Psychonomic Bulletin & Review
2005, 12 (5), 852-857

BRIEF REPORTS

Modes of cognitive control in recognition and source memory: Depth of retrieval

LARRY L. JACOBY, YUJIRO SHIMIZU, KAREN A. DANIELS, and MATTHEW G. RHODES
Washington University, St. Louis, Missouri

Recognition memory is usually regarded as a judgment based on trace strength or familiarity. But recognition may also be accomplished by constraining retrieval so that only sought after information comes to mind (*source-constrained retrieval*). We introduce a *memory-for-foils* paradigm that provides evidence for source-constrained retrieval in recognition memory (Experiment 1) and source memory (Experiment 2). In this paradigm, subjects studied words under deep or shallow encoding conditions and were given a memory test (recognition or source) that required them to discriminate between new items (foils) and either deep or shallow targets. A final recognition test was used to examine memory for the foils. In both experiments, foil memory was superior when subjects attempted to retrieve deep rather than shallow targets on the earlier test. These findings support a source-constrained retrieval view of cognitive control by demonstrating qualitative differences in the basis for memory performance.

Recognition memory is traditionally described as relying on a judgment of unidimensionsal trace strength or familiarity. For example, in global activation models of memory (e.g., Gillund & Shiffrin, 1984), recognition is accomplished by comparing a memory probe's strength, familiarity, or activation against a decision criterion. If the probe's value exceeds criterion, it is accepted as "old"; otherwise, it is rejected as "new." The greater the match between a memory probe and traces in memory, including the match between study and test contexts, the greater the level of familiarity. Thus, the emphasis lies in the *quantitative* relationship between the probe and criterion. In contrast, we argue that recognition is sometimes accomplished by considering the *kind* of memory that is sought. Our position is that recognition often involves *source-constrained retrieval*—the self-initiated use of source information to constrain what comes to mind during retrieval. Specifically, we suggest that processes implemented during study are sometimes reimplemented during retrieval by the rememberer. As a result, source-constrained retrieval influences not only what is accepted or rejected, but also what information is used to make recognition decisions.

Experiment 1 was designed to show that source-constrained retrieval can produce a *qualitative* change in the type of information used for recognition memory judgments and influence the manner in which both old and new test items are processed. During the first phase of Experiment 1 (Figure 1), level of processing (e.g., Craik & Lockhart, 1972) was manipulated by having subjects make pleasantness judgments for words in one list and judgments about the vowels of words in another list. In Phase 2, subjects were given two tests of recognition memory. For one test of recognition memory, they were correctly informed that all "old" words had been judged for pleasantness. For another test, they were correctly informed that all "old" words had been vowel judged. We expected results from these tests to replicate results of prior experiments (e.g., Craik & Lockhart, 1972). That is, judging the pleasantness of words (deep processing) should produce recognition memory performance better than that for vowel judgments (shallow processing).

To gain evidence of source-constrained retrieval, a third phase of Experiment 1 tested recognition memory for *new items* (*foils*) that appeared on the earlier recognition memory tests. By the source-constrained retrieval view, it was predicted that foils from the recognition test for pleasantness-judged items were more likely to be later recognized than foils appearing in the recognition test for vowel-judged items. This follows because recognition is held to be accomplished by constraining retrieval processing in a way that recapitulates study processing. Consequently, when attempting to recognize pleasantness-judged old words, subjects would likely process the meaning of *both* targets and foils, perhaps considering each test word's pleasantness to determine whether they had made a similar judgment previously. In contrast, attempting to recognize vowel-judged, old words would likely rely less on the processing of meaning. The deeper processing of foils when pleasantness-

This research was supported by National Institute on Aging Grant AG13845. We are grateful to Carole Jacoby for her data collection assistance. Correspondence concerning this article should be addressed to L. L. Jacoby, Department of Psychology, Washington University, St. Louis, MO 63130 (e-mail: lljacoby@artsci.wustl.edu).

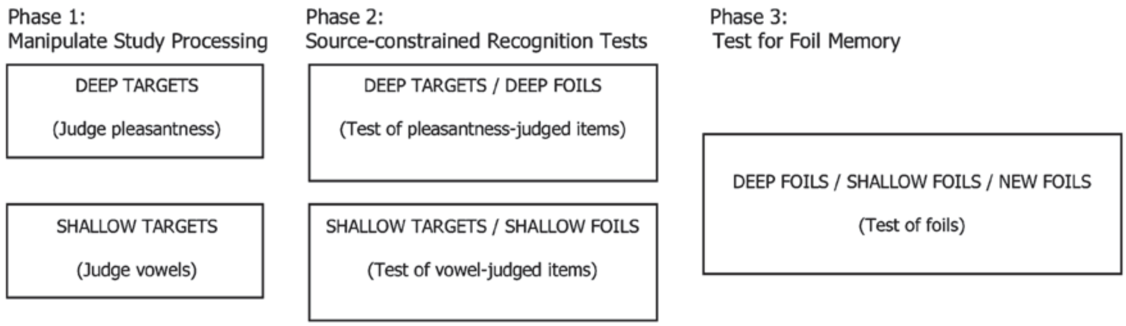

Figure 1. The memory for foils paradigm of Experiment 1. All variables were manipulated within subjects.

judged words were targets was expected to result in higher subsequent recognition memory for those foils.

If subjects simply assess global familiarity when making recognition judgments (e.g., Gillund & Shiffrin, 1984), there would be no reason to expect differential processing of the foils depending on the study processing of targets, and, consequently, no reason to expect differences in subsequent memory for foils. Any predictions based on a familiarity account would likely suggest that increasing the difficulty of recognition decisions would increase the exposure duration of foils and, thereby, enhance their later recognition. Because recognition memory is expected to be poorer after shallow processing, this would suggest that differences in recognition of foils would favor targets that were shallowly processed, a prediction opposite that made by the source-constrained retrieval view. Consequently, better memory for foils when targets were deeply processed would provide direct evidence of differences in *retrieval depth* that reflect source-constrained retrieval.

Source-constrained retrieval can be considered a means of cognitive control. This may be illustrated by considering an example of quality control in a manufacturing situation. An obvious means of quality control is to have inspectors who monitor manufactured products, rejecting those that do not satisfy specified criteria. However, quality control may also be accomplished by increasing production constraints so as to decrease the probability of producing defective products, thereby reducing the need for postproduction inspectors. Similarly, one method of controlling memory is to edit potential responses after they come to mind and select only those that are sufficiently familiar. An alternative method of cognitive control is to tightly constrain retrieval so that only memories from a questioned source come to mind. Thus, the contrast is between gaining control by restricting memory access (*source-constrained retrieval*) in comparison with relying on a postaccess process (*source identification*).

Although it is likely that both forms of cognitive control are sometimes employed (e.g., Burgess & Shallice, 1996; Jacoby, Kelley, & McElree, 1999), prior theorizing has primarily emphasized source identification, evident in work on source memory. Source memory is typically tested by asking subjects to decide whether a test item has originated from one of several sources (e.g., read vs. heard) or is new (for a review, see Johnson, Hashtroudi, & Lindsay, 1993), and it is generally assumed to follow an initial recognition judgment. A finding of effects of source-constrained retrieval for a standard test of recognition memory (Experiment 1) raises the possibility that source-constrained retrieval also operates when subjects must identify the source of previously presented items. This possibility was examined in Experiment 2. We will conclude by discussing the implications of source-constrained retrieval for theories of recognition and source memory.

EXPERIMENT 1

Method

Subjects. Sixteen undergraduates at Washington University participated for course credit. All variables were manipulated within subjects.

Materials and Procedure. In Phase 1, subjects made pleasantness judgments for 36 words in one list and vowel judgments for 36 words in another list, indicating whether a word included an O or U. All words were presented for 1.5 sec each in the center of the screen, and judgments were made as keypresses.

In Phase 2, subjects received deep and shallow recognition memory tests. For the *deep* recognition memory test, words whose pleasantness had been judged were intermixed with an equal number of new words (i.e., foils). Subjects were correctly informed that all of the "old" words in the test list were from the pleasantness-judged list. For a separate, *shallow* recognition memory test, the subjects were correctly informed that all "old" words in the test list were presented in the vowel judgment list. In Phase 3, three types of words appeared in a recognition memory test of foils: 36 *deep foils* (presented as new items in the deep recognition memory test); 36 *shallow foils* (presented as new items in the shallow recognition memory test); and 72 *new foils* (words that were not presented earlier). The subjects were instructed to judge a word as "old" if it had been presented earlier during any phase of the experiment, and to respond "new" only if the word had not been presented earlier. All recognition tests were self-paced, with responses made as keypresses.

The stimuli were 272 words (216 critical and 56 buffers) matched in frequency (1–104 per million, $M = 15.67$; Kučera & Francis, 1967), and length (4–8 letters, $M = 6.67$). The order of orienting tasks in Phase 1 and recognition tests in Phase 2 was fully counterbalanced across subjects, as was the assignment of words to each condition.

Results and Discussion

All statistical analyses used an α level of .05. Consistent with previous work, the probability of correctly calling an item "old" was higher for the deep than for the shallow recognition memory test (.94 vs. .61) whereas false alarms were lower for the deep than for the shallow test (.05 vs. .16). This interaction of level of processing and old/new status was significant [$F(1,15) = 76.50$].

Of primary interest was whether subsequent recognition memory would differ for foils from the deep and shallow tests. A source-constrained retrieval view predicts better recognition memory performance for deep foils. As expected, recognition memory was significantly better for the deep foils than for the shallow foils (.84 vs. .76) [$F(1,15) = 5.67$]. The false alarm rate for new foils was .13.

Perhaps foils that were falsely recognized or whose rejection was slow on the earlier recognition memory test were more likely to be later recognized. Against this possibility, recognition memory was highest for foils from the pleasantness-judged test, which had the lowest false alarm rate on the initial test. In addition, rejection time, which corresponds to processing or study time, was actually significantly shorter (1,101 msec) for deep foils than for shallow foils (1,352 msec) [$t(15) = 2.48$].

Overall, the results provide direct evidence for source-constrained retrieval, as reflected by memory for foils. Specifically, attempting to recognize old items that were deeply processed during study resulted in greater depth of processing at retrieval and thus better memory for foils than did attempting to recognize items that were shallowly processed during study. This was true even though foils were more quickly and accurately rejected when deeply processed old words were targets. Effects on foil memory reveal qualitative differences in bases for recognition memory that cannot be explained by a model that treats trace strength or global familiarity as the sole basis for recognition. Consistent with our findings of differences in foil memory, Rugg, Allan, and Birch (2000) found differences in event-related potentials for new words that depended on whether old words in a recognition test had been studied in a deep or in a shallow encoding task.

In later experiments (Jacoby, Shimizu, Velanova, & Rhodes, 2005), we have replicated the pattern of results reported here for memory for foils using a between-subjects manipulation of level of processing. The between-subjects manipulation allows subjects to know about the prior processing of target items but has the advantage of employing standard recognition memory test instructions, providing further evidence that source-constrained retrieval plays a role in recognition memory.

EXPERIMENT 2

As in Experiment 1, the level of processing of target words was manipulated in Experiment 2. However, source memory was assessed in Experiment 2 by requiring subjects to accept old words from a target source and to exclude old words from a nontarget source (cf. Jacoby, 1999). In one test condition, subjects were instructed to accept pleasantness-judged, old words and to exclude words that were only heard, whereas subjects in a second test condition received the opposite instructions. A third test condition employed a standard recognition memory test for which subjects were instructed to accept *both* pleasantness-judged and heard old words and reject only new words. Test lists were the same for the three conditions with only the instructions varied.

Instructing subjects to respond "yes" *only* for words whose pleasantness had been judged was expected to result in deeper retrieval processing and better subsequent recognition of foils than would be found in the other two test conditions. A finding of effects of test instructions on memory for foils would demonstrate that source-constrained retrieval plays a role in performance of a source memory task. In contrast, others (e.g., Bayen, Murnane, & Erdfelder, 1996) have held that source memory relies solely on a source-identification process that follows recognition memory.

Method

Subjects. Fifty-four undergraduates at Washington University participated for course credit. Eighteen subjects were assigned to one of three test conditions.

Materials and Procedure. During Phase 1, the subjects read and made pleasantness judgments for 30 words in a first list and then heard 30 words in a second list that they were instructed to remember for a subsequent test. The pleasantness judgments were self-paced, and the heard words were presented at a 2-sec rate. In Phase 2, the subjects were given one of three memory tests that included all words from the two encoding lists intermixed with 30 new words. The subjects in a *standard recognition memory* test condition were instructed to say "yes" to old words, accepting both pleasantness-judged words and earlier heard words, and to say "no" to new words. The subjects in a *pleasantness-judged* source memory test condition were instructed to accept *only* old words whose pleasantness had been judged, and to reject both earlier heard words and new words. Finally, subjects in an *earlier heard* source memory test condition were instructed to accept only words that they had heard and to reject both pleasantness-judged and new words. In Phase 3, the subjects from each of the three test conditions were given a recognition memory test for foils, comprising 30 "old" foils presented in Phase 2, and 30 "new" foils.

The stimuli were 120 words, 4–6 letters long, subdivided into four sets matched for frequency (2 to 25 per million according to Kučera & Francis, 1967) and word length. The pleasantness–read task always preceded the auditory–remember task, and the stimuli in these two conditions were fixed. Two final lists were counterbalanced across subjects in such a way that they served equally often as new items in the recognition tests. The presentation of words within a list was random.

Results and Discussion

Table 1 displays hits and false alarms along with corrected recognition (hits minus false alarms) data for each of the three test conditions. Corrected recognition for pleasantness-judged words was significantly higher when only pleasantness-judged words were to be accepted than in the standard recognition memory test condition (.93 vs. .85) [$t(34) = 2.45$, $SE = 2.95$]. Recognition of heard items, although numerically higher when given standard recognition instructions as opposed to instructions to identify only earlier heard items (.66 vs. .56), was not statistically different for the two groups [$t(34) = 1.32$, $SE = 7.83$, $p = .19$].

Table 1
Hits, False Alarms, and Difference Scores for Pleasantness-Judged and Earlier Heard Items for the Standard Recognition Memory Test (Within Subjects) and the Source Memory Tests (Between Subjects) in Experiment 2

Memory Test	Hits	False Alarms	Difference
Standard Recognition			
Pleasantness	.95	.10	.85
Heard	.76		.66
Source			
Pleasantness	.94	.01	.93
Heard	.73	.17	.56

Differences between the standard recognition memory condition and the two source memory test conditions resulted from significant differences in the probability of a false alarm [$F(2,51) = 13.15$, $MS_e = 91.05$]. The largest differences existed between the pleasantness-judged condition and the other two test conditions (ts = 4.81 and 4.61 for the standard recognition and earlier heard test conditions, respectively). Hits did not differ between the standard recognition and source test conditions ($F < 1$). This is particularly noteworthy: If differences in recognition resulted solely from differences in response criterion, one would expect that differences would exist for *both* false alarms and hits. Instead, differences were evident only in false alarms, suggesting that differences in recognition reflect differences in the ability to discriminate between old and new words as well as differences in response criterion.

The finding that differences in criterion did not produce a change in the probability of a hit may simply be due to chance. Alternatively, specifying the source of target items may produce a qualitatively different criterion for rejecting new items while leaving the hit rate unchanged. By the latter account, deciding whether a word was earlier pleasantness judged served as a surer basis for rejecting foils than did judging whether a word was earlier heard. The intermediate false alarm rate for the standard recognition condition might reflect a mix of the two bases for rejection. That is, subjects may have been least likely to recapitulate study processing (e.g., judging the pleasantness of test words) in the earlier heard condition, and engaged in this strategy to a greater extent in the standard recognition condition.

Given that source identification is generally assumed to follow recognition memory (e.g., Bayen et al., 1996), the preceding differences in recognition memory cannot have resulted from differences in source identification. Rather, they reflect qualitatively different bases for recognition memory through the use of source-constrained retrieval. Direct evidence of this may be obtained by examining memory for foils (Table 2). Memory for foils was best when subjects were instructed to endorse only pleasantness-judged items. Specifically, recognition memory for foils was significantly higher in the pleasantness-judged test condition (.76) than in both the standard recognition-memory test condition (.62) [$t(34) = 2.92$] and the earlier heard test condition (.63) [$t(34) = 2.51$]. These effects resulted largely from differences in false alarms [$F(2,15) = 4.59$] with significant differences emerging between the pleasantness-judged condition and the other test conditions (ts = 2.53 and 2.86 for standard recognition and earlier heard tests, respectively). No significant differences among groups were found for hits ($F < 1$). Because there were no differences in hits, differences in foil recognition cannot be explained as solely reflecting a difference in criterion but instead must also reflect a difference in ability to discriminate between old and new words.

As in Experiment 1, the effects on foil memory cannot be accounted for by alternative explanations based on differences in false recognition or rejection time during the initial recognition memory tests. Foil memory was highest in the pleasantness-judged test condition, which had the lowest false alarm rate on the initial test. Arguing against a rejection time account, the earlier heard test condition—a group with significantly *worse* foil recognition performance than that for the pleasantness-judged test group—had the slowest rejection time (1,676 msec) for new items, in comparison with rejection times for the pleasantness judgment (1,397 msec) and standard recognition (1,321 msec) test groups.

In addition to rejecting new words, the subjects in the pleasantness-judged test condition were instructed to reject earlier heard words and the subjects in the earlier heard test condition were instructed to reject pleasantness-judged words. These instructions were the same as those used for *exclusion* conditions in process dissociation research (e.g., Jacoby, 1991). An analysis of exclusion and new-item performance for these two test conditions revealed a significant interaction [$F(1,34) = 11.99$, $MS_e = 51.07$]. False alarms were more likely for earlier heard words than for new words in the pleasantness-judged test condition (.07 vs. .01), whereas in the earlier heard test condition, false alarms were less likely for pleasantness-judged words than for new words (.11 vs. 17). Interactions of this sort have been used as support for a dual-process model which holds that recollection and familiarity serve as alternative bases for recognition memory (see, e.g., Jacoby, 1999). By that model, the better exclusion performance of pleasantness-judged words reflects an advantage in recollection for those words as opposed to earlier heard words.

Table 2
Hits, False Alarms, and Difference Scores for the Foil Recognition Test Following the Three Initial Test Conditions in Experiment 2

Memory Test	Hits	False Alarms	Difference
Standard Recognition	.82	.20	.62
Source			
Pleasantness	.88	.12	.76
Heard	.86	.23	.63

GENERAL DISCUSSION

The results from the present experiments indicate that specifying the source of old items produced a qualitative change in the type of information used for memory judgments. Such effects of source-constrained retrieval were demonstrated in a standard recognition memory test (Experiment 1) and a source memory test (Experiment 2). In both experiments, recognition memory for foils was higher when target items were earlier deeply processed rather than shallowly processed. This was true even though foils were more accurately and quickly rejected when targets had been deeply processed at test. The depth of processing of the foils, not the difficulty of their rejection, was responsible for their subsequent recognition.

Memory for foils provides direct evidence of source-constrained retrieval, evidence that converges with results reported by others. Humphreys and his colleagues (2003) have shown the importance of test instructions for memory access in exclusion tasks such as those used in our Experiment 2, and they have argued that memory decisions are based on a match between a reinstated context, including processing context, and a context that is retrieved using the probe as a cue (cf. Marsh & Hicks, 1998). Effects of source-constrained retrieval reveal qualitative differences in the criteria used to reject foils, showing that the processing of foils is relative to the specification of target information. In contrast, formal models of recognition memory have highlighted the importance of quantitative criteria (e.g., strength of global familiarity in the model proposed by Gillund & Shiffrin, 1984) for accepting an item as "old," largely neglecting qualitative differences in criteria.

The results of Experiment 2 showed that subjects instructed to accept only pleasantness-judged words exhibited both superior recognition memory and better memory for foils than did subjects given standard recognition memory instructions. These data suggest problems for measurement models of source identification that hold that recognition memory, presumably based on familiarity, is independent of source memory (e.g., Bayen et al., 1996). By those models, information about source is retrieved for recognized items in response to the source question but does not play a role in recognition memory per se, an assumption inconsistent with our results.

Effects of source-constrained retrieval on recognition also suggest problems for the inclusion/exclusion variant of the process dissociation procedure that has been used to separate the contributions of recollection and familiarity to recognition memory performance (e.g., Jacoby, 1991). The standard recognition memory condition in Experiment 2 corresponded to an inclusion test, whereas the other two test conditions in that experiment were exclusion tests. As Humphreys et al. (2003) have suggested, asking different memory questions for inclusion and exclusion tests can influence memory access and violate the assumption that familiarity is equivalent for the two test conditions, an assumption that the process dissociation procedure relies on to gain estimates of recollection and familiarity. Our results converge with those reported by Humphreys et al. in suggesting that the equal familiarity assumption is sometimes violated, and we agree with their conclusion that estimates of familiarity gained from the inclusion/exclusion procedure should be interpreted with care, particularly when there are large differences between conditions in the rejection of new items, such as those found in Experiment 2.

Our results show the value of memory for foils as a means of directly measuring reliance on source-constrained retrieval. We have also used effects on memory for foils to examine reliance on cognitively controlled retrieval, rather than familiarity, for recognition memory (Velanova et al., 2003) and to distinguish bases for responding in a working memory task (Speer, Jacoby, & Braver, 2003). We are currently using tests of foil memory to investigate potential age-related deficits in retrieval depth. The relatively poor performance of older adults on tests of source identification (e.g., Spencer & Raz, 1995) might arise, in part, from a diminished ability to constrain retrieval and be open to remediation. Just as remedying faulty quality control would differ depending on whether the problem arose from "production" or from "ineffective inspectors," one would follow very different courses for memory remediation depending on whether problems arose from a deficit in source-constrained retrieval or from a deficit in source identification.

REFERENCES

Bayen, U. J., Murnane, K., & Erdfelder, E. (1996). Source discrimination, item detection, and multinomial models of source monitoring. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **22**, 197-215.

Burgess, P. W., & Shallice, T. (1996). Confabulation and the control of recollection. *Memory*, **4**, 359-411.

Craik, F. I. M., & Lockhart, R. S. (1972). Levels of processing: A framework for memory research. *Journal of Verbal Learning & Verbal Behavior*, **11**, 671-684.

Gillund, G., & Shiffrin, R. M. (1984). A retrieval model for both recognition and recall. *Psychological Review*, **91**, 1-67.

Humphreys, M. S., Dennis, S., Maguire, A. M., Reynolds, K., Bolland, S. W., & Hughes, J. D. (2003). What you get out of memory depends on the question you ask. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **29**, 797-812.

Jacoby, L. L. (1991). A process dissociation framework: Separating automatic from intentional uses of memory. *Journal of Memory & Language*, **30**, 513-541.

Jacoby, L. L. (1999). Ironic effects of repetition: Measuring age-related differences in memory. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **25**, 3-22.

Jacoby, L. L., Kelley, C. M., & McElree, B. D. (1999). The role of cognitive control: Early selection versus late correction. In S. Chaiken & Y. Trope (Eds.), *Dual-process theories in social psychology* (pp. 383-400). New York: Guilford.

Jacoby, L. L., Shimizu, Y., Velanova, K., & Rhodes, M. G. (2005). Age differences in depth of retrieval: Memory for foils. *Journal of Memory & Language*, **52**, 493-504.

Johnson, M. K., Hashtroudi, S., & Lindsay, D. S. (1993). Source monitoring. *Psychological Bulletin*, **114**, 3-28.

Kučera, H., & Francis, W. N. (1967). *Computational analysis of present-day American English*. Providence, RI: Brown University Press.

Marsh, R. L., & Hicks, J. L. (1998). Test formats change source-monitoring decision processes. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **24**, 1137-1151.

Rugg, M. D., Allan, K., & Birch, C. S. (2000). Electrophysiological evidence for the modulation of retrieval orientation by depth of study processing. *Journal of Cognitive Neuroscience*, **12**, 664-678.

Speer, N. K., Jacoby, L. L., & Braver, T. S. (2003). Strategy-dependent changes in memory: Effects on behavior and brain activity. *Cognitive, Affective, & Behavioral Neuroscience*, **3**, 155-167.

Spencer, W. D., & Raz, N. (1995). Differential effects of aging on memory for content and context: A meta-analysis. *Psychology & Aging,* **10**, 527-539.

Velanova, K., Jacoby, L. L., Wheeler, M. E., McAvoy, M. P., Petersen, S. E., & Buckner, R. L. (2003). Functional–anatomic correlates of sustained and transient processing components engaged during controlled retrieval. *Journal of Neuroscience*, **23**, 8460-8470.

(Manuscript received June 23, 2004;
revision accepted for publication December 22, 2004.)