AUDITING: A JOURNAL OF PRACTICE & THEORY
Vol. 23, No. 1
March 2004
pp. 89–105

The Effects of Immediate Context on Auditors' Judgments of Loan Quality

Dennis M. O'Reilly, Robert A. Leitch, and Douglas H. Wedell

SUMMARY: This study examines whether auditors, when making a series of similar, independent judgments are non-normatively influenced by their earlier judgments. Two judgment biases are considered: contrast effect and assimilation effect.

We conducted an experiment where experienced auditors classified a number of commercial loans based on collection risk. The auditors' judgments displayed a significant contrast effect, and range-frequency theory explains a significant portion of the variance in their judgments (Parducci 1965). The results suggest that auditors tended to use the range principle more than the frequency principle to classify loans. This bias has potential implications for audit practice. By comparing auditors to graduate accounting students we find evidence that task experience moderates the magnitude of the judgment bias.

Keywords: auditor judgment; context; contrast effect; range-frequency theory; bias; loan evaluation.

Data Availability: Please contact the first author for data availability.

INTRODUCTION

Auditors make many judgments regarding company attributes during the audit process. Systematic biases in judgments can lead auditors to different conclusions regarding accounting and auditing issues. We examine whether auditors, when making a series of judgments, have their later judgments non-normatively influenced by earlier judgments (referred to as the "immediate context"). The immediate context can lead to both contrast and assimilation effects in auditors' judgments and task-specific experience may moderate these effects. The study of auditors' susceptibility to these effects is important because they can significantly alter auditors' judgments. In addition, we attempt to better understand auditors' underlying cognitive processes when making a series of judgments using Parducci's (1965) range-frequency theory.

We report the results of an experiment in which participants made a series of judgments about the collectibility of commercial loans where the majority of the loans they evaluated were either high quality or low quality. The primary research question is what effect these "contextual" loans have on auditors' subsequent ratings of medium-quality loans.[1] We are interested in whether judgments made early in the series influence judgments made later in the series. This is related to, but different

Dennis M. O'Reilly is an Assistant Professor at Xavier University, and Robert A. Leitch and Douglas H. Wedell are Professors, both at the University of South Carolina.

The author gratefulley acknowledges the helpful comments of Bill Messier, Arnie Wright, and two anonymous reviewers.

[1] As discussed later in the paper, the subject pool contains auditors, bank examiners, loan analysts, and accounting students. However, our primary interest is in the effects of immediate context on auditor decision making.

Submitted: August 2001
Accepted: June 2003

than, studies that examined Hogarth and Einhorn's (1992) belief-adjustment model. The belief-adjustment model is an "update" model where the decision maker is making a single judgment but revising his or her judgment based on new information. We examine the situation where the decision maker makes a series of *independent* judgments.

It is possible that the immediate context of evaluating loans of one quality level can lead to a contrast effect in judgments of loans of other quality levels. Parducci's (1965) range-frequency theory has been proposed as an explanation of how the distribution of recently experienced stimuli can affect judgment and lead to contrast effects. A contrast effect is a shift in judgment *away* from an anchor point on a scale in a direction opposite of that of the preceding stimuli. In the present study, a contrast effect would be reflected in a tendency for auditors to judge a loan as risky after examining a series of high-quality loans, but judge the same loan as sound after examining a series of low-quality loans.

Other theories, however, suggest that immediate context can lead to an assimilation effect in later judgments. Assimilation refers to a shift in judgment *toward* an anchor point on a scale, in the direction of the majority of the preceding stimuli. Several theories suggest that assimilation effects are produced as the result of biases in attention to and encoding of information. For example, the use of cognitive categories that reflect positive loan qualities in analyzing loans early in the series may prime those categories (Srull and Wyer 1980) increasing the probability that those same categories will be used in organizing one's perceptions of loans later in the series. Assimilation might also result from a categorization process, as is often observed in the literature on the use of group stereotypes in judgment (Biernat et al. 1991). In a loan analysis task, evaluating a preponderance of low-quality loans may build the perception that the loaning agency takes on risky loans. This characterization of the loaning agency may be used in evaluating subsequent loans, producing assimilation.

The determination of how context affects auditor judgment is important because auditors spend a large portion of their time making judgments about various attributes of individual companies, such as the correctness of account balances, adequacy of disclosures, effectiveness and reliability of internal controls, and collectibility of receivables. If context effects lead to systematic biases in auditor judgments, then one auditor's judgment about an item may differ from that of another auditor if each views the same item in different immediate contexts. It is also important that we understand the cognitive process that leads to the auditor's judgment and Parducci's (1965) range-frequency theory helps us better understand this process.

The results of our experiment indicate that auditors are strongly affected by the immediate context when making a series of judgments and that their judgments exhibit contrast effects. Specifically, auditors who initially classified a portfolio of high-quality loans proceeded to classify medium-quality loans as being significantly less collectible than did auditors who initially classified a portfolio of low-quality loans. By fitting the predictions of range-frequency theory to the auditors' judgments, the theory is shown to explain more than 90 percent of the variance in the judgments, suggesting that it offers important insight into the auditors' cognitive process. The auditors' judgments displayed no evidence of assimilation effects. Even in cases where they were provided with a reason to relate the loans to each other (i.e., all loans were from the same bank), participants displayed contrast effects in a magnitude similar to those they displayed when not supplied with a reason to relate the loans to each other. We also examined whether task-specific experience reduces auditors' susceptibility to the contrast effects of the immediate context, as suggested by Messier and Tubbs (1994). Auditors experienced in loan classification displayed significantly smaller contrast effects than students with no experience in loan classification.

The remainder of this article is organized as follows. First, the nature of the possible context effects is reviewed and hypotheses are developed. Second, the methodology employed to test the hypotheses is described. Finally, we present the experimental results and discuss their implications.

HYPOTHESES DEVELOPMENT

Contrast Effects

Research has shown that persons who are asked to rate a series of stimuli along some specified dimension normally shift their judgment regarding a current stimulus away from the bulk of their recent experience (i.e., a contrast effect) if the current stimulus is different from the bulk of their previous stimuli. For example, in one experiment participants who initially evaluated the attractiveness of a series of more attractive faces tended to assign lower ratings to faces of average attractiveness than did participants who had initially evaluated a series of less attractive faces (Wedell et al. 1987). This effect has been demonstrated in a wide range of domains, including judgments of academic performance (Mellers and Birnbaum 1983), judgments of the severity of different crimes (Pepitone and DiNubile 1976), and judgments of psychopathology based on clinical case histories (Wedell et al. 1990). Auditors' susceptibility to immediate context may pose a problem for auditors where consensus and consistency are valued in the judgment process.

Parducci's (1965) range-frequency theory suggests that contrast effects arise from the cognitive application of two basic judgmental principles: a range principle and a frequency principle. Range-frequency theory was proposed by Schepanski et al. (1992, 132) as an explanation of how the same stimuli could be judged differently depending on whether they were presented in a between- or within-subjects design due to differences in the contexts brought to mind under each design. According to the range principle, people tend to cognitively value a stimulus in terms of where its value falls between the maximum and minimum stimulus values brought to mind at the time of judgment. The range value of a stimulus then corresponds to the proportion of the subjective range of stimuli lying below that stimulus value and can be algebraically represented as follows:

$$R_{ik} = (S_i - S_{MIN,k})/(S_{MAX,k} - S_{MIN,k}), \quad (1)$$

where S_i is the context independent scale value of stimulus i, and $S_{MAX,k}$ and $S_{MIN,k}$ are the maximum and minimum subjective values defining the range in context k, respectively. Although these maximum and minimum values may be influenced by the set of stimuli observed by the decision maker, they are not always equal to the observed stimuli and must be inferred from the decision-maker's ratings.

According to the range principle, contrast effects may arise from comparing the stimulus to different endpoints defining the range. For example, the assessment of a student's performance on an exam will depend on the range of possible exam scores one considers. If expectations for exam scores ranged from 40 to 80, then a score of 70 would result in a range value of .75 and be considered a better-than-average or good performance. If expectations for exam scores ranged from 60 to 100, on the other hand, then a score of 70 would result in a range value of .25 and be evaluated as a poorer-than-average or a poor performance. It follows that auditors' judgments may also be subject to a range of outcome expectations. Differences in the contextual range observed in practice may lead individual auditors to define the range differently from each other.

Contrast effects can occur even when the subjective range is equated across sets of stimuli. According to the frequency principle, decision makers value a stimulus in terms of its percentile rank. Formally, the frequency value of a stimulus is the proportion of the total number of contextual stimuli that lie below a stimulus value:

$$F_{ik} = (rank_{ik} - 1)/(N_k - 1) \quad (2)$$

where $rank_{ik}$ is the rank of stimulus i in context k, N_k is the highest rank in the contextual set and 1 is the lowest rank in the set. Thus, even if the subjective range is the same across contexts, a stimulus may be judged higher in one context than another if its percentile rank is higher in that context. For example, an exam score of 70 may be judged good if it corresponds to the 90th percentile, but bad if it corresponds to the 10th percentile. It follows that auditors' judgments may also be affected by differences in how a set of items to be judged is distributed within a given range.

Range-frequency theory assumes that both range and frequency principles affect the decision-maker's cognitive process at the time of judgment, so that the final judgment represents a compromise between range and frequency values (Parducci 1995). Thus, the internal judgment of stimulus i in context k (on a 0–1 scale) can be described as a weighted average of its range and frequency values:

$$J_{ik} = wR_{ik} + (1 - w)F_{ik} \quad (3)$$

where the parameter w describes the weighting of the range value and can vary from 0.0 to 1.0 and R_{ik} and F_{ik} are defined in Equations (1) and (2). In psychophysical experiments, the value of w is often close to 0.5, representing a roughly equal compromise between range and frequency principles (Birnbaum 1974; Parducci and Perrett 1971). In social judgment domains, w tends to be much higher, such as $w = .80$ (Wedell 1994).

In cases where a person's judgment must be categorized, range-frequency theory typically assumes a linear relationship between the subjective judgment and the mean category rating:

$$C_{ik} = a + bJ_{ik} \quad (4)$$

where C_{ik} is the category rating of stimulus i in context k, J_{ik} is defined in Equation (3) and a is the value assigned the lowest rating category and b is the range of category values. For example, with a five-point scale such as the one used in this study, $a = 1$ and $b = 4$.[2] Equations (1)–(4) then describe how category ratings can be related to stimulus values given that the subjective scale values (S_i's) of Equation (1) and ranks of Equation (2) are known. The adequacy with which range-frequency theory explains participants' judgments in an experiment can be determined by fitting the predictions of the range-frequency model to those judgments.

Participants were required to classify each of nine loans into one of five ordinally ranked categories. Range-frequency theory predicts that participants' categorical judgments will be affected by the immediate context within which each loan is examined. Further, range-frequency theory predicts that their judgments will display a contrast effect. That is, as the participants work through the portfolio of loans their later judgments, regarding medium-quality loans, will shift away from their earlier judgments. This judgment shift will result in differential ratings for medium-quality loans that are common across all loan portfolios (cells) depending upon the context of the initial set of loans.

H1: When making a series of similar, but independent, judgments auditors, as well as students, will tend to shift later judgments in a direction opposite that of their earlier judgments.

Assimilation Effects

An assimilation effect describes the tendency for people's judgments to shift toward the bulk of their recent experiences. Several theories have been proffered in the psychology literature to explain assimilation effects in judgment. Manis and Paskewitz (1984) theorize that assimilation effects may result from global expectations in which the decision maker expects future stimuli to resemble the stimuli they have recently encountered. This bias is similar to anchoring and adjustment biases discussed by Tversky and Kahneman (1974), Joyce and Biddle (1981), and Kinney and Uecker (1982), among others.

[2] The idea is to match the subjective judgment scale on a 0–1 scale to the rating scale of interest. This means that the intercept will be the lowest category where $J_{ik} = 0$ and the highest category should correspond to $J_{ik} = 1.0$ so that the slope must be $(Cmax - Cmin)/(Jmax - Jmin)$. In the five-category loan-classification scale used in this study this is $(5 - 1)/(1 - 0) = 4$ for the slope and 1 is the intercept.

Other theories of assimilation are based on the idea of "priming" (Higgins et al. 1977; Srull and Wyer 1980). In experiments illustrating assimilative priming, participants are typically incidentally exposed to extreme stimuli in the initial phase of the experiment and then asked to judge an ambiguous stimulus. The initial exposure is assumed to prime the corresponding evaluative category so that it is more readily applied to the ambiguous stimulus, producing an assimilation effect. Priming theories of assimilation imply that assimilation is most likely to occur when the stimuli being judged are ambiguous, with contrast predicted for unambiguous stimuli (Herr 1986).

Assimilation may also result from a general categorization process, as is typically found in the literature on stereotyping (Brewer 1988). In this case, the initial contextual stimuli are used to form a prototypical value for the category. When judging an instance of the category, the judgment shifts toward the value of the category prototype (Anderson 1966; Biernat et al. 1991; Manis and Paskewitz 1984).

The categorical basis of assimilation illustrated here implies the potential importance of the source of the loans as a moderating effect on the hypothesized contrast effect. If contextual (high- or low-quality) loans are attributed to the same (different) loaning agency (agencies) as (than) the medium-quality loans being evaluated, assimilation would (not) be expected. To test this, participants in our experiment were either told that all of the loans being evaluated were originated by the same bank (increased likelihood of assimilation) or that each loan was originated by a different bank (decreased likelihood of assimilation). Since an assimilation effect should result in an effect that is the opposite direction from that of the previously hypothesized contrast effect, we expect it to moderate the contrast effect and result in an interaction.[3]

Based on the above, the following hypothesis is tested:

H2: When making a series of similar, independent judgments *and* given a reason to make overall assumptions about the set of stimuli being evaluated, auditors will shift their later judgments toward their earlier judgments.

Task-Specific Experience

The task selected for this study is the evaluation of commercial loans. Due to the large number of regulations that affect the financial reporting of financial institutions, bank auditors are usually CPAs who specialize in the audits of financial institutions. This specialized experience may limit auditors' susceptibility to context effects in two ways. First, experienced auditors may tend to utilize a broad context based on their past experiences with this task so that the effects of immediate context, in particular contrasting stimuli, are diluted (Messier and Tubbs 1994). In other words, an auditor may invoke base-rate knowledge acquired through experience (Heiman 1990; Libby and Frederick 1990; Tuttle 1996). Second, experienced auditors may have set rules that they use for these classification decisions so that their judgments are less subjective than those made by auditors possessing less experience with the task.

Previous research in accounting suggests that task-specific experience and training (Bonner 1990 and Bonner and Lewis 1990; Frederick et al. 1994) enhance auditors' ability to process information and to detect various patterns in decision-making situations. Focusing on the task of loan analysis, two studies in the accounting literature suggest that experience is associated with improved decision making. Jeffrey (1992) found that auditors with relatively greater levels of task-specific experience were less likely to exhibit escalation of commitment and Wright (2001) found that task

[3] An experiment by Wedell et al. (1987) demonstrated that assimilation and contrast effects can occur simultaneously. Theoretically, these processes can occur in the same judgment because they may operate at different stages of cognition. For example, assimilation may take place prior to range-frequency integration (as in priming based assimilation) or after range-frequency integration (as in anchoring and adjustment). In their experiment, Wedell et al. (1987) demonstrated that assimilation occurred after contrast.

experience is positively associated with less judgment bias and increased judgment consensus in auditors' assessments of loan collectibility.

Based on studies that support the beneficial effects of task experience, we posit that task experience minimizes the susceptibility of auditors to the effects of immediate context. This postulate is tested through the following hypothesis:

H3: When making a series of similar, independent judgments, task experience reduces auditors' susceptibility to the effects of immediate context.

METHOD

Experimental Design

A 2 × 2 × 2 factorial design experiment was used to test the hypotheses. Three between-subjects variables were manipulated in the experiment: *Context*, *Loan Origination*, and *Task Experience*. *Context* was manipulated by providing participants with loan portfolios that included six loans that were either high-quality loans or low-quality loans followed by three medium-quality loans. *Loan Origination* was manipulated by altering whether all of the loans were from the same bank (thus, increasing the possibility of an assimilation effect) or that each loan was from a different bank (thereby reducing the possibility of an assimilation effect). *Task Experience* was manipulated by sampling professionals with experience in loan analysis and graduate students with no experience in loan analysis. The dependent variable is the *average* of the ratings each participant assigned to three medium-quality loans that were common across all experimental cells.

A second group of professionals provided context-free ratings for all the loans used in the experiment. These ratings are necessary for fitting the predictions of range-frequency theory to the first group of professionals' ratings as a means of understanding, rather than simply describing, their judgment and decision processes.

Task

The loan-classification task used in this study is derived from the Office of the Comptroller of the Currency's (OCC) requirement that banks classify all loans into one of five categories:

1. **Current**—normal acceptable banking risk.
2. **Special Mention**—evidence of weakness in the borrower's financial condition or an unrealistic payment schedule.
3. **Substandard**—severely adverse trends or developments of a financial, managerial, economic, or political nature, which require prompt corrective action.
4. **Doubtful**—full repayment of the loan appears to be questionable. Some eventual loss seems likely. Interest is not accrued.
5. **Loss**—loan is regarded as uncollectible and should be written off. (AICPA 1996)

The classification of a loan is a subjective decision made by persons such as bank examiners, loan analysts, and external auditors. Although there is no prescribed evaluation procedure, loan classifications reflect judgments based on ratios, trends, and subjective evidence concerning the borrowing company's management, industry, and future prospects. The loan-classification process is part of the recommended procedure for determining the adequacy of a financial institution's loan loss reserve (AICPA 1996).

We chose to investigate this particular auditing task for two reasons. First, the loan-classification task is semi-structured and indicative of the types of judgments that auditors make. Second, it provides an opportunity to test auditors' susceptibility to the effects of immediate context on a task with which they have experience and one where previous studies have shown that experience results in higher-quality judgments (Jeffrey 1992; Wright 2001).

Materials and Procedure

The instrument used in the experiment provided participants with the AICPA definitions of the five OCC loan-classification categories as well as a portfolio of nine loans. Participants were asked to classify each loan into one of the five OCC categories. In the experiment contextual effects were maximized by presenting the six context loans (high or low quality) first and then followed by the three medium-quality loans that were common to all cells. To obtain context-free ratings for each loan, a second group of professionals received in random order all 15 loans that had been used in the experimental conditions.

For the sake of efficiency, certain variables important to a loan-classification judgment were held constant across all loans. These variables included, among others, the outstanding balance, loan term, and payment schedule. The participants were also told to assume that all borrowers operated in the same industry, were approximately the same size, and had provided the lending bank with audited financial statements.

Loan-specific descriptions consisted of six financial ratios and net sales for a three-year period. Although there are a multitude of financial ratios that could have been used, ratios for this study were selected based on previous research (Abdel-khalik and El-Sheshai 1980; Dietrich and Kaplan 1982)[4] as well as discussions with a credit analyst and two CPAs experienced in auditing banks. The information provided to the participants coincides closely with that required by Wright and Willingham's (1997) computational model of loan loss judgments for loans treated as unsecured. A portion of the instrument is included in the Appendix.

Initial ratings for the 15 loan descriptions were established by use of the Dietrich and Kaplan (1982) model for predicting loan classifications followed by consultation with two CPA firm partners experienced in classifying commercial loans. The goal was to create a portfolio of loans that would cover the full range of the OCC's classification scale. The instrument was pretested by students enrolled in a graduate auditing course (none of whom participated in the experiment) and minor modifications were made to the instrument based on feedback received through the test.

Participants

Two groups of participants completed the experiment. The first group was made up primarily of auditors but included several bank examiners and loan analysts as well. Restricting the data analysis to only those professionals with auditing experience did not qualitatively alter the results. Therefore, we only report results that include all of the professionals. The second group was made up of graduate accounting students. These two groups are referred to as the auditor and student groups, respectively.

Participants in the auditor group were obtained by randomly sampling 420 AICPA members who had designated a banking specialty in their AICPA profile. Each person was randomly assigned to one of four experimental groups and mailed one version of the instrument. Responses were received from 115 members, 113 of who provided usable data (a 27 percent response rate).

Participants in the student group were graduate accounting students from a large state university. Of the 44 students sampled, three did not respond, resulting in a final sample of 41 students, all of whom reported having no prior experience in classifying bank loans. As with the professionals, they were given the case and asked to return it anonymously. Only general demographic information regarding experience was collected.[5]

---

[4] Abdel-khalik and El-Sheshai (1980) allowed loan officers to choose information cues to assist in predicting default on debt. We used the most common information cue choices as a starting point in designing the instrument. Dietrich and Kaplan (1982) regressed a single bank's lending officers' loan classification judgments on various financial variables. The result was a simple linear model that explains and predicts their loan classification judgments. The model consisted of a debt ratio, a cash flow measure, and a variable to capture negative sales trends. Consistent with their model, our instrument provides information about debt levels, cash flows, and sales trends.

[5] Those students who returned the instrument were given a small amount of extra credit for a graduate accounting systems course. A record of those who returned it was kept but was not linked to a specific instrument.

To obtain context-free loan judgments we had a second group of professionals classify all 15 loans used in the experiment. The order of these 15 loans was randomized. These ratings are used both to test the hypotheses and to empirically fit the range-frequency model. This group is referred to as the baseline group. Participants in the baseline group were obtained by randomly sampling an additional 80 AICPA members from the same sample frame noted above. Seventeen members of this group responded (a 21 percent response rate).

Table 1 provides demographic information about both groups of professionals (auditor and baseline) who participated in the study. One hundred twelve (86 percent) of the 130 professionals who participated reported that the source of their banking experience was as an auditor. The remaining 18 participants (14 percent) reported having experience as a bank examiner or loan analyst.

RESULTS

Context Effects and Task Experience

Table 2 presents the results of a three-way between-subjects ANOVA with the dependent variable measured as the average ratings of the three loans that were common across all cells for the auditor and student context groups. Table 3 presents the combined means and standard deviations of the categorical ratings for the three common loans for both the auditor[6] and the student groups. The main effect for context is highly significant ($F = 214.39$; $p < .001$). An examination of the means in Table 3 suggests that the immediate context led to a contrast effect in both auditors' and students' judgments, thus, supporting H1. Auditors who were exposed to a portfolio of mostly low-quality loans rated the three common loans as more sound (1.97) than did those auditors who were exposed to a portfolio of mostly high-quality loans (3.18) ($t = 9.76$, two-tailed $p < .001$). Similarly, student participants in the low-quality context rated the common loans as more sound (1.52) than did students in the high-quality context (3.81) ($t = 10.93$, two-tailed $p < .001$).[7]

TABLE 1
Demographic Information of Professionals by Auditor and Baseline Group

	Auditor Group Context Loans[a]		Baseline Group	
Number of Participants by Source of Task Experience	High-Quality	Low-Quality	No-Context	Total
Auditor	43	54	15	112
Bank examiner/Loan analyst	10	6	2	18
Mean experience analyzing commercial loans	14 years	11 years	9 years	

[a] Context is determined by whether the participant reviewed a portfolio of loans that were either high-quality or low-quality.

[6] Qualitatively similar results follow when the data set is limited to ratings made by auditors with other professionals excluded.

[7] These results are further supported by nonparametric tests that show significant differences in median category ratings for each of the three common loans between the high-quality and low-quality contexts for both the auditors and the students. The direction of the differences suggests the presence of contrast effects. To further examine whether both contextual groups displayed contrast effects we compared auditors' context-influenced loan ratings to the context-free loan ratings of the baseline group. Mann-Whitney U nonparametric tests reveal that five of the six comparisons show contrast effects that are significant at the .05 level. These results suggest that both contextual conditions led to contrast effects in auditors' judgments. Results of nonparametric tests are available from the authors.

TABLE 2
ANOVA Results for the Mean Loan Ratings[a]

Source of Variance	df	Mean Square	F-statistic	p-value
Context[b]	1	91.398	214.389	< 0.001
Loan Origin (Assimilation)[c]	1	0.596	1.399	0.239
Task Experience[d]	1	0.180	0.423	0.516
Context × *Loan Origin*	1	0.001	0.144	0.992
Context × *Task Experience*	1	8.419	19.749	< 0.001
Loan Origin × *Task Experience*	1	0.056	0.131	0.718
Context × *Task Experience* × *Loan Origin*	1	0.092	0.217	0.642
Error	146	0.426		
Adjusted R^2 = .501				

[a] Three loan descriptions were common across all cells.
[b] *Context* is determined by whether the participant was provided with a portfolio of loans that were primarily high quality or primarily low quality.
[c] The loans were described as either being all from the same bank or all from different banks.
[d] Experience is a dichotomous measure between professional auditors and students.

TABLE 3
Mean and Standard Deviations of Participants' Category Ratings[a] for the Three Common Loans

	Professional Auditors Loan Origin			Students Loan Origin		
Context: Quality of Non-Common Loans	**Same Bank**	**Different Bank**	**Combined**	**Same Bank**	**Different Bank**	**Combined**
Low-Quality						
Mean	2.034	1.903	1.97	1.59	1.43	1.52
Std. Deviation	0.62	0.71	0.66	0.36	0.45	0.40
n	29	31	60	13	10	23
High-Quality						
Mean	3.32	3.08	3.18	3.83	3.79	3.81
Std. Deviation	0.71	0.61	0.66	0.77	0.92	0.82
n	23	30	53	10	8	18

[a] Participants classified each loan into one of five ordinally ranked categories: 1 = Current, 2 = Special Mention, 3 = Substandard, 4 = Doubtful, and 5 = Loss.

See Table 2 for variable definitions.

To examine whether task experience moderates auditors' susceptibility to the effects of immediate context we compared the loan ratings of the auditor and student groups. Hypothesis 3 predicts an interaction between context and experience and Table 2 shows that a significant interaction occurred ($F = 19.749$, $p < .001$). The mean ratings shown in Table 3 suggest that the source of the interaction is that the auditors' judgments display smaller contrast effects than those of the student participants.[8] Thus, we find support for H3—task-specific experience reduces, but does not eliminate, auditors' susceptibility to the effects of immediate context.

Moderating Effect of Assimilation

The above analysis found clear evidence of a contrast effect in auditor and student context group's judgments. However, it is possible that assimilation occurred as well but was overwhelmed by the contrast effect. Recall that many of the theories that predict assimilation require the decision maker to have a basis for making an overall assessment of the stimuli. In the experiment, approximately one-half of each of the two context groups was told that all of the loans were from the same bank while the other half of each group was told that each loan was from a different bank. Participants in the "same bank" condition, therefore, have a reason to make overall assumptions about the set of loans being evaluated that the other participants lack. If the former group produced assimilation effects significantly greater than the latter group, then a significant interaction of context and loan origin should be present.

Table 2 indicates that neither the main effect for loan origin nor its interaction with context is significant ($F = 0.144$; $p = .992$). Therefore, H2 is not supported. Table 3 shows that the mean ratings for the three common loans did not differ much within contextual condition depending on whether the loans were from the same or different banks. Thus, there is no support for any moderating assimilation effect related to the origin of the loans. This lack of support should be considered tentative, however, because the manipulation of loan origin may have lacked operational strength.[9]

Fit of the Range-Frequency Model

Once the data revealed that the context groups' judgments had displayed a contrast effect, we further analyzed the data by fitting the predictions of range-frequency theory to the auditor and student groups' loan judgments. This is done separately for comparative purposes. A close fit would suggest that the theory is providing insight into how participants in these groups cognitively processed the loan information under the two context conditions.

First, we used the mean ratings of the baseline group of professionals who were exposed to the full set of 15 loans as scale values for (S_i's) in Equation (1). This assumption seems justified because the distribution of loans for the full set appeared to be uniform so that mean ratings should reflect scale values. In this way, the 15 loans were placed on the five-category OCC classification scale with 1.0 corresponding to the highest-quality loan (S_{MIN}) and 5.0 corresponding to the lowest-quality loan (S_{MAX}).

[8] Nonparametric tests of the differences between the ratings of auditors and students for each of the three common loans were performed. The differences are significant at the .05 level for two of the three loans for each context. Consistent with the results of the parametric tests, the differences indicate that contrast effects are significantly smaller in the auditors' judgments than in the students' judgments. To explore whether greater levels of experience are associated with lesser degrees of judgment bias, regression analysis was conducted on the auditors' loan judgments using years of task experience as a continuous variable. Results of the regression analysis (not reported) indicated that this variable did not have explanatory power ($p = 0.464$) nor did it interact with the context variable. Furthermore, partitioning the data based on various levels of experience did not alter the results qualitatively.

[9] No manipulation check was performed to assess the strength of the loan origin manipulation.

We assumed that when an end stimulus of the contextual range presented to the auditor and student context groups matched an end stimulus of the full range condition, the corresponding subjective minimum or maximum also matched. Therefore, S_{MIN} was assumed to equal 1.0 for the high-quality portfolio condition because this group of participants was exposed to the high-quality loans. Similarly, S_{MAX} was assumed to equal 5.0 for the low-quality portfolio condition because this group of participants was exposed to the low-quality loans. Thus, only three parameters were fit to the 18 data points: the value of the weighting factor (w), S_{MIN} for the low-quality context and S_{MAX} for the high-quality context. The data points consisted of the six high-quality and six low-quality loans plus three medium-quality loans. We used an iterative nonlinear regression procedure (with a least squares loss function) to fit the data, combining Equations (1)–(4) and substituting for the mean category rating:

$$C_{ik} = 1 + 4\,[w\,(S_i - S_{MIN,k})/(S_{MAX,k} - S_{MIN,k}) + (1 - w)(rank_{ik} - 1)/(N_k - 1)] \quad (5)$$

where the intercept a = 1 and the slope b = 4.[10]

Figure 1 presents the fit of the range-frequency model to the data for the auditor and student groups. Note that the spacing of loans along the horizontal axis corresponds to the loan-quality scale values, which were equated to mean ratings from the baseline group with low numbers representing higher-quality loan ratings. The open and solid points are the mean ratings and the dashed lines are the range-frequency predictions for each context, with only three parameters free to vary. The minimum and maximum range values are those that the model posits auditors and students have in mind as they classify the loans in each context. They are represented in Figure 1 by where the dashed lines for each context would have crossed the lower and upper horizontal axis respectively. The diagonal represents the ratings the loans received in the full range condition of the baseline group of professionals.

The data are fairly well described by the model, with 94.6 percent and 91.0 percent (R^2) of the variance in the mean ratings predicted by the model for auditor and student groups, respectively.[11] The slightly better fit to the auditors' data reflects the greater correspondence of the auditors' assessments to the initial scale values. As shown in Figure 1 and Table 4, the auditors' minimum predicted scale value in the low-quality context differs little from that in the high-quality and full range contexts (1.30 versus 1.00). This suggests that regardless of whether high-quality loans were presented to raters, their subjective minimum was cognitively anchored on the same type of low-risk loans. In contrast, as shown in Figure 1 and Table 4, the auditors' maximum predicted value of the range inferred in the high-quality context (3.42) was considerably smaller than that assumed for the full range and low-quality (5.00). In summary, the minimum subjective range values for the low-quality context is not far from the baseline diagonal (representing the full range of loans), whereas the maximum range value for the high-quality context is far from the diagonal. This suggests that while the lower end of the range (indicating a high-quality loan) is not easily shifted for auditors, the upper end of the range (indicating a low-quality loan) is more susceptible to context effects.

[10] This equation is similar to Equation 2 in Schepanski et al. (1992) where this expression is in terms of a five-point categorical scale from 1 to 5. See Equation (4) and footnote 2 for the transformation to a categorical scale and the derivation of the intercept and slope. By subscripting S_{MIN} and S_{MAX} we fix these parameters in one context and let them vary in the other context. To fit the data to the model we assigned "marker" (dummy) variables to indicate which context a specific rating belongs to. We used these marker variables to isolate different parameter values. In the equation below, which is an extension of Equation (5), marker variable "low" = 1 when in the low context and 0 when in the high context, and marker variable "high" = 1 in the high context and 0 in the low context. The resulting equation (derived from Equation (5)) that we actually use in the iterative regression is:
$C_{ik} = 1 + 4\,[w * (S_i - high*1 - low*S_{MINLC})/(high*S_{MAXHC} + low*5 - high*1 - low*S_{MINLC}) + (1 - w)(rank_{ik} - 1)(N_k - 1)]$,
where S_{MINLC} is a variable representing S_{MIN} when context is low-quality loans and S_{MAXHC} is a variable representing S_{MAX} when the context is high-quality loans.

[11] The student-participants showed slightly less agreement, but the correlation between their ratings and the scale values is still high within each context. These are r = 0.952 for the low-quality context and r = 0.961 for the high-quality context. This means that students have the ability and knowledge to use the relevant data to rate the loans on the criteria provided and to combine the data appropriately within each context.

FIGURE 1
Fit of the Range-Frequency Model to Participants' Ratings

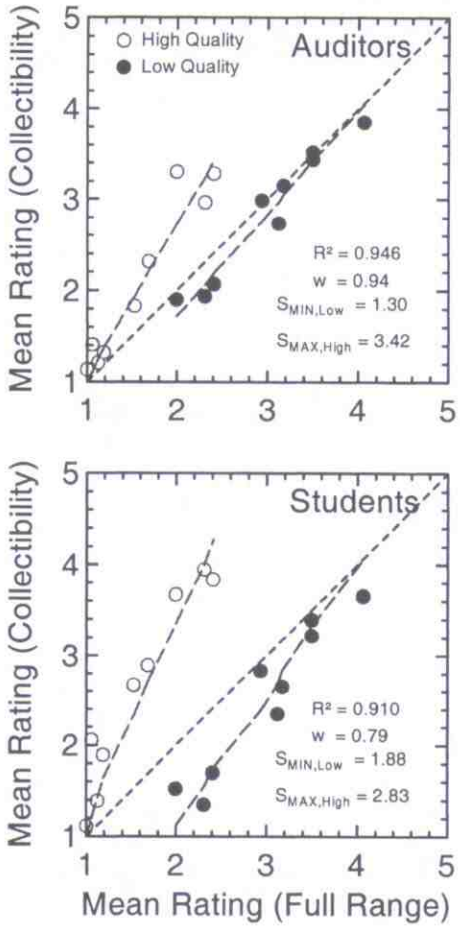

TABLE 4
Parameters for Range-Frequency Analysis

Sample	$S_{MIN,LC}$	$S_{MIN,HC}$	$S_{MAX,LC}$	$S_{MAX,HC}$	w	R^2
Auditors	1.30	1.00 F	5.00 F	3.42	.94	.946
Students	1.88	1.00 F	5.00 F	2.83	.79	.910

F = fixed parameter.

In contrast, the students' minimum predicted scale value in the low-quality context differs markedly from that in the high-quality and full range contexts (1.88 versus 1.00) for the students. This suggests that the students, lacking task experience, did not cognitively anchor their subjective minimum on the same type of high-quality loans, as did the professional auditors. Further, the students' maximum predicted value of the range inferred in the high-quality context (2.83) was even

lower than that of the auditors and like the auditors was smaller than that assumed for the full range and low-quality contexts (5.00). In summary, both the minimum and maximum subjective range values are far from the baseline diagonal for the students. This suggests that, unlike professional auditors who were experienced in the task, both the lower and upper ends of the range are very susceptible to context effects for the students.

In addition, students seemed to adjust to the range of cases presented in each context more strongly than did the auditors. In the high-quality condition the range-frequency fit indicates that students set the maximum, defining 5 on their scale, to 2.83, whereas auditors set it at 3.42. Thus, while both participant groups showed a strong tendency to set the lowest quality loan category (5) at a moderate risk level, students did so to a much greater extent. In the low-quality condition students set the minimum, defining a high-quality loan (1) on their scale to 1.88, whereas auditors set it at 1.30. Thus, the students showed a greater tendency to adjust their scale of judgment to the values presented. Thus, we conclude that students inexperienced in this task are more susceptible to context effects.[12]

Finally, the value of $w = .94$ (Equations (4) and (5)) for the auditors indicates that they primarily used the range principle to analyze the loans.[13] In contrast, $w = .79$ for the students indicating a greater weighting of frequency values .21 (1 – .79) versus .06 (1 – .94) for the auditors. It seems that students tended to use the rating categories equally using the frequency principle, whereas the greater frequency of previous rating cases tended to mute this for the auditors experienced with this task.

DISCUSSION

This study investigates whether auditors, when making a series of independent judgments, display biases in their later judgments that are systematically related to those made early in the series (the immediate context), and whether range-frequency theory explains the cognitive process that led to the observed judgment shifts. Our results indicate that auditors' judgments were significantly affected by the immediate context and they displayed contrast effects that are consistent with the predictions of range-frequency theory. We also find some evidence that task experience moderates the effects of immediate context. Auditors who were experienced with the rating task displayed significantly smaller contrast effects than did inexperienced students.

The loan-classification judgments and the cognitive process used by the participants are explained in large part by the range-frequency model, which had an $R^2 = .95$ for the auditors and an $R^2 = .91$ for the students. The results suggest that, in this setting, auditors were more susceptible to the range effect than to the frequency effect and that the largest effect was on those auditors who were exposed to an immediate context of high-quality loans. The auditors' individual work experiences may have led to an asymmetric extension of the subjective range across contexts. Specifically, auditors in the high-quality context may have reviewed relatively few loans during their careers that would warrant being assigned into the worst categories, leading them to be more reliant on the set of loans presented in the experiment for imagining the qualities that such a loan would possess. Conversely, auditors in the low-quality context would have observed many loans that would properly be classified as "current" and therefore, were better able to mentally consider the full range of categories though no "current" loans were presented to them in the experiment. The results for the students clearly indicate that they did not benefit from any task-specific experience as they shifted their subjective scale similarly in both contexts.

[12] We were unable to identify any differences among the auditors that were associated with differences in years of experience. This suggests that any protection from the effects of immediate context provided by task experience is obtained early in an auditor's experience and does not appreciably increase over time.

[13] The 95 percent confidence interval for w varied from 0.74 to 1.14, still indicating that auditors primarily used the range principle to analyze the loans.

Our results have audit implications. If these were real financial institutions, then several of the loans would likely have been over- or underrated, potentially leading to the bank's loan loss reserve being under- or overstated, respectively. This suggests the need by auditors to reduce their susceptibility to context effects. For example, to reduce range effects that may arise from skewed immediate contexts auditors should periodically review loans that exemplify the full range of the OCC's scale. Further, auditors should consider developing decision-making models (e.g., Wright and Willingham 1997) to aid auditors in achieving a greater level of consensus and consistency in assessing loan collectibility.

Our study addresses some basic questions concerning audit judgment and decision making such as the level of consensus in auditor judgment, the bias present in the judgment process, and the level of calibration in the judgment process (Bell and Wright 1995). We find that auditors can conclude very different outcomes, be subject to considerable bias, and exhibit poor calibration given different immediate contexts. We speculate that the contrast effect found in this study is not unique to the loan-classification task and that similar results would occur in other audit settings where an auditor is called upon to make a series of judgments such as establishing materiality, assessing inherent risk, evaluating internal control, evaluating analytical procedure results, evaluating results of testing, etc. (Bell and Wright 1995).

This study is subject to several limitations. First, to minimize the time requirements of the task, the participants had to make loan-classification judgments based solely on the information provided. In an actual loan-evaluation setting, auditors have a multitude of quantitative and qualitative information available to them. Second, participants made a series of loan-classification decisions over a short time interval. It is probable that in practice such decisions would be made over a longer period, which would likely dilute the effects of the immediate context. The external validity could be enhanced via future studies that consider alternative or additional information for loan-rating decisions. Third, the absence of an observable assimilation effect could be due to a lack of strength in the manipulation. Fourth, our results suggest that the range effect is primarily responsible for the contrast effect displayed by the participants, but this finding is tentative given that the experimental groups were given a restricted range. A future study to more clearly isolate and test the extent of the influence of the frequency principle would require an experimental design in which the range of loans presented was equal across conditions and only the preponderance of low-quality or high-quality loans being different. Finally, we conducted our experiment on one auditing task. Although we speculate that the contrast effects exhibited in this study would also occur in other audit settings, additional research is necessary to determine the generalizability of this study's findings.

APPENDIX

General Loan Information

The accompanying booklet includes limited financial information for a non-random sample of 9 commercial loans made by various banks, each with total assets less than $500 million.

You should assume that each loan was made by a *different* bank.

To make your judgment task more efficient, please assume the following information applies to each loan that you evaluate:

> The loan has an outstanding balance of $750,000. The borrower has additional debt from other lenders.
>
> The original term of the loan was 6 years. There are 4 years remaining on the loan.
>
> The collateral on the loan has *materially* depreciated in value over the past year.
>
> The terms of the loan require the borrower to make monthly payments to the bank to pay interest and principal on the loan.
>
> The loan is the first and only loan the bank has ever made to the borrower.
>
> The borrower is a manufacturer of small components that are used in a wide variety of products.
>
> All financial information provided is based on audited financial statements of the borrower.

The back of this page provides the AICPA's description of the five loan-classification categories that we ask you to use to classify each loan.

Loan-Classification Categories

The American Institute of Certified Public Accountants (AICPA) publishes an audit and accounting guide for Banks and Savings Institutions. The risk classification categories cited by the AICPA's audit guide (1996) are as follows:

1. **Current**—Normal acceptable banking risk
2. **Special Mention—**potential weaknesses that deserve management's close attention. If left uncorrected, these potential weaknesses may result in deterioration of the repayment prospects for the loan at some future date.
3. **Substandard**—inadequately protected by the current sound worth and paying capacity of the borrower or of the collateral pledged, if any. Loan has a well-defined weakness or weaknesses that jeopardize the liquidation of the debt.
4. **Doubtful**—possesses all the weaknesses inherent in those classified as substandard, with the added characteristic that the weaknesses make collection or liquidation in full, on the basis of currently existing facts, conditions, and values, highly improbable.
5. **Loss**—the loan is regarded as uncollectible and of such little value that its continuance as a bankable asset is not warranted.

In this experiment, you are asked to classify a series of loans using only a limited set of financial information for each loan. The following scale will be used for your classifications:

Current	**Special Mention**	**Substandard**	**Doubtful**	**Loss**

For each loan that you analyze, simply place an X in the box above the category in which you would place that loan.

Example of a Loan Description

Bayview Bank and Trust Co.

Borrower: Alpha Products **Payment Status:** > 120 days past due

Ratio	19x3	19x2	19x1
Current Assets / Current Liabilities	.95	1.16	1.31
Costs of Goods Sold / Average Inventory	2.0	1.9	1.9
Total Debt / Stockholders' Equity	2.55	2.14	1.64
Income From Operations / Net Sales	–.08	–.04	.01
Income From Operations / Stockholders' Equity	–.35	–.15	.02
Cash Flow From Operations / Interest Expense	–.88	–.03	1.18
Total Sales (in thousands)	$12,000	$11,429	$10,884

Based on the above information, I would classify the loan (as of 19x3) as:

Current	**Special Mention**	**Substandard**	**Doubtful**	**Loss**

REFERENCES

Abdel-khalik, A. R., and K. M. El-Sheshai. 1980. Information choice and utilization in an experiment on default prediction. *Journal of Accounting Research* 18 (Autumn): 325–342.

American Institute of Certified Public Accountants (AICPA). 1996. *AICPA Audit and Accounting Guide: Banks and Savings Institutions*. New York, NY: AICPA.

Anderson, N. H. 1966. Component ratings in impression formation. *Psychonomic Science* 6: 279–280.

Bell, T. B., and A. M. Wright. 1995. *Auditing Practice, Research, and Education*. New York, NY: AICPA.

Biernat, M., M. Manis, and T. E. Nelson. 1991. Stereotypes and standards of judgment. *Journal of Personality and Social Psychology* 60: 485–499.

Birnbaum, M. 1974. Using contextual effects to derive psycho-physical scales. *Perception and Psychophysics* 15: 89–96.

Bonner, S. E. 1990. Experience effect in auditing: The role of task-specific knowledge. *The Accounting Review* 65 (January): 72–92.

———, and B. L. Lewis. 1990. Determinants of auditor expertise. *Journal of Accounting Research* 28 (Supplement): 1–20.

Brewer, M. B. 1988. A dual process model of impression formation. In *Advances in Social Cognition*, edited by T. K. Srull, and R. S. Wyer, Jr., 1–36. Hillsdale, NJ: Erlbaum.

Dietrich, J. R., and R. S. Kaplan. 1982. Empirical analysis of the commercial loan classification decision. *The Accounting Review* (January): 18–38.

Frederick, D. M., V. B. Heiman-Hoffman, and R. Libby. 1994. The structure of auditor's knowledge of financial statement errors. *Auditing: A Journal of Practice & Theory* 13 (Spring): 1–21.

Heiman, V. B. 1990. Auditors' assessments of the likelihood of error explanations in analytical review. *The*

Accounting Review 65 (October): 875–890.

Herr, P. M. 1986. Consequences of priming: Judgment and behavior. *Journal of Personality and Social Psychology* 51: 1106–1115.

Higgins, E. T., W. S. Rholes, and C. R. Jones. 1977. Category accessibility and impression formation. *Journal of Experimental Social Psychology* 13: 141–154.

Hogarth, R., and H. Einhorn. 1992. Order effects in belief updating: The belief adjustment model. *Cognitive Psychology* 24: 1–55.

Jeffrey, C. 1992. The relation of judgment, personal involvement, and experience in the audit of bank loans. *The Accounting Review* 65 (October): 802–819.

Joyce, E. J., and G. C. Biddle. 1981. Anchoring and adjustment in probabilistic inference in auditing. *Journal of Accounting Research* 19 (Spring): 120–145.

Kinney, W. R., and W. C. Uecker. 1982. Mitigating the consequences of anchoring in auditor adjustments. *The Accounting Review* 57 (January): 55–69.

Libby, R., and D. M. Frederick. 1990. Experience and the ability to explain audit findings. *Journal of Accounting Research* 28 (Autumn): 348–367.

Manis, M., and J. R. Paskewitz. 1984. Judging psychopathology: Expectation and contrast. *Journal of Experimental Social Psychology* 20: 363–381.

Mellers, B., and M. Birnbaum. 1983. Contextual effects in social judgment. *Journal of Experimental Social Psychology* 19: 157–171.

Messier, W. F., and R. M. Tubbs. 1994. Recency effects in belief revision: The impact of audit experience and the review process. *Auditing: A Journal of Practice & Theory* 13 (Spring): 57–72.

Parducci, A. 1965. Category judgment: A range-frequency model. *Psychological Review* 72: 407–418.

———, and L. F. Perrett. 1971. Category rating scales: Effects of relative spacing and frequency. *Journal of Experimental Psychology Monograph* 89: 427–452.

———. 1995. *Happiness, Pleasure and Judgment: The Contextual Theory and its Applications*. Mahwah, NJ: Erlbaum.

Pepitone, A., and M. DiNubile. 1976. Contrast effects in judgments of crime severity and the punishment of criminal violators. *Journal of Personality and Social Psychology* 33: 448–459.

Schepanski, A., R. M. Tubbs, and R. A. Grimlund. 1992. Issues of concern regarding within- and between-subjects designs in behavioral accounting research. *Journal of Accounting Literature* 11: 121–150.

Srull, T. K., and S. R. Wyer. 1980. Category accessibility and social perception: Some implications for the study of person memory and interpersonal judgments. *Journal of Personality and Social Psychology* 38: 841–856.

Tuttle, B. 1996. Using base rate frequency perceptions to diagnose financial statement error causes. *Auditing: A Journal of Practice & Theory* 15 (Spring): 104–121.

Tversky, A., and D. Kahneman. 1974. Judgment under uncertainty: Heuristics and biases. *Science* 185: 1124–1131.

Wedell, D. H., A. Parducci, and R. Geiselman. 1987. A formal analysis of ratings of physical attractiveness: Successive contrast and simultaneous assimilation. *Journal of Experimental Social Psychology* 23: 230–249.

———, A. Parducci, and M. Lane. 1990. Reducing the dependence of clinical judgment on the immediate context: Effects of number of categories and type of anchors. *Journal of Personality and Social Psychology* 58(2): 319–329.

———. 1994. Contextual contrast in evaluative judgments: Test of pre- versus post-integration models of contrast. *Journal of Personality and Social Psychology* 66: 1007–1019.

Wright, W. F., and J. J. Willingham. 1997. A computational model of loan loss judgments. *Auditing: A Journal of Practice & Theory* 16 (Spring): 99–113.

———. 2001. Task experience as a predictor of superior loan loss judgments. *Auditing: A Journal of Practice & Theory* 20 (March): 147–155