



The effect of audit seniors' decisions on working paper documentation and on partners' decisions

David N. Ricchiute *

College of Business Administration, University of Notre Dame, Notre Dame, IN 46556-0399, USA

Abstract

Prior research reports that the memories of working paper preparers may be biased toward evidence consistent with their prior decisions, but that reviewers exposed to the *same* set of evidence can mitigate the bias by evaluating inconsistent evidence. This study tests whether audit seniors' decisions bias their ability to recognize evidence to document in working papers, and whether biased documentation affects the decisions of audit partners who are exposed only to the *subset* of evidence that seniors recognize and document. Experiment 1 confirms that audit seniors' prior decisions bias their memories for evidence to document in working papers, and also creates materials for experiment 2. Experiment 2 offers a new insight: when exposed only to the evidence that seniors recognize and document, partners make decisions biased in the direction of the seniors' decisions, since not all of the inconsistent evidence is documented. Experimental procedures controlled for four alternative interpretations: justification, evidence order, recency, and primacy. © 1999 Elsevier Science Ltd. All rights reserved.

1. Introduction

Prior research reports that auditors' memories are biased (e.g. Moeckel & Plumlee, 1989), and that the memory performance of reviewers exposed to the same set of evidence is superior for evidence that is inconsistent with preparers' decisions (Libby & Trotman, 1993). This study extends Libby and Trotman by focusing on audit decision making when the reviewer evaluates a *subset* of the evidence that a preparer evaluated: I test whether audit seniors' decisions affect their memory for the evidence they will document in working

papers, and whether the seniors' documentation affects audit partners' decisions. The results lead to two findings. First, audit seniors' decisions interact with pre-decision evidence in memory to bias their post-decision recognition of the evidence they will document in working papers. Second, but more important to audit effectiveness, the evidence that seniors recognize and document in working papers biases audit partners' decisions in the direction of the seniors' decisions.

Post-decision recognition bias is important because some public accounting firms document working paper evidence after a decision is reached by the staff member initially responsible for the decision, but *before* the decision is reviewed by the staff member finally responsible for the decision

* Fax: +1-219-631-5255; e-mail: ricchiute.1@nd.edu

(Cushing & Loebbecke, 1986). For example, the firm participating in this study empowers an audit senior to document evidence that supports his or her going-concern decision. The engagement partner's review of the senior's decision is based on a subset of the evidence the senior evaluated: the evidence documented in the working papers. The full set of evidence the senior evaluated is not fully available to the partner. The link among a senior's decision, the senior's memory for the evidence evaluated to reach the decision, and the evidence the senior will document in the working papers has implications for partner decision making. At the extreme, a partner may reach a different decision with the full set of evidence.

Post-decision recognition bias I expect because prior research in auditing (Libby & Trotman, 1993; Tan, 1995) and in psychology (Pennington & Hastie, 1988) reports that informed decision makers' post-decision recognition behavior is superior for facts consistent with their decisions. These results offer reasonable expectations for auditor behavior in this study. For example, the experimental settings and the experience levels of subjects participating in Libby and Trotman, in Tan, and in this study are similar, and both the going-concern decision task and the verdict decision task rely on evidence that has dependent probative implications for a decision.

One hundred subjects from a major public accounting firm participated in the experiments. Forty audit seniors participated in experiment 1, and 60 audit partners participated in experiment 2. Experiment 1 establishes that audit seniors are more likely to recognize and document sentences that are consistent with their prior substantial-doubt decisions. The results confirm Libby and Trotman (1993) in a recognition memory task, and are important because they set up, and provide experimental materials for, experiment 2. Experiment 2 extends Libby and Trotman: Libby and Trotman found that, when exposed to the *same* set of evidence that preparers evaluated, reviewers can mitigate the effects of memory bias by evaluating inconsistent evidence. Experiment 2 finds that, when exposed only to the *subset* of evidence that seniors document, reviewers' decisions are biased in the direction of seniors' decisions.

The results of the study have important implications for working paper preparation in practice: A staff member initially responsible for an audit decision task should document in the audit working papers (or make available for review) all evidence relevant to the decision, thereby allowing a reviewer to evaluate evidence that is inconsistent with the preparer's decision.

2. Motivation

Prior research has lent insight into the review of completed working papers. For example, research reveals that the review process reduces judgment variance (Trotman & Yetton, 1985), that managers (seniors) detect conceptual (mechanical) errors more accurately than seniors (managers) (Ramsey, 1994), and that reviewers who are familiar with preparers detect more errors in audit conclusions (Asare & McDaniel, 1996). However, little is known about partner decision making when working papers document a subset of the evidence that preparers evaluated. As Libby and Trotman (1993, p. 563) point out, "if initial decision makers recall more consistent than inconsistent items and then use these items in the workpapers to support their decision, the reviewer will receive a biased set of facts". The purpose of this study is to test whether audit seniors' decisions bias what they will remember and document in working papers, and whether the seniors' biased documentation biases partners' decisions.

2.1. Audit senior decision making and memory

Audit seniors gather and evaluate large volumes of evidence to reach audit decisions, and document decision-relevant evidence in working papers. However, seniors sometimes document evidence in working papers *after* an audit decision is reached, not when the evidence is gathered and evaluated. For example, in the participating firm, audit seniors gather and evaluate evidence relevant to the going-concern decision, and then document evidence supporting their decision *before* the working papers are reviewed by the

engagement partner.¹ The completeness of the working papers an engagement partner reviews depends in part on audit seniors' memories for the evidence they gather and evaluate.

Prior research in auditing offers a reasonable expectation that audit seniors' memories may be inaccurate. For example, Moeckel and Plumlee (1989) found that, following delay, auditors are as confident in their incomplete and inaccurate recognitions of working paper evidence as they are in their accurate recognitions. However, more important to the expectations in this study, prior research both in auditing and in psychology reports that a decision maker's post-decision memory is superior for facts consistent with their decisions. In auditing, Libby and Trotman (1993) and Tan (1995) addressed initial decision makers' memories for evidence consistent and inconsistent with their audit decisions. Libby and Trotman found (experiment 1) that initial decision makers' relative recall was in the direction of evidence consistent with their prior decisions.² And Tan (1995) found (experiment 1) that, compared to auditors newly assigned to an engagement, auditors previously assigned to an engagement recalled more evidence consistent with audit decisions. In psychology, Pennington and Hastie (1988) found (experiment 1) that subjects in a juror decision task were more likely to correctly recognize evidence sentences that they had seen (targets) and that were consistent with their prior verdict decisions. More revealing, the subjects were more

likely to incorrectly recognize plausible sentences that they had not seen (lures) but that were consistent with their prior verdict decisions.

Because no single theory has emerged as the optimal predictor of auditor recognition behavior, Libby and Trotman (1993), Tan (1995), and Pennington and Hastie (1988) were relied upon to motivate reasonable expectations for audit seniors' recognition bias for three reasons. First, the experimental settings in Libby and Trotman and in Tan are similar to the setting in this study: financial distress, financial viability, and substantial-doubt decisions, respectively. Second, the experience levels of subjects in Libby and Trotman (experiment 1: mean = 3.95 years) and the rank of subjects in Tan (experiment 1: audit seniors) are consistent with the experience levels and ranks of subjects participating in this study (experiment 1: audit seniors). Third, Pennington and Hastie's materials (evidence sentences) and decision task (*yes* there is reasonable doubt vs *no* there is not) map well to an audit seniors' going-concern decision task (*yes* there is substantial doubt vs *no* there is not). As a result, I expect that audit seniors are more likely to correctly (incorrectly) recognize target (lure) sentences that were (were not) presented as evidence and (but) that are consistent with their substantial-doubt decisions. Experiment 1 tests this expectation.

The auditing literature offers two related perspectives which, like recognition bias, can drive working papers biased toward a preparer's decision but which, unlike recognition bias, are conscious intentional behaviors: justification and persuasion. The decision to document evidence in working papers may be motivated by *justification* (e.g. Koonce, Anderson, & Marchant, 1995; Tan, 1995), an accountable senior's attempt to justify his or her decision with confirming documentation. Consistent with justification, Rich, Solomon, & Trotman (1997) characterize the working paper preparation and review process from a *persuasion* perspective,³ arguing that preparers may selectively remember, incompletely document, or frame the content of documented evidence to persuade—and thereby enhance their reputation with—reviewers. To rule out the alternative interpretation that the results of the study may be driven by justification

¹ The participating firm's field reference manual reads in part as follows: "For some audit decisions, it is advisable for the in-charge auditor to evaluate the relevant audit evidence before the evidence and the related audit decision are documented in the audit workpapers. Examples of these types of decisions include ... substantial doubt about the ability of an entity to continue as a going concern".

² Libby and Trotman (1993) also found (experiment 1) that reviewers recalled more evidence that was inconsistent with initial decision makers' judgments.

³ Rich et al. (1997, p. 485) acknowledge the consistency between justification and the persuasion perspective: "The notion that there are incentives for preparers to act strategically is consistent with research viewing the review process as an accountability-inducing mechanism."

(or persuasion), experiment 1 manipulates justification between subjects.

2.2. Partner decision making

The effects of audit seniors' recognition memories on audit partners' decision effectiveness depends largely on two issues: partner review strategies, and partner awareness of evidence that is inconsistent with audit seniors' decisions. The first, *review strategies*, has been addressed in prior audit research on memory. Libby and Trotman (1993) found (experiment 2) that auditors acting as reviewers (mean experience = 4.2 years) recalled more items that were inconsistent with preparers' financial-distress decisions, and that the reviewers' recall and subsequent decisions were positively and significantly associated. This result suggests that the review process may act as a control over the biased memories of working paper preparers.

⁴ Both *recall* and *recognition* are germane to this study. In experiment 1 (task one), the subjects' substantial-doubt decisions require that they rely on recall, since the probative value of each newly revealed evidence sentence depends on the recalled content of sentences previously seen. However, the purpose of the experiment is not to measure recall. Rather, the purpose is to classify the subjects into substantial-doubt decision categories (*yes* vs *no*). Recognition is germane to task three (recognition memory test) for three reasons. First, like Moeckel and Plumlee (1989, p. 655), my purpose is to "probe the contents of the subjects' memories most directly, without the confounding by retrieval strategies that would occur using a recall test". Second, a national office partner who helped develop the experimental materials advised (and I agreed) that, although recall is relevant, recognition is used widely in complex tasks for which the evidence is gathered, seen, and is then available for reinspection. Third, Dellarosa and Bourne Jr. (1984) predicted and found (experiment 1) that recognition is *less* likely than recall to produce an effect in tasks related to text, and therefore provides a stronger test of the expectations in this research. The applicability of recall and recognition in applied experimental research is not without controversy (for example: in auditing, see Choo & Trotman, 1991, note 9; in marketing, see Singh, Rothschild & Churchill, 1988). See Haist, Shinamura, & Squire (1992) for a discussion, and experiments, on the relationship between recall and recognition memory.

⁵ There is no significant difference in the results of experiment 1 when the eight subjects who had *not* made and documented a substantial doubt decision are deleted from the analysis.

However, as Libby and Trotman (1993) point out, the generalizeability of their experiment 2 is limited because partners are *not aware of inconsistent evidence* in all audit decision tasks. For example, if the expectation for experiment 1 in this study is supported (i.e. if seniors are more likely to recognize evidence that is consistent with their prior decisions), then the assumption that the full set of evidence will be available to partners does not necessarily generalize to going-concern decisions in the participating firm. Rather, audit seniors' recognition bias would systematically reduce the amount of inconsistent evidence available to the partner for review. As a result, I expect that audit partners are more likely to decide *yes* (*no*) when exposed only to the subset of evidence sentences that the seniors who decided *yes* (*no*) in experiment 1 recognize and document. Experiment 2 tests this expectation.

3. Experiment 1: audit seniors' decision making and memory

Experiment 1 tests whether audit seniors' recognition⁴ memories are affected by their prior audit decisions. Experiment 1 also collects data (task three: recognition memory test; task four: documentation) crucial to experiment 2, which provides evidence on the study's most important contribution to audit effectiveness: if audit seniors' prior decisions bias their memories for evidence to document in working papers, does the biased documentation affect audit partners' decisions?

3.1. Experiment 1

Subjects for experiment 1 were 40 audit seniors attending a continuing education seminar sponsored by the participating firm; 32 had made and documented a substantial-doubt decision in practice.⁵ The seniors ranged in experience from 32 to 43 months (mean = 37 months).

3.2. Materials

The experimental materials for task one consisted of evidence sentences adapted from Ric-

chiute (1992) but edited by a national office partner, a manager, and the author to the same number of syllables and to the active voice. Editing equalized the form of each sentence, thereby balancing the memorability of the sentences that might be attributed to form only. The sentences involve a substantial-doubt decision and are appropriate to this study for two reasons. First, the substantial-doubt decision is typically made in practice from voluminous working paper evidence that requires the use of memory, and is sufficiently unstructured to exploit differences in the behavior of more experienced (partners) and less experienced (seniors) practitioners (Abdolmohammadi & Wright, 1987). Second, no one sentence was judged by the national office partner and manager to be so strongly persuasive that it might alone drive premature decisions (for example, a sentence that announced bankruptcy proceedings). Premature decisions could have detracted the subjects' attention from the remaining sentences, potentially contaminating the recognition memory test, the third task in experiment 1.

Fifteen of the 60 sentences refer to events consistent with the *yes* decision, for example, "The company will report a net loss for the second year in a row." Fifteen refer to events consistent with the *no* decision, for example, "Management signed a long term contract with a new domestic customer." Fifteen refer to events consistent with *both* decisions, for example, "The company contemplates issuing new convertible bonds next

year."⁶ And fifteen refer to events consistent with *neither* decision, for example, "Independence has not been impaired by any audit staff assignments." Appendix A lists all 60 sentences by decision membership: *yes*, *no*, *both*, and *neither*. The national office partner, manager, and two audit practice partners independently judged, and agreed to, the decision membership of each sentence. The firm representatives who helped edit and judge the evidence sentences did not participate in the experiments.

3.3. Procedure

Experiment 1 was held in a computer-equipped training room. The author was present at the experimental site, but not in the training room. Rather, a national office partner greeted the subjects, invited them to sit at individual work stations, and introduced the experiment. The partner explained that the experiment involved working paper documentation and that the firm would use the results of the experiment to help reconsider firm policy about working paper documentation standards.

The partner told the subjects that they would view evidence sentences, one at a time, about a wire and cable company and then make a binary (yes/no) substantial-doubt decision.⁷ The subjects worked individually, advancing to the next sentence by depressing the keyboard spacebar. Following display, an evidence sentence could not be retrieved.

3.4. Tasks

After reading all 60 sentences, each subject performed the first task: indicate *yes*, there is, or *no*, there is not, substantial doubt about the company's ability to continue as a going concern, the task imposed on U.S. auditors by SAS No. 59 (Anon, 1988). The decisions—*yes* or *no*—were used to determine whether the subjects' recognition memories, measured in the third task, were affected by the substantial-doubt decisions they had made in the first task. The sentences were ordered randomly for each subject to control for the effects of order, recency, and primacy; no two subjects read the sentences in the same order.

⁶ The sentence refers both to the *yes* and to the *no* decisions in this sense: to the *yes* decision because offering bonds results in additional debt (and, therefore, additional demands on cash for interest and principal payments); and to the *no* decision because the bond proceeds provide a means to finance existing liabilities. In short, the sentences that refer to both decisions carry bad news (the *yes* decision) and good news (the *no* decision).

⁷ Although experimental questions posed as probability distributions (or multiple-point Likert scales) are common in research on auditors' going-concern decisions (e.g., Choo & Trotman, 1991), the participating firm argued strongly that anything other than a binary *yes/no* decision would distort task realism since, in practice, the firm demands that audit seniors decide whether there is, or is not, substantial doubt and then document the decision in the working papers. At no time in practice does the firm cast the decision into a probability distribution.

To clear short-term memory, the second task, a distractor, asked that the subjects respond to demographic questions (e.g. years of experience, experience with substantial-doubt decisions, experience within rank and in the profession, education, degrees, certification). The length of the distractor task was determined from a pilot study. Twenty audit seniors participated in a pilot study that tested two alternative delay schemes for the time lag between the administrations of task one (substantial-doubt decision) and task three (recognition memory test): *same day* administration (time lag = 20 min distractor-task delay) vs *next day* administration (time lag = 24 h delay). Ten seniors were assigned at random to the *same day* condition, and 10 to the *next day* condition.

Same day administration was chosen for three reasons. First, there was no significant difference in recognition behavior across treatment groups. Second, *same day* administration rules out the suspicion that subjects may have contaminated the results of the recognition memory test by discussing the experiment between the administrations of tasks one and three (24 h delay). And third, the *same day* administration would be less intrusive on the time demands of the participating seniors. At this point in the experiment, the subjects did not know that the next task was a memory test.

In the recognition memory test, the third task, 60 individual sentences were displayed on the subjects' computer screens one at a time: 32 (about half) were sentences that had been displayed among the original 60 (target sentences) and 28 were lures (seven for each of the four decision memberships: *yes*, *no*, *both*, and *neither*). Lure sentences were used to cross validate the subjects' recognition behavior on target sentences, thereby offering a stronger test of recognition bias for working paper evidence. That is, if the subjects correctly recognize more targets *and* incorrectly recognize more lures referring to their decisions, then their behavior on lures would cross validate their behavior on targets, thereby allowing for a stronger claim that the subjects' prior substantial-doubt decisions are not likely independent of their recognition behavior.

The lure sentences were drafted by a national office partner, a manager, and the author; were the

same number of syllables as the targets; and are listed in Appendix B. Rather than rearrangements of the original sentences (e.g. "The company incorporated forty five years ago." vs "Forty five years ago the company incorporated."), the lures were plausible case-relevant sentences that altered the meaning of the original sentences. For example, "Long term commercial bank loan covenants restrict some fixed asset disposals" (Appendix A, sentence number 10), an original evidence sentence consistent with the *yes* decision, was replaced in task three with the lure, "Long term commercial bank loan covenants on assets have been violated," also consistent with the *yes* decision. In this case, the original target sentence indicates that *covenants restrict* some disposals, whereas the lure indicates that *covenants have been violated*: two different meanings.

The subjects were instructed on their computer screens that they would view evidence sentences and that, for each displayed sentence, the task was to indicate whether they had seen the sentence among the original 60, not whether the sentence was plausible. Specifically, the subjects were told to respond to each sentence by pressing the key *I* if they had seen the sentence and *0* if not.

For each sentence they recognized, the subjects completed a fourth, two-question, task: (a) indicate whether they would document the evidence in the working papers (*I* = yes, *2* = no), and (b) if they would document the sentence, rate the importance of the sentence as working paper evidence (a 25 point response scale anchored by "Important evidence *against* substantial doubt" and "Important evidence *for* substantial doubt"; the midpoint was "Not important.")

Interpretation of the results of the fourth task is threatened by the alternative explanation that the results may be driven by justification (or persuasion). To rule out this interpretation, justification was manipulated similar to Koonce et al. (1995, p. 373–374): 20 of the seniors (mean experience = 37 months) were assigned to group 1 (the *review* condition), and 20 (mean experience = 38 months) were assigned to group 2 (the *no review* condition). However, because the objective was to determine whether the experimental results were necessarily driven by justification, rather than to lend insight into subordinate justification behavior, I did not manipulate

variables that are common to justification studies but that require significant numbers of participants to manipulate between subjects. For example, like Koonce et al., review condition was manipulated between subjects (review vs no review), but unlike Koonce et al., the type of evidence that the subjects reviewed was not manipulated. In the *review* condition, the 20 subjects were instructed on their computer screens prior to task one that the national office would review their responses. The *review* condition subjects were asked to reveal their names, telephone numbers, and office and e-mail addresses (all 20 subjects replied). About five weeks after the administration of experiment 1, the national office communicated with the seniors about their responses and the purpose (and results) of the study. No instructions about review were given to the *no review* condition subjects.

In summary, the subjects in experiment 1 performed four tasks: (1) decide substantial doubt (*yes/no*) based on 60 randomly ordered evidence sentences, (2) provide demographic data (distractor task), (3) for each of the target and lure sentences, indicate whether they had seen the sentence among the original sentences, and (4) indicate whether to document the recognized sentence in the working papers, and rate the importance of the documented sentence as working paper evidence.

3.5. Results: experiment 1

3.5.1. Recognition

Table 1, Panel A, reports the mean percentage of correct recognitions (i.e. the percentage of correctly recognized target sentences in each decision category) and the mean percentage of incorrect recognitions (i.e. the percentage of incorrectly recognized lure sentences) by review condition.⁸ For example, in Panel A the 24 seniors (13 *review* condition seniors + 11 *no review* seniors) who decided *yes* there is substantial doubt correctly

recognized the 16 target sentences that refer to the *yes* decision (8 *yes* targets + 8 *both* targets) in 86.4% of the trials, but incorrectly recognized the 14 lure sentences referring to the *yes* decision (7 *yes* lures + 7 *both* lures) in 26.3% of the trials. Although more seniors judged *yes* (24) there is substantial doubt than *no* (16), the pattern of *yes* and *no* decisions is not reliably different from 20 (50%) *yes* and 20 (50%) *no* ($X^2(1, n=40)=1.6$, $p>0.20$), thereby suggesting that the persuasiveness of the *yes* and *no* sentences was roughly equivalent.

The direction of the mean percentages of *correct recognitions* in Table 1, Panel A (left column), confirms that, as expected, the higher percentages of correct recognitions are associated with the seniors' decisions. That is, the 24 seniors who decided *yes* there is substantial doubt correctly recognized 86.4% of the target sentences referring to evidence in the *yes* decision, which is more than the 74.2% correctly recognized target sentences referring to the *no* decision. Consistent with this finding, the reverse is true for the 16 seniors who decided *no*. The *incorrect recognitions* for lure sentences reported in Table 1, Panel A (right column), provide further evidence that the seniors' recognition behavior is consistent with the conclusion that audit seniors' substantial-doubt decisions affect their memory for the evidence they will document in working papers. Observe in Table 1, Panel A (right column), that, as expected, the seniors' percentage of incorrect recognitions for lure sentences is higher for the evidence sentences corresponding to their substantial-doubt decisions. In addition, the *review* condition and the *no review* condition subjects' pattern of correct and incorrect recognitions is similar within substantial-doubt decision categories. For example, the 13 *review* condition subjects and the 11 *no review* subjects who decided *yes*, correctly recognized 86.2% and 86.6%, respectively, of the target sentences referring to the *yes* decision.

Table 1, Panel B, reports a multivariate analysis of variance (*MANOVA*) that includes two independent variables, the seniors' *yes/no* substantial doubt decisions and review condition, and four dependent variables: the percentage of correctly recognized target sentences referring to the *yes*

⁸ Subjects' response times in the recognition memory test were measured in milliseconds per syllable as a proxy for the cognitive effort expended on targets and on lures. The overall mean response times for targets and for lures are not significantly different ($p>0.20$), suggesting that the subjects were no more taxed by lures than by targets.

Table 1

Mean percentage of correct/incorrect recognitions and MANOVA. Experiment 1: audit seniors

| Panel A: Mean percentage (standard deviation) of correct/incorrect recognitions (rounded) | | Correct recognitions | | Incorrect recognitions | |
|---|-------------------------|-----------------------|-------------|------------------------|-------------|
| | | Targets referring to: | | Lures referring to: | |
| Substantial Doubt Decisions | <i>n</i> | Yes Decision | No Decision | Yes Decision | No Decision |
| <i>Yes</i> | | | | | |
| Review condition | 13 | 86.2 (5.3) | 74.4 (6.7) | 26.7 (3.3) | 13.9 (2.3) |
| No review condition | 11 | 86.6 (5.0) | 73.9 (3.9) | 25.7 (2.4) | 13.6 (2.9) |
| | 24 | 86.4 (5.1) | 74.2 (5.5) | 26.3 (2.9) | 13.8 (2.6) |
| <i>No</i> | | | | | |
| Review condition | 7 | 73.5 (6.9) | 83.0 (3.7) | 14.9 (3.0) | 26.9 (3.0) |
| No review condition | 9 | 73.3 (5.3) | 84.9 (6.2) | 14.7 (3.3) | 27.2 (2.8) |
| | 16 | 73.4 (5.8) | 84.1 (5.2) | 14.8 (3.1) | 27.1 (2.9) |
| Panel B: MANOVA | | | | | |
| Independent variable | Dependent variables | SS | F | p | |
| Substantial doubt decision | Correct recognitions: | | | | |
| | Targets referring to | | | | |
| | Yes decision | 910.442 | 30.304 | 0.001 | |
| | No decision | 1589.338 | 51.880 | 0.001 | |
| Review condition | Incorrect recognitions: | | | | |
| | Lures referring to | | | | |
| | Yes decision | 1696.240 | 215.238 | 0.001 | |
| | No decision | 1122.212 | 113.998 | 0.001 | |
| Decision x Review condition | Correct recognitions: | | | | |
| | Targets referring to | | | | |
| | Yes decision | 4.658 | 0.155 | 0.696 | |
| | No decision | 2.921 | 0.095 | 0.759 | |
| | Incorrect recognitions: | | | | |
| | Lures referring to | | | | |
| | Yes decision | 12.221 | 1.551 | 0.221 | |
| | No decision | 25.203 | 2.560 | 0.118 | |
| | Correct recognitions: | | | | |
| | Targets referring to: | | | | |
| | Yes decision | 13.171 | 0.438 | 0.512 | |
| | No decision | 1.349 | 0.044 | 0.835 | |
| | Incorrect recognitions: | | | | |
| | Lures referring to | | | | |
| | Yes decision | 20.303 | 2.576 | 0.117 | |
| | No decision | 4.520 | 0.459 | 0.502 | |

decision, the percentage of correctly recognized target sentences referring to the *no* decision, the percentage of incorrectly recognized lure sentences referring to the *yes* decision, and the percentage of incorrectly recognized lure sentences referring to the *no* decision. For the independent variable *substantial doubt decision*, all four dependent variables are significant ($p < 0.001$); Roy's criterion, the

proportion of variance accounted for by the independent variable, is 0.91; and Hotelling's trace criterion (10.748), a multivariate test of significance, is significant ($p < 0.001$). However, for the independent variable *review condition*, none of the four dependent variables are significant ($p > 0.118$), Roy's criterion is 0.11, and Hotelling's trace criterion is insignificant (0.122, $p > 0.416$),

suggesting that the recognition results were not likely driven by the seniors' attempts to justify their substantial doubt decisions. Panel B also reveals that the interaction between substantial-doubt decision and review condition is not significant for any of the four dependent variables ($p > 0.117$).

A 2 (yes/no Substantial Doubt decision) \times 2 (Target/Lure sentences) \times 2 (yes/no Decision Membership of evidence sentences) ANOVA was used to test for an interaction between the seniors' pooled decisions (i.e. review condition + no review condition) and the decision membership (yes, no, both, neither) of the evidence sentences they recognized. The dependent variable is the mean percentage of correct/incorrect recognitions. The ANOVA reveals a significant Substantial Doubt \times Decision Membership interaction ($F = 3.96$, $MS_e = 5784.02$, $p < 0.05$), indicating that the seniors' recognition behavior is not independent of their substantial-doubt decisions. This interaction, the results of the MANOVA (Table 1, Panel B), and the direction of the mean percentage of correct/incorrect recognitions (Table 1, Panel A) indicate that the seniors are significantly more likely to correctly recognize target sentences or to incorrectly recognize lures that are consistent with their substantial-doubt decisions. These results suggest that the seniors' recognition memories are not likely independent of their prior decisions.

3.5.2. Documentation and importance ratings

The 24 seniors who decided yes, there is substantial doubt, all recognized (task three) and documented (task four) 44 of the 60 evidence sentences, and the 16 seniors who decided no all recognized and documented 41 of the evidence sentences.⁹ The sentences documented by the seniors who decided yes and no comprise the experimental materials for experiment 2, groups 2 and 3 (and are discussed later and reported at Appendix C).

Consistent with the relatively high percentage of sentences documented across decision categories (yes: $44/60 = 73\%$; no: $41/60 = 68\%$), the seniors' documentation preferences did not differ significantly for 54 of the 60 of the evidence sentences (90%). Table 2, Panel A, reports the six sentences (60–54) for which the seniors who decided yes, and the seniors who decided no, differed significantly ($p \leq 0.05$) about whether to document the sentence. Two of the six sentences (numbers 5 and 10) refer to events consistent with the yes decision (Appendix A), and in Panel A were documented significantly more often by the seniors who decided yes, all of whom documented the two sentences. Three of the six sentences (17, 18 and 24) refer to events consistent with the no decision (Appendix A), and in Panel A were documented significantly more often by the seniors who decided no, all of whom documented the three sentences. Neither review condition ($p > 0.108$) nor the Substantial Doubt \times Review Condition interaction ($p > 0.101$) was significant for any of the 60 evidence sentences. The results indicate that, apart from six sentences, the seniors' substantial-doubt decisions did not weigh significantly on their decisions to document individual evidence sentences. However, the seniors' substantial doubt decisions did weigh more on which sentences were recognized by the 24 seniors who decided yes and by the 16 seniors who decided no. For example, the seniors who decided yes all documented 15 of the yes sentences and 10 of the no sentences, and the seniors who decided no all documented 14 of the no sentences and 9 of the yes sentences (see Appendix C). Table 2, Panel B, reports mean importance ratings for evidence sentences corresponding to the yes and to the no decisions, categorized by the seniors' decisions, and the sentences rated most important. Clearly, the recognized evidence sentences corresponding to the seniors' substantial-doubt decisions received much higher ratings (19.92 for seniors deciding yes, 20.50 for seniors deciding no) than did evidence sentences corresponding to the opposing decision (3.91 and 3.66, respectively). A 2 (yes/no Substantial Doubt decision) \times 2 (yes/no Decision Membership of evidence sentences) ANOVA reveals a significant interaction ($p < 0.001$). Together, the pattern of results for documentation and

⁹None of the 16 (19) sentences that were not recognized and documented by all 24 (16) of the seniors who decided yes (no) were rated on average more important by the seniors who did recognize and document them.

Table 2

Sentences documented significantly more often by seniors who decided “yes” and “no” and importance ratings. Experimental 1: audit seniors

Panel A: Sentences documented significantly more often by seniors who decided “Yes” and “No”

| Sentence | Sentence documented significantly more often by seniors who decided | | <i>F</i> | <i>p</i> |
|--|---|----|----------|----------|
| | Yes | No | | |
| 5 Management signed an uneconomic purchase commitment for copper. | x | | 4.715 | 0.037 |
| 10 Long term commercial bank loan covenants restrict some fixed asset disposals. | x | | 4.715 | 0.037 |
| 17 Total assets have not changed significantly over the prior year. | | x | 4.819 | 0.035 |
| 18 The debt to equity ratio is down compared to the prior year | | x | 6.292 | 0.017 |
| 24 Management signed a long term contract with a new domestic customer. | | x | 7.961 | 0.008 |
| 39 The company has not fully insured the replacement cost of assets. | x | | 7.379 | 0.010 |

F statistic insignificant ($p > 0.05$) for all other evidence sentences.

Panel B: Mean importance ratings

| Substantial doubt Decision | Evidence sentences referring to | |
|----------------------------|---------------------------------|-------------|
| | Yes Decision | No Decision |
| Yes $n = 24^a$ | 19.92 | 3.91 |
| No $n = 16^b$ | 3.66 | 20.50 |

^a Seniors who decided yes rated sentence numbers 1, 7, 12, 13, 14, 16, 19, 20, 29, and 50 most important for the working papers.

^b Seniors who decided no rated sentence numbers 17, 18, 24, 28, 33, 34, 37, 42, 44, and 45 most important for the working papers.

for importance ratings indicates that, although the seniors' substantial-doubt decisions did not drive significantly different working paper documentation preferences on the vast majority of evidence sentences, the seniors found significantly more important to the working papers those documented sentences that were consistent with their decisions.

4. Experiment 2: partner decision making

Experiment 2 addresses the key question left unresolved by the audit seniors' recognition bias documented in experiment 1: Does audit seniors' recognition bias affect audit partners' decisions? More specifically, would audit partners exposed to the *full set* of evidence sentences make different sub-

stantial-doubt decisions than partners exposed only to the *subset* of sentences that the seniors in experiment 1 recognized (task three) and reported they would document in the working papers (task four)?

4.1. Subjects

Subjects for experiment 2 were 60 audit partners attending an audit technology meeting sponsored by the participating firm's national office. The subjects ranged in experience from 14 to 20 years (mean = 15 years); all 60 had made and documented substantial-doubt decisions.

4.2. Materials

The experimental materials varied by treatment group, and are summarized in Appendix C. In

group 1 (the base-rate group), the materials consisted of the full set of 60 evidence sentences from experiment 1. The materials in groups 2 and 3 consisted of subsets of the 60 evidence sentences: In group 2, the materials consisted of the 44 evidence sentences recognized (task three) and documented (task four) by all of the 24 seniors in experiment 1 who decided *yes* there is substantial doubt. In group 3, the materials consisted of the 41 evidence sentences recognized and documented by all of the 16 seniors in experiment 1 who decided *no*.

Summarizing the experimental materials (Appendix C), group 2 (group 3) saw 100 (60)% of the *yes* sentences (Appendix C: 15 *yes* sentences seen/15 available = 100% (9 *yes* sentences/15 = 60%)), 67 (93)% of the *no* sentences, 67 (67)% of the *both* sentences, and 60 (53)% of the *neither* sentences. Thus, the distribution of sentences across groups was fairly equivalent: (a) group 3 saw almost as many *no* sentences (14) as group 2 saw *yes* sentences (15), (b) group 3 saw almost as many *yes* sentences (9) as group 2 saw *no* sentences (10), and (c) each group saw roughly the same number of *both* (10, 10) and *neither* (9, 8) sentences. Six *yes* sentences were seen by group 2 (but not by group 3), and related to changes in management (sentence numbers 13, 15), labor relations (12), demand/market share (4, 5), and tangible assets (10). The six *yes* sentences included the two (5, 10) that the seniors in experiment 1 who decided *yes* documented significantly more often than the seniors who decided *no* (Table 2, Panel A). Four *no* sentences were seen by group 3 (but not by group 2), and related to demand/market share (sentence number 24), financial results (17, 18), and litigation, claims and assessments (28). The four *no* sentences included the three (17, 18, 24) that the seniors in experiment 1 who decided *no* documented significantly more often than the seniors who decided *yes* (Table 2, Panel A).

Twenty audit partners were assigned to each group. To counterbalance experience level across groups, subjects were assigned to groups based on years of experience (mean = 15.2 years per group).

4.3. Procedure and task

The experiment was held at the same site and in the same city as experiment 1. All three treatments were completed on the same morning about one month after experiment 1: the group 1 treatment was administered first, group 2 second, and group 3 third.

The author was not present at the experimental site. Experiment 2 was administered by a partner (but not the same partner who administered experiment 1)¹⁰ and followed the same procedure as experiment 1, task one. However, the sentences in each group were displayed simultaneously on the subjects' screens, since experiment 2 did not impose a recognition memory task. Although inconsistent with audit practice, the partners were not told the seniors' decisions in experiment 1. This was done to rule out the alternative interpretation that the partners' decisions in groups 2 and 3 may have been biased by the seniors' decisions in experiment 1.

After reading the sentences, the subjects in each group performed a single task: indicate *yes* there is, or *no* there is not, substantial doubt about the company's ability to continue as a going concern, the same task imposed in experiment 1. Within group, the evidence sentences were ordered randomly for each subject.

4.4. Results: experiment 2

Table 3 reports substantial-doubt decisions by treatment group. The decisions indicate that, as expected, in comparison to group 1 (the base rate group), the decisions in group 2 and in group 3 shifted in the direction of the audit seniors' decisions in experiment 1. That is, in group 1, the group that had access to the full set of 60 evidence sentences, 9 partners decided *yes* there is substantial doubt and 11 decided *no*. However, 14 partners decided *yes* (6 *no*) in group 2, the treatment group that had access to six *yes* sentences not seen by group 3, two of which (sentences 5 and 10) the seniors in experiment 1 who decided *yes* documented significantly more often than the seniors who decided *no*. And 15 partners decided *no* (5 *yes*) in group 3, the treatment group that had access to four *no* sentences not seen by group

¹⁰The partner who administered experiment 2 reports to, and was rehearsed by, the national office partner who had administered experiment 1.

Table 3

Substantial doubt decisions by evidence-sentence treatment group. Experiment 2: audit partners

| Evidence-sentence treatment | | Subsets of sentences recognised and documented in experiment 1 by audit seniors who made | |
|-----------------------------|--|--|-------------------------------|
| Substantial doubt Decisions | Full set of evidence sentences (Group 1, $n = 20$) | Yes | No |
| | | Decision (Group 2, $n = 20$) | Decision (Group 3, $n = 20$) |
| Yes | 9 | 14 | 5 |
| No. | 11 | 6 | 15 |

Overall $X^2(2, n = 60) = 8.4$, $p < 0.015$ Group 1 vs Group 2 $X^2(1, n = 40) = 3.4$, $p < 0.065$.Group 1 vs Group 3 $X^2(1, n = 40) = 5.2$, $p < 0.023$ Group 2 vs Group 3 $X^2(1, n = 40) = 8.2$, $p < 0.004$.

2, three of which (17, 18 and 24) the seniors in experiment 1 who decided *no* documented significantly more often than the seniors who decided *yes*.

The pattern of decisions across all three groups is significantly different ($X^2(2, n = 60) = 8.4$, $p < 0.015$). The difference in decisions between groups 1 and 2 is marginally significant ($X^2(1, n = 40) = 3.4$, $p < 0.065$), and between groups 1 and 3 is significant ($X^2(1, n = 40) = 5.2$, $p < 0.023$). The difference in decisions between groups 2 and 3—the two groups who read only the evidence sentences that the seniors in experiment 1 recognized and documented—is significant ($X^2(1, n = 40) = 8.2$, $p < 0.004$). The results of experiment 2 indicate that audit partners' substantial-doubt decisions are biased by the evidence that audit seniors recognize and document in working papers. Because the sentences were displayed simultaneously on the subjects' screens, there is no *a priori* reason to believe that differences in behavior across groups is driven by the different numbers of sentences in each group.

5. Summary and discussion

This study tests whether audit seniors' decisions bias working paper documentation, and whether

biased documentation affects the decisions of audit partners who are exposed only to a *subset* of the evidence that seniors evaluated. Experiment 1 provides evidence that audit seniors' prior decisions affect their ability to recognize evidence to document in working papers. Although the seniors' prior decisions did not affect their documentation preferences for most of the evidence, they found that evidence consistent with their decisions was more important to document. The results of experiment 1 confirm Libby and Trotman (1993): together, the two papers offer evidence that the effects of audit seniors' prior decisions on memory are robust both to recall tasks (Libby and Trotman) and to recognition tasks (this paper). Experiment 2 provides evidence that, when exposed only to a *subset* of the evidence that seniors evaluated, audit partners make substantial-doubt decisions biased in the direction of the seniors' decisions. The results extend Libby and Trotman who found that, when exposed to the *same* set of evidence that preparers evaluated, reviewers can mitigate the effects of memory bias by evaluating inconsistent evidence. The results of this study suggest that public accounting firms may undermine audit effectiveness by empowering audit seniors to document evidence that supports their decisions *before* the decision is reviewed by partners. If documented audit evidence is biased toward an audit senior's decision, the audit

partner's decision may also be biased. Documenting before partner review increases the likelihood that the working papers will exclude evidence that is inconsistent with audit seniors' decisions, evidence that prior research (Libby and Trotman) reveals can serve as a control for biased memory if included in the review process.

The results of the justification manipulation in experiment 1 are inconsistent with Koonce et al. (1995) and Tan (1995, experiment 2). Koonce et al. found that auditors who expect a review will document more justifications, and Tan found that auditors who are aware of a review are more "vigilant". However, the results reported here do not necessarily contradict prior research for two reasons. First, the Koonce et al. experiments are set in a different audit issue (audit time-budget planning). Second, both prior studies use experimental manipulations for justification that are more rigorous than mine. For example, in addition to review condition, Koonce et al. also manipulated evidence between-subjects on three levels: corroborating evidence, inconsistent evidence, and no evidence.

Several limitations bear on the interpretation of the results reported in this paper. First, the procedure and tasks imposed in this study do not completely map realistically to the substantial-doubt decision task in practice. For example, the experimental setting in experiment 1 did not capture memory tasks in the same way that audit practice would impose the tasks, and the partners in experiment 2 were not told the seniors' decisions in experiment 1. However, as Swieringa and Weick (1982, p. 63) point out, what matters in applied research is not whether the experimental setting is real, but whether the setting elicits real behavior. Consistent, for example, with Earley, Wojnarowski, & Prest (1987) in psychology and with Kennedy (1993) in auditing, the author attempted in a debriefing question to measure indirectly whether the subjects' effort in the experiments may have overcome the artificial experimental setting. The mean reported effort ratings (7 = high effort) in experiments 1 and 2, respectively, were reasonably high: 6.24 and 6.17.

Second, experiment 1 (task four) collected binary data (1 = yes, 2 = no) about whether the

seniors would document recognized evidence sentences in the working papers. However, the task does not capture how and in what form the seniors would document the evidence, thereby overlooking the fact that in practice there are degrees of documentation which map to degrees of persuasiveness. For example, evidence sentence number 13, "The chief financial officer resigned from the company late last year" (a *yes* sentence recognized and documented by all 24 seniors who decided *yes*, there is substantial doubt), could be documented in the working papers simply as a "yes" response in a questionnaire or more elaborately (and more persuasively) in a memorandum for the files that summarizes the CFO's influential link to local financial markets. Binary data also limits the ability of the design to probe subtle differences in the working paper preferences of the seniors who decided *yes* and the seniors who decided *no*. For example, while interval data (25 point response scale) detected significantly different importance ratings, binary data failed to detect significantly different documentation patterns for 54 of the 60 evidence sentences, suggesting that the documentation results of experiment 1 may be an artifact of the design.

Third, experiment 2 assumes unrealistically that audit partners have access only to working paper evidence and that evidence is uniformly voluminous across engagements. However, a national office partner speculated that, compared to large engagements, an audit partner's substantial-doubt decision in a middle-market engagement would likely be vested in far fewer working paper cues, thereby reducing the demands on memory for working paper evidence. Fourth, despite the statistically significant difference in the pattern of correct recognitions in experiment 1, the participating seniors were generally able to discriminate between evidence sentences that had and had not been presented. For example, the 24 seniors who decided *yes* there is substantial doubt correctly recognized 86.4% of the targets referring to the *yes* decision and 74.2% of the targets referring to the *no* decision, yielding failure rates of 13.6% (100.0–86.4) and 25.8% (100.0–74.2). Gardiner and Tulving (1980) argue that it is not the level of recognition failure (e.g. 25.8%) that is important,

but rather the extent of recognition failure given the overall level of recognition (e.g. 34.8%: 25.8/74.2). Nevertheless, the results raise the question, Do statistical differences in an experimental setting necessarily map to economic differences in practice? An answer awaits research that focuses more sharply on the size of the recognition-bias effect documented in this paper.

Fifth, in practice, an audit manager's review of an audit senior's working papers could lead to questions that would compel the senior to discover and document evidence that might be inconsistent with the senior's decision, particularly if the manager did not have prior involvement with the engagement (Tan, 1995, experiment 1). The participating firm was reluctant to administer an experiment bearing more directly on manager reviews because firm policy dictates that the senior is responsible to gather evidence for the going-concern decision, suggesting that the manager would not likely be aware of evidence not documented by the senior. However, to the extent that policy differs in other firms, the effect of manager reviews on the amount of evidence documented in working papers is both an open question and a limitation on the implications of this study.

Acknowledgements

The author expresses appreciation to the Deloitte & Touche Foundation for financial support, to the participating firm for subjects and software, and to Brian Ballou, Fred Phillips, Sandra Vera-Muñoz and two anonymous reviewers for insightful comments on earlier versions of this paper.

Appendix A

Evidence sentences classified by decision membership experiment 1

Yes decision

- 1 The company will report a net loss for the second year in a row.
- 2 The current ratio is the lowest in the industry this year.

- 3 Domestic and world wide demand for wire and cable products is down.
- 4 Competition from foreign imports has risen rather sharply this year.
- 5 Management signed an uneconomic purchase commitment for copper.
- 6 A principal customer has not yet placed a large expected order.
- 7 A patent for a unique production process will expire next year.
- 8 The company was denied credit from a long time copper supplier.
- 9 A newspaper article reported world copper production declines.
- 10 Long term commercial bank loan covenants restrict some fixed asset disposals.
- 11 The company defaulted on an interest payment for a bank loan.
- 12 Employees organized a work stoppage about fringe benefits this year.
- 13 The chief financial officer resigned from the company late last year.
- 14 The president is considering a lucrative employment offer.
- 15 The vice president for manufacturing resigned effective next month.

No Decision

- 16 Sales revenues have increased by five percent over the prior year.
- 17 Total assets have not changed significantly over the prior year.
- 18 The debt to equity ratio is down compared to the prior year.
- 19 No trade customer receivable accounts are ninety days or older.
- 20 No new long term liabilities have been incurred in the current year.
- 21 Days sales in receivables is less than each of the prior five years.
- 22 The company has no long term bond indebtedness outstanding this year.
- 23 Management agreed in substance to ship to a new foreign company.
- 24 Management signed a long term contract with a new domestic customer.
- 25 Unused productive assets are maintained and readily marketable.

- 26 There are no arrangements with banks that require compensating balances.
- 27 Confirmations reveal no unknown debt to financial institutions.
- 28 No violations of laws require disclosure as loss contingencies.
- 29 The company has complied with all agreements that are contractual.
- 30 A lawyer's letter indicates that there is no current litigation.

Both Decisions

- 31 The company has not factored any trade customer receivables.
- 32 The company contemplates issuing new convertible bonds next year.
- 33 The company plans a public offering of securities next year.
- 34 A dividend has not been declared in seven consecutive quarters.
- 35 The company plans a fifty percent reduction in research costs.
- 36 The company postponed a product development consulting project.
- 37 The company operates at seventy percent of capacity.
- 38 The company delayed scheduled maintenance on unused fixed assets.
- 39 The company has not fully insured the replacement cost of assets.
- 40 Management canceled a personal injury insurance policy.
- 41 The company renegotiated a long term commercial bank loan.
- 42 The company has forgiven a personal loan to a director.
- 43 A newspaper reported the threat of a patent infringement suit.
- 44 The company sought to buy a company that has loss carryovers.
- 45 A newspaper reported that production workers voted not to strike.

Neither Decision

- 46 The company was incorporated about forty five years ago.
- 47 The company's common and preferred stock is traded over the counter.

- 48 The company manufactures and repairs wire and cable products.
- 49 The engagement letter called for an audit and federal tax return.
- 50 Independence has not been impaired by any audit staff assignments.
- 51 The company imposed no restrictions on any audit procedures.
- 52 Interim audit procedures indicate that inherent risk is low.
- 53 Interim audit tests of controls indicate that control risk is low.
- 54 Audit procedures indicate acceptable detection risk is low.
- 55 The interim estimate of materiality is moderate.
- 56 Tolerable deviation rates exceed estimates for all controls.
- 57 Substantive tests have been performed for all financial statement assertions.
- 58 Tolerable error exceeded projected error for all accounts.
- 59 Two wire and cable competitors reported net losses last year.
- 60 All employee claims for workmans compensation insurance have been paid.

Appendix B

Lure sentences classified by decision membership and by related target sentence numbers (Appendix A): experiment 1

Yes decision

- 1 The company will report negative cash flows for the second straight year.
- 2 Current assets will barely exceed current liabilities this year.
- 6 New orders from continuing customers are below expectations.
- 7 A patent application for a new production process has been scrapped.
- 9 A newspaper article reports domestic copper supply declines.
- 10 Long term commercial bank loan covenants on assets have been violated.
- 12 Labor union officials have threatened a strike over fringe benefits.

No decision

- 16 Management expects sales volume to increase by five percent next year.
- 18 Debt and current liabilities are down compared to the prior year.
- 19 No trade accounts payable balances are more than thirty days past due.
- 21 Inventory turned over more than in any of the prior five years.
- 23 Management is seeking new factory orders from foreign companies.
- 28 No violations of contracts will qualify as loss contingencies.
- 30 A lawyer's letter indicates no unasserted claims or assessments.

Both decisions

- 34 The company has expressed no plans to declare a dividend next year.
- 35 The company plans a reduction in pollution control expenses.
- 36 The company postponed an employee benefits consulting project.
- 37 The company had seventy percent excess capacity this year.
- 40 Management canceled a vehicle collision insurance policy.
- 42 A related party has agreed to make a loan to the company.
- 45 A trade journal reports that factory employees will not likely strike.

Neither decision

- 47 The company's common and preferred stock is traded in foreign markets.
- 50 Independence has not been impaired by any financial arrangements.
- 51 The company has not imposed restrictions on access to evidence.
- 54 Audit procedures indicate that allowable audit risk is low.
- 55 Planned materiality exceeds projected monetary error.
- 59 Two wire and cable competitors reported poor cash flows last year.
- 60 No new claims for workmans compensation insurance are being processed.

Appendix C

Evidence sentence numbers (Appendix A) seen by groups 1, 2, and 3: Experiment 2

| | <u>Group</u> | | | | <u>Group</u> | | |
|--------------|--------------|-----------------|-----------------|------------------|--------------|-----------------|-----------------|
| Yes decision | 1 | 2 ¹¹ | 3 ¹² | Both decisions | 1 | 2 ¹¹ | 3 ¹² |
| 1 | x | x | x | 31 | x | | x |
| 2 | x | x | x | 32 | x | x | |
| 3 | x | x | x | 33 | x | | x |
| 4 | x | x | | 34 | x | x | x |
| 5 | x | x | | 35 | x | x | x |
| 6 | x | x | x | 36 | x | x | x |
| 7 | x | x | x | 37 | x | x | x |
| 8 | x | x | x | 38 | x | | |
| 9 | x | x | x | 39 | x | x | |
| 10 | x | x | | 40 | x | x | |
| 11 | x | x | x | 41 | x | x | x |
| 12 | x | x | | 42 | x | | x |
| 13 | x | x | | 43 | x | x | |
| 14 | x | x | x | 44 | x | x | x |
| 15 | <u>x</u> | <u>x</u> | | <u>45</u> | <u>x</u> | | <u>x</u> |
| | 15 | 15 | 9 | | 15 | 10 | 10 |
| No decision | | | | Neither decision | | | |
| 16 | x | x | x | 46 | x | x | x |
| 17 | x | | x | 47 | x | x | x |
| 18 | x | | x | 48 | x | x | x |
| 19 | x | x | x | 49 | x | x | x |
| 20 | x | x | x | 50 | x | x | |
| 21 | x | x | x | 51 | x | x | x |
| 22 | x | x | x | 52 | x | x | x |
| 23 | x | x | x | 53 | x | | |
| 24 | x | | x | 54 | x | | x |
| 25 | x | x | x | 55 | x | x | x |
| 26 | x | x | x | 56 | x | | |
| 27 | x | | | 57 | x | | |
| 28 | x | | x | 58 | x | | |
| 29 | x | x | x | 59 | x | | |
| 30 | <u>x</u> | <u>x</u> | <u>x</u> | <u>60</u> | <u>x</u> | <u>x</u> | |
| | 15 | 10 | 14 | | 15 | 9 | 8 |

Number of sentences seen by each group:

| | <u>Group</u> | | |
|---------|--------------|-----------|-----------|
| | <u>1</u> | <u>2</u> | <u>3</u> |
| Yes | 15 | 15 | 9 |
| No | 15 | 10 | 14 |
| Both | 15 | 10 | 10 |
| Neither | <u>15</u> | <u>9</u> | <u>8</u> |
| | <u>60</u> | <u>44</u> | <u>41</u> |

¹¹Sentences recognized (Task 3) and documented (Task 4) by all of the 24 seniors in experiment 1 who decided *yes*, there is substantial doubt. ¹²Sentences recognized (Task 3) and documented (Task 4) by all of the 16 seniors in experiment 1 who decided *no*, there is not substantial doubt.

References

- Abdolmohammadi, M. J., & Wright, A. (1987). An examination of the effects of experience and task complexity on audit judgments. *The Accounting Review*, 1–13.
- Anon (1988). *Statement on auditing standards No. 59: the auditor's consideration of an entity's ability to continue as a going concern*. American Institute of Certified Public Accountants.
- Asare, S. K., & McDaniel, L. S. (1996). The effects of familiarity with the preparer and task complexity on the effectiveness of the audit review process. *The Accounting Review*, 139–159.
- Choo F., & Trotman, K. T. (1991). The relationship between knowledge structure and judgments for experienced and inexperienced auditors. *The Accounting Review*, 464–485.
- Cushing, B.E., & Loebbecke, J. K. (1986). *Comparison of Audit Methodologies of Large Accounting Firms*. AAA.
- Dellarosa, D., & Bourne Jr, L. E. (1984). Decisions and memory: differential retrievability of consistent and contradictory evidence. *Journal of Verbal Learning and Verbal Behavior*, 669–682.
- Earley, P.C., Wojnaroski, P., & Prest, W. (1987). Task planning and energy expended: exploration of how goals influence performance. *Journal of Applied Psychology*, 107–114.
- Gardiner, J.M., & Tulving, E. (1980). Exceptions to recognition failure of recallable words. *Journal of Verbal Learning and Verbal Behavior*, 194–209.
- Haist, F., Shimamura, A. P., & Squire, L. R. (1992). On the relationship between recall and recognition memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 691–702.
- Kennedy, J. (1993). Debiasing audit judgment with accountability: a framework and experimental results. *Journal of Accounting Research*, 231–245.
- Koonce, L., Anderson, U., & Marchant, G. (1995). Justification of decisions in auditing. *Journal of Accounting Research*, 369–384.
- Libby, R., & Trotman, K. T. (1993). The review process as a control for differential recall of evidence in auditor judgment. *Accounting, Organizations and Society*, 559–574.
- Moeckel, C.L., & Plumlee, R. D. (1989). Auditors' confidence in recognition of audit evidence. *The Accounting Review*, 653–666.
- Pennington, N., & Hastie, R. (1988). Explanation-based decision making: effects of memory structure on judgment. *Journal of Experimental Psychology: Learning, Memory and Cognition*, 521–533.
- Ramsey, R. J. (1994). Senior/manager differences in audit workpaper performance. *Journal of Accounting Research*, 127–135.
- Ricchiute, D. N. (1992). Working paper order effects and auditors' going-concern decisions. *The Accounting Review*, 46–58.
- Rich, J.S., Solomon, I., & Trotman, K. T. (1997). The audit review process: a characterization from the persuasion perspective. *Accounting, Organizations and Society*, 481–505.
- Singh, S.N., Rothschild, M. L., & Churchill Jr, G. A. (1988). Recognition versus recall as measures of television commercial forgetting. *Journal of Marketing Research*, 72–80.
- Swieringa, R.J., & Weick, K. E. (1982). An assessment of laboratory experiments in accounting. *Journal of Accounting Research*, Supplement, 56–101.
- Tan, H. (1995). Effects of expectations, prior involvement, and review awareness on memory for audit evidence and judgment. *Journal of Accounting Research*, 113–135.
- Trotman, K.T., & Yetton, P. W. (1985). The effect of the review process on auditor judgments. *Journal of Accounting Research*, 256–267.
- Tulving, E., & Schacter, D. L. (1990). Priming and human memory systems. *Science*, 301–306.