International Journal of Accounting and Taxation
June 2017, Vol. 5, No. 1, pp. 48-67
ISSN: 2372-4978 (Print), 2372-4986 (Online)

Published by American Research Institute for Policy Development
DOI: 10.15640/ijat.v5n1a5
URL: https://doi.org/10.15640/ijat.v5n1a5

Information Saliency, Auditors' Analytical Assessments and Learning

Minwoo Lee[1] & Hyun S. Hwang[2]

Abstract

This study performed an experimental investigation using expert subjects on how different information conditions affect auditors' probability judgments regarding customer account default in a setting of analytical procedure. This study achieved significant improvements in experimental design issues identified in previous studies in this area, and also addressed unexplored issues. The study found the saliency level of prior probability of the account default negatively affected the magnitude of probability assessment errors. These results based on expert subjects are different from those with non-expert subjects in the previous literature. Previously, the saliency of prior probability was not found to significantly affect the magnitude of probability assessment error. Despite that the subjects were found to generate bias in probability judgments, the magnitude of the error decreased over time with the subjects' learning from feedback. This learning effect was significant in the condition where the prior probability of default was more salient. This significant learning effect resulted in improvement in probability assessments over time with a reduction of the probability judgment error by a substantial amount.

Key words: Probability assessment, Prior probability, Base-rate fallacy, Signals, Learning

[1] Department of Accounting, Western Kentucky University, 1906 College Heights, Blvd. #11061, Bowling Green, KY 42101-1061, USA

[2] Kiwoom Securities, Co. Ltd., 18, Yeouinaru-ro 4-gil, Yeongdeungpo-gu, Seoul, Korea

1. Introduction

1.1. Analytical procedures and probability assessments

In the United States, since the passage of the Sarbanes-Oxley Act (SOX) and the creation of the Public Accounting Oversight Board (PCAOB) auditors have faced greater challenge to identify accounts in a company's financial statements which are prone to misstatement. For providing a certain level of assurance as to the financial statements, an audit is performed on the basis of established criteria and the Generally Accepted Accounting Principles and PCAOB's the Generally Accepted Auditing Standards. In this process, auditors make judgments and provide opinions based on information and test results that are not perfectly accurate regarding the underlying state of nature. After SOX, auditors face increased legal liability for material errors in financial statements globally (Chung, Farrar, Puri, & Throne, 2010), hence, their ability to correctly make judgments in audit is crucial. Auditors use analytical procedures as a diagnostic process to identify and investigate irregularities in financial information (Koonce, 1993). Pinho (2014) found that analytical procedures are an effective and efficient method to detect misrepresentations. There are two types of analytical procedures, quantitative and judgment-based. Fraser, Hatherly, and Lin (1997) and Lin, Fraser, and Hatherly (2000) found that judgment-based procedures that judgment-based procedures outperform quantitative ones in terms of efficiency and effectiveness in error detection.

However, to the extent that analytical procedures require auditor's judgments, the usefulness of judgment-based analytical procedures depends on how accurate the auditors' judgments are. The research on the auditor judgment and decision-making literature has encompassed a number of auditing aspects. Reviews of this literature can be found from Solomon and Trotman (2003) and Nelson and Tan (2005). Among the various auditing aspects, this study focuses on the effect of informational environment on probability assessments in the process by which auditors evaluate evidence from analytical procedures. In a number of situations, auditors have to assess the likelihood of an event such as an account misstatement. This probability assessment is made in the so-called "cab problem" (Treversky & Kahneman, 1974) setting, which is known to generate base rate fallacy (BRF). BRF occurs when an individual either ignores or places less weight on the normal chance (prior probability) in favor of specific case evidence, rather than combining the two, as prescribed by Bayes' Theorem. Consider an audit situation for a client company.[3] One of the customers' accounts of is past-due with a large balance. Normally, there is a 25% chance that receivables as old as this particular one and having a similar balance are not collected.

[3] The same description of the situation was presented to the subjects of this study.

There is a mechanism which is 60% accurate in predicting whether credit customers will pay or not. If the mechanism predicts that the customer will not pay, then what is the probability that the customer will default on the account? If an auditor commits the BRF error and considers only the 60% accuracy rate of the mechanism (specific case evidence) while ignoring the 25% normal chance of default, (s)he will assess the probability to be 60%, which is incorrect.

The correct answer is 38%. It is determined by Bayes' Theorem, which computes the conditional posterior probability by reflecting the normal chance of default as well as the accuracy rate of the mechanism. Note that in the above situation the BRF probability is above 50% while the correct probability is below 50%. If 50% is the threshold probability level that triggers further actions, depending on the correctness of the probability assessment further actions will or will not be taken. Therefore, as Kinney (1987) suggested, auditors' inaccurate probability assessments could result in costly over-auditing or risky under-auditing. There is ample evidence that individuals are biased in making judgments. For example, individuals were found to suffer from BRF when predicting the likelihood of uncertain events (Grether, 1992; Tversky & Kahneman, 1974). Martins (2006) affirmed that the biases in human probabilistic reasoning can be explained by heuristics employed by the individual. In the accounting literature, auditors have been shown substantially biased in making judgments (Shanteau, 1989) or at least partially irrational (Frederick, 1991; Heiman-Hoffman & Moser, 1995).

1.2. Research issues in this study

With the ample evidence from the literature pointing that the quality of auditors' judgments and decisions is not always optimal, judgment and decision-making research seeking to understand and improve behavior and decisions-making of auditors should carry substantial value to researchers, auditors, regulators, and users of accounting information (Bonner, 1999). This study intends to contribute to this literature by exploring new issues. This study is to investigate how experts' probability judgment biases are affected by different informational conditions in the setting of analytical procedures. Our study pertains to the probability of a customer account default, by which related accounts might be misstated. The study focuses on the saliency of the prior probability of the default.

The saliency is the main factor that determines informational condition in this study. As discussed, individuals are found to ignore or under-weigh the prior probability in their judgments. Moser (1989) and Lee, Ross, and Little (2012) suggested that, in the judgment situation, the saliency of information determines the extent to which the information is considered. Desirably, the prior probability information should be fully considered for correct probability assessments. It is of interest to investigate whether and how the probability assessment and its errors are actually affected by the saliency of the prior probability.

Also, this study examines potential learning effect in the long-run judgment process. This particular research question draws on the argument of Grether (1978) that evidence of judgment bias is not conclusive unless there is an opportunity to learn from experience. Ganguly, Kagel, and Moser (1994 & 2000) confirmed that learning occur over time in an investment decision-making situation. To find whether the subjects' probability judgment errors decrease over time in the setting of analytical procedures, this study performs an experiment in a multi-period setting to provide the subjects with feedback each period. Three aspects of learning are investigated: the existence of learning effect, the effect of different information conditions on learning, and the magnitude of learning.

1.3. Methodological Issues and Improvements

In the literature, there is very little research that specifically addresses the BRF issue in the probability judgment setting of analytical procedures. Lee et al. (2012) investigated using student subjects how different information conditions affect the auditors' probability judgment errors. The subjects' judgement errors were not found to be affected by the saliency of the prior probability. Instead, they were found to rely on the predictions of analytical procedures. Therefore, if they received predictions from a more accurate mechanism, they generated smaller errors. The results were consistent with BRF.

Noteworthy is that compared to the knowledge, skill, and experience required to perform a task as complex as an audit, those possessed by student subjects are limited. The task complexity and skill (Bonner, 1994) and experience (Shanteau, 1989) may have affected the performances of the student subjects. In this regard, this study employs expert subjects who have experience in the field of auditing/accounting or business. In a working paper by Lee, Little, and Hunt (2016), the subjects of their experiment were experts in business or accounting. They found that expert subjects are less subject to BRF and therefore less biased when the prior probability is more salient. Our study uses expert subjects. At the same time, it addresses methodological weaknesses of the Lee set al. study and makes improvements. Methodologically, our study has advantages over previous studies in the following aspects. Although using expert subjects has advantages in experiments, there remains a question whether the experiment should be done on an individual or group basis. An audit is usually performed by a group of practitioners.

However, the specific task of assessing probabilities of an event as part of analytical procedures may be done either by a group or an individual. A working paper by Lee (2017) confirmed this by surveying relevant literature. To reflect these two possibilities, this study uses both individual and group participants in the experiment. Regarding the description of the judgment-making situation used in the experiment, this study presents to subjects a clear situation with a cause-and-effect link.

In dealing with the main issue of determining the probability of an account misstatement, neither Lee et al. (2012) nor Lee et al. (2016) gave the subjects the cause of the account misstatement in describing the situation. Unclear situations missing with a cause may affect subjects' judgment performances (Lee, 2017). The last issue pertains to experimental design. Both Lee et al. (2012) and Lee et al. (2016) studies produced experimental situations where the correct Bayesian probabilities are all under 50% or mostly the BRF and Bayesian probabilities are on the same side (both are either above or below 50%). This (near) uniformity in experimental situations makes learning hampered (Lee, 2017). Our study resolved this issue by finding a combination of two prior probabilities and accuracy rate of the analytical procedure mechanism. Given that there are a limited number of studies investigating the effect of informational environment on the auditor's probability judgement, our knowledge in this specific area is also limited. With those improvements noted above, this study provides new experimental results to contribute to the auditing judgment and decision-making literature. This study has practical implications also. Based on the understanding of the probability assessment aspects such as errors and their causes along with the potential learning effect, practitioners can design effective training programs to enhance the quality of their professional judgments.

2. Experimental Design

2.1. Task and experimental manipulation

This study performed an experiment employing a situation which is known to produce BRF. The subjects were presented an analytical procedures setting where an auditor evaluates the accounts receivable account regarding whether the account is overstated. In related previous studies such as Lee et al. (2012; 2016), the decision making situations were quite abstract in that they stated an account could be possibly misstated without given a probable cause. In our experiment, the situation was described specifically. We asked the subject to consider the following situation:

You are an auditor. At your client company, one of the customers' accounts is past-due with a large balance. Normally, there is a x percent probability that receivables as old as this particular one and having a similar balance are not collected. You use an independent mechanism, such as credit agency report, to determine whether the customer is solvent but tardy in paying or simply insolvent. This mechanism signals either "Default" or "Paying" as to the collectability of the account. The signals are known to have a 60% accuracy rate in predicting whether the customer will pay or not.

We asked the subjects to assess the probability that the customer will default. Intentionally, we did not tell the subjects how long the account was past-due or how big the account balance was to ensure that they would not develop any prejudice in making the probability judgments. If the "Default" state will occur later, currently the receivables are overstated. On the other hand, the "Paying" state implies no overstatement.

Figure 1. Experimental Manipulation

Condition 1	Condition 2
(n = 11)	(n = 11)
Prior Probability	Prior Probability
- More Salient (25%)	- Less Salient (65%)
AP Signal Accuracy (60%)	AP Signal Accuracy (60%)

Therefore, effectively this experiment dealt with auditors' judgments regarding an account misstatement. The experiment created two information conditions to which equal numbers of the subjects were randomly assigned. In the first condition, we told the subjects that x, the normal chance of the account default mentioned in the description of the situation above, was 25%. In the other condition, it was known to the subjects to be 65%. These two levels of the prior probability of the account default have different degree of saliency: 25% is more salient than 65%.[4] We manipulated the saliency of the prior probability as the between-subject factor at these two levels. Figure 1 depicts this experimental manipulation.[5]

2.2. Subjects

In some 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 business or accounting/auditing experience. The vast majority of participants were aware of how analytical procedures work in audits. Therefore, the subjects performed a familiar task in the experiment.

---

[4] The saliency of a probability pertains to how definite the probability is in predicting what state will occur. It is measured by the extent to which the probability is different from 50%. Therefore, 25% (25% difference from 50%) is more salient than 65% (15% difference). Note that a 50% probability is least definitive about what state will occur in the future. A 65% probability tells that the event in question has a 65% likelihood and a 35% unlikelihood. If the probability changes to 25%, it 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 predicting the future events. Accordingly, a probability is considered more salient or conspicuous if it is further away from 50%.

[5] We wanted to have a reasonable number of subjects in each condition. If there is another manipulation factor, given the number of total subjects, less number of subjects is assigned to each condition. Because of the difficulty of recruiting expert subjects, this study focused on the effect of prior probability saliency; therefore only one factor was manipulated. This one factor approach actually resolved the last experimental issue discussed in the Introduction.

Table 1. Subject Composition

Field	Occupation		
Professional	Certified public accountants	7	
	Accounting staff	9	
	Financial analysts	8	24
Academic	Accounting professors	7	
	Business professors	3	
	Accounting graduate students	4	14
Total			38

A total of 38 experts participated in this experiment. They were experienced individuals having relevant academic or practical background. Specifically, the participants had occupations such as certified public accountants, financial analysts, accounting or business professors, accounting staffs, and graduate accounting students. Table 1 shows the composition of the participants. Among the 38 subjects, we randomly selected 24 subjects to create eight groups with three members. We instructed them to make the judgments as a group. Usually, audits are performed as a group task. However, as part of an audit, the probability judgments similar to the one in this experiment can be made by the group as a whole or by an individual in the group. Therefore, we used a mix of individuals and groups. In this experiment, the subjects consisted of 14 individual participants and eight group participants. Both of the individual and group participants were randomly and evenly assigned to one of the two information conditions.

2.3. Procedures

At the start of the experiment, we informed the subjects of the prior probability (normal chance) that receivables which are similar to the customer's account in question would not be defaulted. Depending on the information condition manipulated, the prior probability was either 25% or 65%. The experiment ran for 20 periods, for one client company each period. We told the subjects that in each period from a mechanism of analytical procedures (AP), they would receive a signal (prediction) stated as either "Default" or "Paying" depending on whether the customer was expected to default or pay. Also known to the subjects was the accuracy rate of the AP signals, which was 60% for both "Default and "Paying" signals. Each period, we asked the subjects to assess the probability that the customer would default on the account.

Table 2. Bayesian and Base Rate Fallacy Probability Assessments of Default

Condition	Prior Prob./ AP Accuracy	AP Signal	Bayesian Probability [a]	Total BRF Probability [a]	Diff [b]
1	(25%/60%)	Default	33%	60%	27%
		Paying	18%	40%	22%
2	(65%/60%)	Default	74%	60%	14%
		Paying	55%	40%	15%

Note. [a] Assessed probability that the actual state is "Default"
[b] Difference between Total BRF and Bayesian probabilities

We also provided information about the makeup of the client companies. At the beginning, the subjects became aware that they would assess the probability of default for ten randomly selected companies for which the AP *signaled (predicted)* "Default" and another ten randomly selected companies for which the AP *signaled* "Paying." These 20 companies were presented to the subjects, one for each period.

In each of the 20 periods, the experiment was performed with the following procedures:

1. An AP *signal*, either "Default" or "Paying," is presented to the subjects.
2. The subjects assess the probability in percentage that the client company's customer will default on the account in question.
3. The *actual* state, either "Default" or "Paying," becomes known to the subjects.

Each period, we provided the subjects with the AP signal in a context-specific setting that was known to generate biases in probability assessments. Note that after the subjects made the probability judgments, the actual state was revealed. The revelation of actual state provided the subjects with the opportunity to learn from feedback experience by taking the actual state information into consideration for their probability judgments in subsequent periods. We ran the experiment for 20 periods to investigate whether such learning occurred over time. After the subjects made the judgments for all 20 periods, we revealed the correct probabilities for individual periods. Then, we made payment to the groups based on their performance in assessing the probabilities. Table 2 presents the correct probabilities, which are Bayesian probabilities, along with total BRF probabilities for each combination of the prior probabilities and AP signals. A total BRF probability is the one assessed with the prior probability completely ignored.

2.4. Generation of signals and actual states and payment scheme

Before performing this experiment, we generated AP signals and actual states to be revealed in the experiment. This information was generated was based on the prior probabilities of the account default and the accuracy rate of AP signals given to the subjects. The method used by Ganguly et al. (1994) for an investment decision setting was adopted to generate the AP signals and actual states. Regarding payment, we paid the subjects based on performance, which was measured by the absolute value of difference between the subject's assessed probability and the correct answer. The experimental currency was lira. For each period, the payment was 100 liras less one lira for each percentage point of the difference defined above. The total payment for the experiment was the sum of all payments over the 20 periods.

In the majority of past experimental studies, payment was made for only one period which was randomly selected. This method has a disadvantage in keeping the subjects motivated to make probability assessments seriously. Given that the probability that any given period is chosen for the payment purpose is only 5%, it is possible that the subject may not do their best in any period.Unlike most previous studies, by paying all periods, our study could keep the subjects' motivated to think hard throughout all periods in the experiment. As an extra incentive, each of the top three performers in each prior probability saliency condition was paid additional 5,000 liras. In case less than 1,200 liras were earned by a group, as a participation fee 1,200 liras would be paid. We converted the experimental currency of liras into cash based on an approximate rate of $1 per 100 liras.[6]

3. Experimental Results

3.1. Probability assessment errors and saliency of prior probability

Since the correct probabilities differ for different combinations of the prior probability saliency levels and AP signals, we measure the accuracy of the subjects' probability assessments in terms of the magnitude of the probability assessment error.

[6] For this experiment, we had a limited amount of the budget, which was not enough to sufficiently pay experts. However, payment was not a major issue to most of them. They were more interested in knowing how their performance compared to those of others. We promised to inform them of their relative performance. This was the main incentive for the majority of the participants to actively participate in the experiment and keep themselves motivated during the experiment.

Table 3. Probability Assessment Errors of Expert Subjects

Mean Absolute Deviations from Bayesian

	AP Signal		
Prior Probability	Default	Paying	Overall
25% (More salient)	8.859	9.339	9.099
65% (Less salient)	10.346	12.120	11.233
Overall	9.602	10.729	10.166

Analysis of Variance

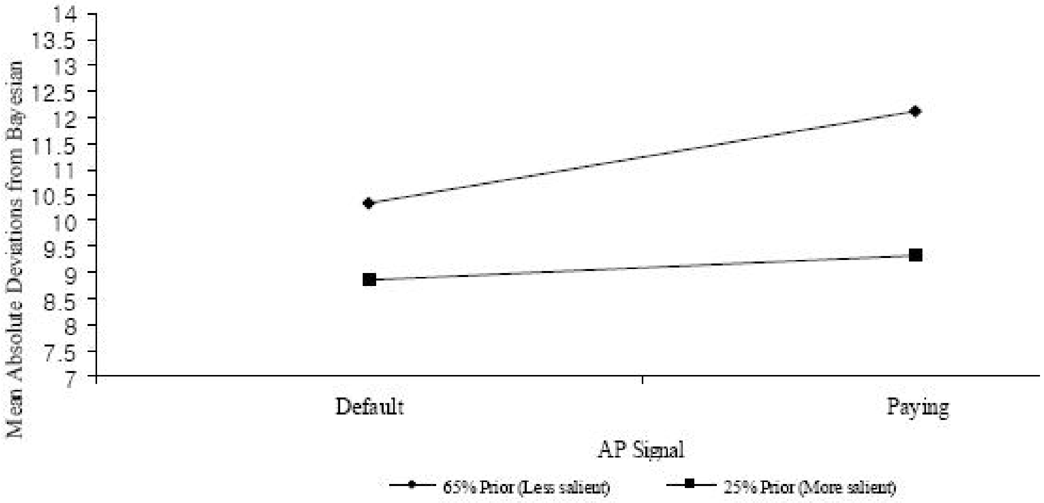
Factor	df	Mean Square	F-value	Significance
Prior Probability	1	501.029	7.492	.006
AP Signal	1	139.678	2.089	.149
Prior Prob. x AP Signal	1	45.982	.688	.407

Note. All error amounts are in % point.

The probability assessment error is defined by the absolute value of the difference between the probability of the account default assessed by the subject and the correct Bayesian probability (presented in Table 2 for each combination). The top panel of Table 3 reports the means of the subjects' probability judgment errors for each of the combinations. The bottom panel shows the results of an analysis of variance (ANOVA) on the magnitude of the error with two factors, the saliency of the prior probability (between-subject factor) and the AP signal (within-subject factor). Figure 2 depicts the errors for different combinations of these two factors and the interaction between them.

Overall, the subjects' mean probability assessment error was 10.166% point. On average, the subjects deviated from the Bayesian probabilities by this magnitude. If the subjects are rational, their errors should be smaller when the prior probability is more salient. As expected, we found that the saliency level of the prior probability and the magnitude of the error were inversely related. (9.099% point for more salient vs. 11.233% point for less salient prior probability conditions). The ANOVA confirms that this difference was significant (p-value = .006). Consistent with Moser (1989), this result suggests that the more salient prior probability appeared to be more valuable information to the subjects and was incorporated in their probability judgments to a greater extent.

Figure 2. Probability Judgment Errors on Different Information Conditions



As to the AP signal, the mean error was smaller when the AP signals "Default" than "Paying." However, the difference between the two were not found significant (*p*-value = .149). Table 3 also reports an interaction effect between the prior probability saliency and AP signals. In both "Default" and "Paying" AP signal situations, the mean error associated with the less salient (65%) prior was greater than that with the more salient (25%) prior. The difference between the two errors was smaller in the "Paying" signal case than the "Default" signal case. However, this interaction effect was not statistically significant.

Lee et al. (2012) performed a similar experiment using student subjects. The overall average probability assessment error by the student subjects was 18.040% point, which is much greater than 10.166% point (in Table 3) with the expert subjects in our study. Note that the two studies have different settings. The description of the judgment situation was clearer in our study and the two studies had different probability values for less salient prior probabilities. Given that, it is not possible to directly compare the amounts of probability assessment errors made by the students and experts. However, in each of two directly comparable situations ("Default" and "Paying" signal cases in the 25% prior probability condition), the mean error of this study was significantly smaller than that of the Lee et al. study. There is additional indirect evidence. The less salient prior probability value is set at 65% in this study while that is set at 38% in Lee et al. (2012) study.

As explained previously, 65% is less salient than 38%. Given the same accuracy rate of AP signal in both studies, with other things held constant, the magnitude of the probability assessment error of this study should be greater, since it is associated with less salient prior probability.

However, the mean error of in this study was smaller than that of the Lee et al. study for both “Default” and “Paying” cases. This could provide additional support to the suggestion by Bonner (1994) and Shanteau (1989) that experienced subjects generate smaller amount of judgment biases. Direct evidence is available regarding the effect of the prior probability saliency on the magnitude of the probability assessment error. As shown in Table 3, the expert subjects were rational in that they generated smaller amount of error when the prior probability was more salient. That means they reflected the more conspicuous prior probability to a greater extent in their probability assessments. The negative relationship between the prior probability saliency and the magnitude of the error was also confirmed by Lee et al. (2016). However, this desirable negative relationship did not exist with the student subjects. Instead, the student subjects were found to rely on the AP signals. Since they heavily depended on the AP signals, without much considering the prior probability, the results were such that students’ probability assessments were more accurate when the AP signals were more accurate, and vice versa. Our results imply that the expert subjects were more rational than the student subjects who suffered more from BRF.

3.2. Learning from feedback in different information conditions

It was found that to a certain extent, the experts did make errors in making probability judgments. Given that, we investigated whether the errors decreased over time. Grether (1978) emphasized the importance of learning opportunities in the judgment research. To address the learning effect, this experiment revealed the actual state, “Default” or “Paying” each period after the subjects assessed the probability of the account default. Thus, the subjects had the opportunity to learn from this feedback information to decrease the errors as the experimental periods progressed.

To test the learning effect, we employed the following regression:

$$ER_j = \alpha + \beta\, PN_j + \varepsilon_j \tag{1}$$

where:

ER_j = probability assessment error for subject j

PN_j = period number (1, 2,...,10) for subject j

ε_j = error term for subject j

Table 4. Learning Effect-Regression of Probability Judgment Errors

Condition	α (*t*-stat)	β (*t*-stat)	Signif. of β	Adj.R^2
More Salient (25%) Prior Prob.				
Default AP Signal	12.627 (8.691)	-0.892 (-1.948)	.054	.092
Paying AP Signal	11.654 (7.549)	-1.047 (-2.207)	.029	.078
Less Salient (65%) Prior Prob.				
Default AP Signal	10.291 (8.976)	-0.698 (-1.620)	.108	.047
Paying AP Signal	12.059 (10.013)	-0.812 (-1.893)	.061	.059

PN values were assigned separately for "Default" and "Paying" cases. The first period in which "Default" is signaled was assigned 1, and the last occurrence was identified as 10. For the "Paying" signal cases, PN values were assigned in the same way.

We ran the regression separately for different combinations of prior probability saliency levels and AP signals. If learning occurs, the amount of the probability assessment error becomes smaller as the subjects get into later periods. Therefore, the β coefficient in Equation 1 should be negative. Table 4 provides the evidence of learning. In each of four combinations of prior probability saliency levels and AP signals, β was negative, suggesting that learning occurred to decrease the error over time. To determine whether the learning was significant, the significance level of the estimated β coefficient was shown in each of the four combinations. It was found significant learning occurred only in the condition where the prior probability was more salient. In that condition, the β coefficients were significant (*p*-values of .054[7] and .029 for "Default" and "Paying" signals, respectively.

In the 25% (more salient) prior probability condition, the subjects' probability assessments decreased over time as the experimental periods progressed. Each period, after the subjects assessed the probability of default, they were informed of what state, either "Default" or "Paying" actually occurred. The subjects were found to utilize this information to revise their judgments to assess the probability more accurately in subsequent periods.

[7] This .054 *p*-value is virtually the same as .05, and therefore considered as significant at the 5% level.

Lee (2017) provided an explanation as to why more significant learning occurs in the condition where the prior probability is more salient. He suggests that if the prior probability is more salient, the actual states occur in a manner that is less consistent with the AP signals. Then, individuals can notice this inconsistency more easily in making their probability assessments. This inconsistency serves as feedback information. Lee et al. (2016) found that significant learning occurred when the prior probability was less salient. Reasons for this anomaly could be their less unambiguous description of judgment situation in their study (discussed in the Introduction).

Since their description was not as specific as that of our study, to their subject the experimental task may have appeared less familiar. As Smith and Kida (1991) suggests, this task unfamiliarity might be the cause of judgmental bias and irregularity found in their study.

3.3. Extent of improvement in probability assessments

Given that significant learning occurred, it is of interest to find whether the learning effect resulted in a substantial improvement in probability judgments of the subjects. To address this issue, we examined the cumulative effect of learning over nine (first to tenth) periods. An analysis was performed for the situations where the learning effect was significant. In addition to "Default" and "Paying" AP signal cases within the more salient prior probability condition, the "Paying" AP signal case in the less salient prior probability condition was also included in this analysis. That was because this case had a *quite*[8] significant β (learning effect).

Table 5 reports the magnitude of improvement in probability assessments due to learning. When the prior probability was more salient and the AP signal was "Default," the learning improved the probability assessment by reducing the probability assessment error by an average amount of 8.028% point (-0.892 β coefficient x 9 periods). In the same way, the learning amounts of other situations can be computed. To determine whether the improvements were substantial, in each situation, we compared the amount of improvement due to learning and the difference between the Bayesian and total BRF probabilities.

[8] Generally, we interpret that significance levels of 5% (.005) or smaller to be "significant" and that of 10% (.100) to be "weakly significant." The significance level of β for "Paying" AP signal case in less salient prior probability condition is .061 (6.1%) as reported in Table 4. It is much closer to 5% than 10%.

Table 5. Learning Effect over Time

Condition	(A) Bayes Prob.	(B) Tot.BRF Prob.	(C) Diff.	(D) β in Eq (1)	(E) Learning amount	(F) Learning Ratio
More Salient (25%) Prior						
Default AP Prediction	33%	60%	27	-0.892	-8.028	0.297
Paying AP Prediction	18%	40%	22	-1.047	-9.423	0.428
Less Salient (65%) Prior						
Paying AP Prediction	55%	40%	15	-0.812	-7.308	0.487

Note: (C) = absolute value of {(A) – (B)} in % point
(E) = (D) x 9 in % point
(F) = Absolute value of (E)/(C)

Again take the case of 25% prior probability with "Default" AP signal. If an individual suffers from total BRF (completely ignoring the prior probability), (s)he initially assesses the probability of the account default at 60%, which is 27% point away from the correct Bayesian probability. Over the next nine periods, this individual reduces the error by 8.028 % point. Compared to this 27% point, the 8.028% point is of a substantial amount of learning. The 27% point difference above between the total BRF and Bayesian probabilities is the maximum possible error. If an individual does not completely suffer from BRF, the difference between the starting probability assessment and Bayesian probability will be smaller than 27%. Then the 8.028% point improvement will be more substantial. The same argument can be made for other situations and substantial learning effect can be confirmed.

For a relative measure of substantial learning, a learning ratio can be computed by dividing the learning amount by this difference. For the three situations in Table 5, the learning ratios ranged from 0.297 to 0.487, suggesting that about 30% to 49% of the difference (errors associated with total BRF) can be reduced substantially by learning from feedback. Again, if an individual's probability assessment is better than that associated with total BRF, the difference between the starting probability assessment and Bayesian probability will be smaller. Then, the learning ratio would be greater.

This learning effect could be a "game changer." Suppose that auditors take the *next action*, such as a write-off a customer's account, if the assessed probability of default is over 50%. In two situations (first and last) in Table 5, the Bayesian and total BRF probabilities are on the different sides of the 50% borderline (i.e, one is over 50% and the other below 50%).

In the two situations, the Bayesian and total BRF probabilities lead to different actions (e.g., one leading to write-off and the other, no write-off). In cases where an individual starts with some reasonable, if not accurate, probability assessment, there may be a good chance that the improvement resulting from learning could make subsequently assessed probabilities cross the 50% borderline toward the correct Bayesian probability. Then the next action would be “correct.”

4. Conclusion and Discussion

4.1 Summary of findings

This study performed an experimental investigation using expert subjects on how different information conditions affect auditors’ probability judgments regarding customer account default in a setting of analytical procedure.This study achieved significant improvements in experimental design issues identified in previous studies in this area, and also addressed unexplored issues. The study found the saliency level of prior probability of the account default negatively affected the magnitude of probability assessment errors. These results based on expert subjects are different from those with non-expert subjects in the previous literature. Previously, the saliency of prior probability was not found to significantly affect the magnitude of probability assessment error.

Despite that the subjects were found to generate bias in probability judgments, the magnitude of the error decreased over time with the subjects’ learning from feedback. This learning effect was significant in the condition where the prior probability of default was more salient. This significant learning effect resulted in improvement in probability assessments over time with a reduction of the probability judgment error by a substantial amount.

4.2. Unresolved issues and future research directions

Probability assessments of specific events are an important aspect of analytical procedures in auditing. Since the judgment process in this assessment task is complex, our understanding in this particular area is limited. For future research, some unresolved issues in this study and others are discussed for performing relevant and rigorous studies in the future. For a better understanding of the judgmental process in analytical procedures, two potential limitations of this study are discussed.

4.2.1. Collective judgment and decision-making

Mostly an audit performance is a combined effort among practitioners. Therefore, it is be desirable that group probability judgments are investigated. If a research intends to investigate auditors' performance in probability judgments, it would be desirable to use a mix of individuals and groups. This was the approach adopted by this study. Finding appropriate relative proportions between the individual and groups in the mix can be an issue. Admittedly, due to budget restrictions, insufficient number of groups was included in the mix. Future studies based on a good mix will find more meaningful outcomes.Regarding researching on group judgements, a research on auditor's collective probability judgments based on experts has its own merits, because there has been little research attempt. There are a number of research questions to address about group probability judgments. Those include questions about how the auditors form collective opinions, which opinion aggregation method leads to the best performance, etc.

Also, a research investigating the probability judgments can be done using expert groups to compare the results with those based on individual experts. Camerer (1987) suggested the cancellation hypothesis that although individuals are subject to judgment biases, the aggregate judgments could still be rational. That is because individual biases are made randomly and the random biases are likely to cancel each other. Given this hypothesis, it is worthwhile to explore expert judgments in the context of collective decision-making and investigate whether group judgments outperform individual judgments. As Trotman, Bauer, and Humphreys (2015) surveyed, group judgment and decision-making studies in auditing has dealt with a number of issues. However, no research has addressed experts' collective probability judgments regarding a specific state of nature.

A research on group probability judgement requires finding appropriate measure of group probability assessments. An understanding of how auditors form aggregate opinions is essential in finding appropriate measures of collective judgments. As audits involve group-decision making processes such as hierarchical review, brainstorming, and consultation (Trotman et al., 2015), a conceptual model of collective audit review process by Rich, Solomon, and Trotman (1997) would be especially important in finding the group measures. In a non-auditing literature, Claussen and Roisland (2010), addressed the issue of how to combine individual judgments.

As to expert collective judgments, Albert, Donnet, Guihenneuc-Jouyaux, Low-Choy, Mengersen, and Rousseau (2012) proposed a method of aggregating individual opinions. In conjunction with these studies, the Rich et al.'s conceptual model would help find appropriate group measures for future studies.

4.2.2. Specific-state judgments

Auditors' probability assessment leads to their judgments regarding what state will occur. Based on their state judgments, the next course of actions is determined.[9] Therefore, making correct state judgments is an important process in providing accurate financial information. The relationship between the probability assessment and the specific state predictions has not been clearly understood. Future studies may explore this unclear link. According to Eger and Dickhaut (1982), actions made based on probability judgments are less biased than the probability judgments. Given that auditors are found to generate errors in probability assessments, it may be of interest to investigate how the errors made in the probability assessments are related to the accuracy of their state judgments.

Two more issues could be addressed relating to the state judgments. It is not known how different information conditions affect the accuracy of state predictions. One can investigate the roles of accuracy rates of analytical procedures along with the prior probabilities in predicting future events. The other question would be regarding the threshold probability level that triggers a certain state judgment. For example, if the assessed probability of a customer default is 40%, do we have to make a judgment that this account will be collectible because the probability is below 50%, or do we have to assume that this account will be uncollectible to act conservatively? Future research may do a challenging work of finding threshold probabilities for efficient and effective audits.

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.

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

[9] For example, if a customer's account is judged uncollectible, related misstated accounts such as accounts receivable and uncollectible accounts expense have to be adjusted.

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

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.

Chung, J., Farrar, J., Puri, P. & Throne, L. (2010). Auditor liability to third parties after Sarbanes-Oxley: An international comparison of regulatory and legal reforms. *Journal of International Accounting Auditing and Taxation,* 19 (1), 66-78.

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-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.

Kinney, W.R., Jr. (1987). Attention directing analytical review using accounting ratios. *Auditing: A Journal of Practice and Theory,* 6 (1), 59-73.

Koonce, L. (1993). A cognitive characterization of audit analytical review. *Auditing: A Journal of Practice & Theory* 12 (Supplement): 57-76.

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., & 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.

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.

Tversky, A., & Kahneman, D. (1974). Judgment under uncertainty: Heuristics and biases. *Science,* 185 (4), 1124-1131.