<.0001
(3) Less salient prior	0.773	8.090	<.0001
(4) Misstatement AR predictions	0.782	8.360	<.0001
(5) No misstatement AR predictions	0.836	9.978	<.0001
(6) More salient prior, Misstatement AR prediction only	0.836	7.056	<.0001
(7) More salient prior, No misstatement AR prediction only	0.855	7.437	<.0001
(8) Less salient prior, Misstatement AR prediction only	0.727	4.767	<.0001
(9) Less salient prior, No misstatement AR prediction only	0.818	6.674	<.0001

Note: [a] based on two-sided binomial test of H_0: proportion = 0.5

Robustness Test for State Prediction Accuracy

With the Bayesian and total BRF probabilities in the experiment shown in Table 2, associated Bayesian and total BRF state judgments can be inferred. A total BRF state judgment is the one which is made solely based on the accuracy rate of AR prediction with no consideration given to the prior probability. It can be inferred that in the situations where the AR prediction is "No misstatement" (two of four experimental situations), the Bayesian and total BRF state judgments are identical. Therefore, it is not clear whether the high accuracy rates reported in Table 5 were due to the experimental setup or the subjects' judgments which were made closer to the Bayesian judgments.

To check for the possibility that the reported accuracy rates were driven by the manipulation of the experimental conditions, the following robustness test was performed for situations where (1) the Bayesian and total BRF state predictions are different from each other; and (2) the associated Bayesian probability is different enough from 50% or the total BRF probability. These two requirements make the robustness check feasible.

Only one information situation satisfies the two requirements above. Consider the situation where the prior probability of misstatement is 25% and AR prediction is "Misstatement." The Bayesian state judgment would be "No misstatement" (with a 33% probability of misstatement), while the total BRF judgment would be "Misstatement" (60% probability).

In this situation, as Table 3 shows, on average, the subjects' probability assessments were different from the Bayesian assessment by 11.827% point. The total BRF probability is 27% point higher than the Bayesian probability (60% less 33%) and any probability triggering the inaccurate state judgment is at least 17% point higher (50%+ less 33%) than the Bayesian probability. The subjects' mean error of 11.872% point (error or difference from the correct Bayesian probability) is much smaller than the 27% point or 17% point difference explained above. Thus, it can be interpreted that the subjects in this situation made the probability judgments quite accurately and their highly accurate state prediction performance was not driven by the experiment's setup.

None of other situations satisfies the two requirements above. Thus, no appropriate robustness test could be performed in those situations. Given that, the robustness test results reported above at least provide partial support that the subjects' performances in predicting states was attributed to their own probability judgment performances, not to the experimental design.

Comparisons of Accuracy Rates across Situations

It may be of interest to find whether different information circumstances affected the accuracy rates in state judgments differently. Table 6 reports the results of the tests comparing the mean accuracy rates and the significances of differences in means. First, the mean accuracy rate was higher when the prior saliency level was higher (0.845 for Case 2 vs. 0.773 for Case 3 in Table 5). The difference is almost significant at the 5% level (p-value = .052 in Table 6). However, the difference in means between the "Misstatement" and "No misstatement" cases (Cases 4 and 5) was not significant.

TABLE 6
COMPARISONS OF MEANS IN DIFFERENT CASES

Comparison of Cases [a]	Mean Difference [b]	Std. Error Difference	t-value	Signif.
Prior Probability More vs. Less Salient (Cases 2 and 3)	0.073	0.037	1.945	.052
AR Predictions Misstatement vs. No misstatement (Cases 4 and 5)	0.055	0.037	1.456	.146
Prior Probability More vs. Less Salient for Misstatement only (Cases 6 vs. 8)	0.109	0.055	1.967	.050
Prior Probability More vs. Less Salient for No misstatement only (Cases 7 vs. 9)	0.036	0.050	0.727	.468

Note: [a] Cases identified in Table 5
[b] Difference is case mentioned first less case mentioned second

The significant difference in state judgment accuracy rates between different prior probability saliency levels (Cases 2 and 3) could be explained by the corresponding significant difference of probability judgment errors reported in Table 3 (12.709% point for Case 2 vs. 14.427% point for Case 3). When the prior probability is more salient, the probability judgment error was smaller and state judgment accuracy rate was higher. Similarly, the insignificant difference in accuracy rates between two AR prediction cases could be explained by the insignificant difference in probability judgment errors in corresponding cases.

Given that the prior saliency level made a difference in the state judgment accuracy while the AR predictions did not, further investigations were performed to compare the state judgment accuracy across the two prior saliency levels with AR predictions held constant. When the AR predictions were "Misstatements", the mean accuracy rate is significantly higher (p-value = .050 in Table 6) with the more salient prior (Case 6 vs. Case 8). However, within the "No misstatement" cases the accuracy rates were not significantly different between the prior saliency level cases (Case 7 vs. Case 9). Therefore, the significant difference in the mean accuracy rates between two different prior probability saliency levels was attributed to the significant difference in the accuracy rates with "Misstatement" cases only. Again, the directions and significances of differences in the state judgment accuracy rates were consistent with those in probability judgment errors reported in Table 3.

Additional tests were performed to determine if the mean accuracy rates were different from those implied by the total BRF or Bayesian judgments. The t-test results, which are unreported, uniformly confirmed that for each applicable comparison, there was a significant difference between the reported accuracy rate and the accuracy rate implied by total BRF or Bayesian judgments. Therefore, it can be concluded that subjects made state judgments in a manner that neither random, total BRF, nor Bayesian judgments can explain.

CONCLUSION AND DISCUSSION

Summary of Findings

This study investigated how different information conditions affected auditors' probability and state judgments regarding an account misstatement in an analytical review setting by using expert subjects. It was found that expert subjects did make errors in making the probability judgments. This study also found evidence that the saliency level of prior probabilities of the misstatement negatively affected the magnitude of probability judgment errors. However, the AR predictions were not found to be a significant factor in determining the probability judgment errors. The results are different from those reported in the previous literature based on non-expert subjects. They found that the magnitudes of probability judgment errors were not affected by the saliency of prior probability.

This study also examined learning effect. In the conditions where the prior probability was less salient, the subjects' probability judgments errors decreased over time and thus their judgments became more Bayesian over time. This learning effect from feedback was shown to be statistically significant and sufficiently large to be meaningful. However, the learning effect was not significant in the situation where the prior probability was more salient.

Finally, this study investigated state judgments resulting from probability judgments. Despite that the subjects made of errors in probability judgments, they showed high levels of accuracy in state judgments. The accuracy in predicting states were found significantly higher if the prior probability was more salient. However, AR predictions were not found to significantly affect the accuracy rates. Also, the results show that reported accuracy rates were significantly different from those implied by any of random guessing, Bayesian judgments, and total BRF judgments.

Limitations and Future Research Directions

Probability and state judgments regarding specific events are integral parts of an analytical review in auditing. In researching the auditors' judgments in analytical review further in the future, it should be noted that there are some limitations with this study. Two issues are discussed to enable future researchers to contribute to a better understanding of the judgments in an analytical review.

Collective Judgments

While performing an audit involves a combined effort by groups of auditors and experts, most audit judgment and decision-making literature has concentrated on individual judgments. Therefore, it is worthwhile to explore the probability and state judgments in the context of collective decision making. In the non-audit context, Camerer (1987) suggested that although individuals are subject to judgment biases, the aggregate outcomes could still be rational since individual biases are random and thus can be canceled by each other. However, in the audit setting, group judgments were not found to outperform individual judgments (Bamber, 1983; Trotman, Yetton, & Zimmer, 1983). Therefore, given this contradicting evidence, it is of particular interest to investigate if group judgments are superior to individual judgments in assessing probability and making state predictions regarding an account misstatement.

There are several studies have dealt with group decision-making regarding the review process of an audit (see a survey of studies by Trotman, Bauer, & Humphreys, 2015). However, despite the relevance of collective decisions in audits, there has been no research using experts on collective probability judgments regarding a specific state of nature. Part of the reason may have been that it is challenging to find appropriate measure of collective probability judgments in analytical review.

Whyte and Sebenius (1997) supported the use of an average in aggregating individual opinions. The rationale for the average of individual assessments is the social decision scheme theory that predicts that groups tend to reach consensus by averaging the pre-group estimates of individuals. However, audits involve group-decision making processes such as hierarchical review, brainstorming, and consultation (Trotman et al., 2015). In this environment, the simple average may not be an appropriate measure of group decisions. Pauly and van Hees (2006), Dietrich and List (2007), and Claussen and Roisland (2010), among others, addressed the issue of how to measure group judgment or aggregate individual judgments as a collective judgment in the non-auditing literature. For expert judgments, Albert, Donnet, Guihenneuc-Jouyaux, Low-Choy, Mengersen, and Rousseau (2012) proposed an approach that combines individual opinions.

In conjunction with these studies, a conceptual model of collective audit review process by Rich, Solomon, and Trotman (1997) would give insight for finding appropriate group measures. Since the results of group probability and state judgments could be sensitive to the group measure used, future studies can enhance the quality of research by using a rigorous measure.

Experimental Design Tradeoffs

Two issues with the experiment's setups in this study are worth noting. First, this experiment manipulated the prior probability of the account misstatement at two levels, 25% (more salient) and 38% (less salient), while holding the accuracy rate of AR predictions constant. To fully address the effects of different information conditions on probability judgement errors and state prediction accuracy, another manipulation on AR accuracy would be desirable. However, there is one thing to note.

It was found in this study that the introduction of additional manipulation could create many situations where it is not clear whether the accurate state judgments are the results of c*orrect* probability assessments or *irrational* BRF assessments, since their state judgments are identical. Because of this ambiguity, even in this study which does not perform an AR accuracy manipulation, interpreting the outcome of subjects' state judgments could be difficult depending on the information scenario. This issue was discussed earlier with the robustness test issue.

In this experiment, as can be seen or inferred in Table 2, the Bayesian state judgements are "Misstatement" in all four situations, and in the 38% prior probability condition the differences between the Bayesian and total BRF probabilities are quite small. Ideally, the experimental design should result in equal numbers of "Misstatement" and "No misstatement" Bayesian state judgments and the difference between Bayesian and total BRF probability assessments are large enough for the robustness test in all situations. Thus far, no combination of prior probability levels and AR accuracy rates leading to that perfect experimental design has been found. Some, not all, of the above requirements for the ideal design could be achieved by increasing one of the prior probabilities to around 60%. However, it may not be realistic to assume a high level of prior probability for an account misstatement beyond a certain level. This remains as a challenge to be overcome in future experiments.

Secondly, the threshold probability can create an issue in making specific state judgments. This study used 50% as the lowest probability level that triggers a "Misstatement" state judgment. However, one could raise a question on whether further investigative actions should be taken only if the probability of misstatement is greater than or equal to 50%. Even though 50% threshold probability is simple, given that an audit should be performed in a conservative manner, a lower threshold probability may be desirable. A survey of auditors might help to identify a realistic threshold probability to be used in future studies.

ENDNOTES

1. Given the difficulty of recruiting expert subjects, to ensure a reasonable number of subjects in each condition and to focus on the effect of prior probability saliency, only one factor was manipulated.
2. Saliency is defined by the extent to which a probability differs from 50%. Therefore, 25% (25% saliency) is more salient than 38% (12% saliency). It refers to how definite a probability is in predicting what state will occur. Note that a 50% probability provides the least amount of information about what state will occur in the future. That is because the occurrence and nonoccurrence of the event is equally likely. A probability is more salient or conspicuous if it is further away from 50%. For example, a 38% probability tells that the event in question has a 38% likelihood and a 62% unlikelihood. As opposed to the 38% probability, a 25% probability implies that the event is less likely (25%) and the opposing event is more likely (75%) to occur. Therefore, the 25% probability is more definite in telling which state will occur.
3. If not informed, the subjects might have a tendency to believe that the AR accuracy rate was more reliable for "Misstatement" predictions if the test sample included more misstatements, and vice versa. For a similar reason, in an experiment of investment decision, Ganguly et al. (1994) revealed the composition of the test sample.
4. The majority of past experimental studies randomly selected one period and paid the subjects only for that period. If only one period is selected for payment, the probability that any given period matters to the subjects is 5%. Then, in any period, the subjects may not do their best in assessing the probability. To keep the subjects motivated to think hard each period, this experiment considered their performances for all periods.
5. Our budget for the experiment was not of an amount to sufficiently pay expert subjects. Therefore, to make the subjects serious about the experiment, the gift cards were given to top performers. To most subjects, money was not an important issue. They were more interested in knowing their relative performance among the participants. Promising to inform them of their ranked performance added further incentive to actively participate in the experiment and make serious judgments.
6. Given that the correct probabilities vary depending on the prior probability saliency levels and the AR predictions, it is not feasible to measure how accurate the subjects were in assessing the probability consistently across different conditions. The alternative measure is the probability judgment error. This measure makes it possible to compare the subjects' performances across different conditions.
7. Note that as Bayesian probabilities are correct answers, the state judgments made on the basis of Bayesian probabilities are always correct. If both Bayesian and total BRF probabilities are on the same side (either less than 50% or no less than 50%), the total BRF judgment (i.e., judgment that completely ignores prior probabilities) is also correct since it results in the same state prediction as the Bayesian judgment. If these probabilities are on the different sides, the total BRF judgment is always incorrect. Refer to Table 2 for Bayesian and total BRF probabilities.

REFERENCES

Albert, I, Donnet, S., Guihenneuc-Jouyaux, C., Low-Choy, S., Mengersen, K., & Rousseau, J. (2012). Combining expert opinions in prior elicitation. *Bayesian Analysis* 7 (3), 503-532.

American Accounting Association. Committee on Basic Auditing Concepts (1973). *Statement of basic auditing concepts*, Sarasota, FL: American Accounting Association.

American Institute of Certified Public Accountants. (1988). *Statement on Auditing Standards No. 56 analytical procedures*. Retrieved June 7, 2015 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 November 10, 2015 from https://www.aicpa.org/Research/Standards/AuditAttest/DownloadableDocuments/AU-00316.pdf

Bamber, E.M. (1983). Expert judgment in the audit team. *Journal of Accounting Research,* 21 (2), 396-412.

Blocher, E., Krull, G.W., Tashman, G.W., & Yates, S.V.N. (2002). Updating analytical procedures. *The CPA Journal,* 72 (11), 36-41.

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.

Brody, R. G., Coulter, J.M., & Daneshfar, A. (2003). Auditor probability judgments: Discounting unspecified possibilities. *Theory and Decision,* 54 (2), 85-104.

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

Claussen, C.A & Roisland, O. (2010). A quantitative discursive dilemma. *Social Choice and Welfare*, 35 (1), 49-64.

Dietrich, F, & List. C. (2007). Arrow's theorem in judgment aggregation. *Social Choice and Welfare*, 29 (1), 19-33.

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.

Fraser, I., Hatherly, D., & Lin, K. (1997). An empirical investigation of the use of analytical review by external auditors, *British Accounting Review,* 29 (1), 35-47.

Frederick, D. M. (1991). Auditors' representation and retrieval of knowledge. *The Accounting Review,* 66 (2), 240-258.

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.

Grether, D.M. (1978). Recent psychological studies of behavior under uncertainty. *American Economic Review,* 68 (2), 70-77.

Heiman, V.B. (1990). Auditors' assessments on the likelihood of error explanations in analytical review. *The Accounting Review,* 65 (4), 875-890.

Heiman, V.B. (1992). Testing Bayes' rule and the representativeness heuristic: some experimental evidence. *Journal of Economic Behavior and Organization,* 17 (1), 31-57.

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.

International Federation of Accountants. (2009). *International Standard on Auditing No. 620 using the work of auditor's expert.* Retrieved March 23, 2016 from http://www.ifac.org/system/files/downloads/a035-2010-iaasb-handbook-isa-620.pdf

Joyce, E., & Biddle, G. (1981). Are auditors' judgments sufficiently regressive? *Journal of Accounting Research,* 19 (2), 323-349.
Kinney, W.R., Jr. (1987). Attention directing analytical review using accounting ratios. *Auditing: A Journal of Practice and Theory,* 6 (1), 59-73.
Lee, M. & Colbert, J. L. (1997). Analytical procedures: management tools for monitoring controls. *Management Decision*, 35 (5), 392-397.
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.
Lin K. & Fraser, I. (2003). The use of analytical procedures by external auditors in Canada, *Journal of International Accounting, Auditing and Taxation*, 2003, 12 (2), 153-168.
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.
Martins, A.C.R. (2006). Probability biases as Bayesian inference. *Judgment and Decision Making,* 1 (2), 108-117.
Moser, D.V. (1989). The effect of output interference, availability, and accounting information on investors' predictive judgments. *The Accounting Review,* 64 (3), 433-448.
Nelson, M. & Tan, H.T. (2005). Judgment and decision making research in auditing: A task, person, and interpersonal interaction perspective. *Auditing: A Journal of Practice and Theory*, 24 (S-1), 41-71.
Pauly, M. & van Hees, M. (2006). Logical constraints on judgment aggregation. *Journal of Philosophical Logic,* 35 (6), 569-585.
Pinho, C. (2014). The usefulness of analytical procedures. *International Journal of Business and Social Research*, 4 (8), 25-33.
Rich, J.S., Solomon, I., Trotman, K.T., (1997). Multi-auditor judgment/ decision making research: A decade later. *Journal of Accounting Literature*, 16 (1), 86-126.
Shanteau, J. (1989). Cognitive heuristics and biases in behavioral auditing: review, comments and observations. *Accounting, Organizations and Society,* 14 (1-2), 165-177.
Smith, V.F. & Kida. T. (1991). Expertise and task realism in auditing. *Psychologica Bulletin,* 109 (3), 71-93.
Solomon, I. & Trotman, K.T. (2003). Experimental judgment and decision research in auditing: The first 25 years of *AOS. Accounting, Organizations and Society,* 28 (4), 395-412.
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.
Trotman, K.T., Yetton, P.W., & Zimmer, I. R. (1983). Individual and group judgments of internal control systems. *Journal of Accounting Research*, 21 (1), 286-292.
Tversky, A., & Kahneman, D. (1974). Judgment under uncertainty: Heuristics and biases. *Science,* 185 (4), 1124-1131.
Whyte, G. & Sebenius, J. (1997). The effect of multiple anchors on anchoring in individual and group judgment. *Organizational Behavior and Human Decision Process,* 69 (1), 75-85.
Wright, A. & Ashton, R. (1989). Identifying audit adjustments with attention-directing procedures. *The Accounting Review,* 64 (4), 710-728.