Journal of Memory and Language 61 (2009) 470–479

Contents lists available at ScienceDirect

Journal of Memory and Language

journal homepage: www.elsevier.com/locate/jml

Retrieval constraints on the front end create differences in recollection on a subsequent test

Richard L. Marsh [a,*], J. Thadeus Meeks [b], Gabriel I. Cook [c], Arlo Clark-Foos [d], Jason L. Hicks [e], Gene A. Brewer [a]

[a] Department of Psychology, University of Georgia, USA
[b] Department of Psychology, Southern Illinois University Edwardsville, USA
[c] Department of Psychology, Claremont McKenna College, USA
[d] Department of Psychology, University of Michigan-Dearborn, USA
[e] Department of Psychology, Louisiana State University, USA

ARTICLE INFO

Article history:
Received 20 November 2008
Revision received 15 June 2009
Available online 19 August 2009

Keywords:
Memory
Recollection
Familiarity
Retrieval modes
Recognition memory

ABSTRACT

Four experiments were conducted to investigate how the cognitive control of memory retrieval selects particular qualitative characteristics as a consequence of instantiating a retrieval mode for recognition memory. Adapting the memory for foils paradigm from Jacoby, Shimizu, Daniels, and Rhodes (Jacoby, L. L., Shimizu, Y., Daniels, K. A., & Rhodes, M. G. (2005a). Modes of cognitive control in recognition and source memory: Depth of retrieval. *Psychonomic Bulletin & Review*, *12*, 852–857), we demonstrate that inspecting distractors under different retrieval modes leaves more versus less recollective details in memory. Participants also reported using different rejection strategies under different retrieval modes. The result that retrieval modes are based on selecting qualitative characteristics of the study phase to aid recognition memory was demonstrated by showing that participants cannot establish unique retrieval modes based solely on trace strength or relative judgments of strength. Whether early selection versus late correction criteria will be applied during a memory test depends on the particular retrieval mode established by the characteristics of the recognition memory test.

Introduction

Tulving (1983) has argued that when scholars test human memory they place their participants in a "retrieval mode" which can be defined as a cognitive readiness to evaluate stimuli for evidence of having occurred within a particular place and/or time. Thus, a retrieval mode applies only to episodic memory tasks and not to other unaware uses of memory such as performing a skilled task or other automated activity. Whether episodic retrieval, guided through a retrieval mode, will be successful or not depends on the ecphoric synergy created by the retrieval cues and their interaction with the memory trace itself (i.e., transfer appropriate processing; Rajaram, Srinivas, & Roediger, 1998; Tulving, 1976). Different kinds of memory tests will have different requirements as in, say, comparing free recall to item recognition performance. Over the years, such coarse changes in the retrieval mode have been studied as differing levels of performance across a variety of different types of memory tests. By contrast, the topic of this article concerns finer grained changes in retrieval modes that have a consequential impact on subsequent recognition memory performance.

Jacoby and his colleagues have argued for a theory of cognitive control over memory retrieval which involves two basic sets of processes: late correction and early selection (e.g., Jacoby, Kelley, & McElree, 1999). Late correction involves automatically evaluating a stimulus and making a

* Corresponding author. Address: Department of Psychology, University of Georgia, Athens, GA 30602-3013, USA. Fax: +1 706 542 3275.
E-mail address: rlmarsh@uga.edu (R.L. Marsh).

0749-596X/$ - see front matter © 2009 Elsevier Inc. All rights reserved.
doi:10.1016/j.jml.2009.06.005

memory decision late in the process of retrieval. Because we will be using recognition paradigms in this article, an example of late correction would be automatically evaluating a copy cue and deciding late in the process that not enough evidence was retrieved to label the item as having been studied earlier. By contrast, early-selection processes change the way memory is examined in the first place. For example, the retrieval mode created by an item recognition test may not elicit the same level of detail from memory as attempting to render a source judgment on the very same item. Source tests tend to require more differentiated input (Johnson, Hashtroudi, & Lindsay, 1993). Thus, the retrieval modes of these two episodic memory tasks are likely to elicit different kinds of information from virtually the moment retrieval begins. Jacoby and his colleagues have recently capitalized on a paradigm of testing memory twice to demonstrate that early-selection processes during a first memory test can have consequences for distractor items inserted into a subsequent memory test that targets them as the old items.

To elaborate, during an encoding phase, a levels of processing manipulation is administered that makes some items deeply studied and other items shallowly studied (Jacoby, Shimizu, Velanova, & Rhodes, 2005b; Jacoby et al., 2005a; Shimizu & Jacoby, 2005). In a first test, distractor items are tested along with only deeply studied items; and different distractor items are tested with only shallowly studied items. Because participants are informed which class of items they are inspecting, the two retrieval modes should be different for the deep versus the shallow tests. The consequences of this change in retrieval modes can then be observed by placing all of the old distractors from both of the first tests with new distractors in a single, final recognition test for which participants are told to claim anything as old if it occurred anywhere in the experiment. In this final test, Jacoby and his colleagues found that more of the old distractors from the deep test were claimed to be old than the distractors from the shallow test. So, during the first recognition test, the consequence of evaluating distractors for deep characteristics was better memory for those distractors on a final recognition test. The cognitive control mechanism of early selection must have been guiding retrieval for qualitative characteristics that were most consistent with deeply encoded items during the first test, and this change in retrieval mode ultimately affected final recognition performance for the distractors evaluated in this fashion. Even though Jacoby and his colleagues were interested in the differential evaluation of the distractor items, the retrieval context may also alter the manner in which studied items are phenomenologically inspected on an initial test (Bodner & Lindsay, 2003; McCabe & Balota, 2007).

A similar finding has been reported recently in the cognitive neuroscience literature under the label ‘‘retrieval orientation” rather than early selection (e.g., Herron & Rugg, 2003; Hornberger, Rugg, & Henson, 2006; Rugg & Wilding, 2000). In these studies, participants might be tested on their memory for pictures or auditory verbal labels of the pictures in separate between-subjects conditions. Thus, what was studied at encoding was semantically equivalent, only the modality was different. During the item recognition tests, the neuropsychological correlates associated with processing new items were different for the two types of tests. Presumably, participants who studied pictures were examining memory for pictorial details whereas those who encoded words were examining memory for acoustic or lexical representations; and these different retrieval processes and criteria were observed in the way that the distractors were evaluated. Whether one chooses to label this effect early selection or a change in retrieval orientation really does not matter; the overarching point is that there exist much finer grained changes in retrieval modes that translate to observable neuropsychological differences during those first tests and differences in subsequent recognition memory performance for the foils from those initial tests. At a neural level, these differences in subsequent recognition memory performance seem to arise from qualitatively different signatures of activation dependent on the initial evaluation of those foils.

For us, the current state of this literature suggests that different retrieval modes are guided by early-selection processes. A retrieval mode involves a mental agenda to evaluate test candidates for certain classes of information in much the same way that one would perform when taking a source test (e.g., Johnson et al., 1993). For this reason, Jacoby and his colleagues have labeled their findings source-constrained retrieval. In this sense, different retrieval modes set the evaluative criteria that are used for performing the memory task. These same criteria are applied to new items and inspecting them in different ways (due to being in different retrieval modes) leaves differences in their later memorability. One question that remains, however, is whether or not the consequences of being in a retrieval mode leaves the distractor items from the first test imbued with different levels of familiarity. Under this familiarity scenario, Jacoby and his colleagues found that distractors tested with deep items later feel more familiar on the final recognition test than those inspected for shallower attributes. Of course, this need not be the case because there could be an experiential basis for the better final recognition performance. Under this scenario, participants would be remembering qualitatively different information (e.g., idiosyncratic recollection) for distractors from the deep first test as compared with the shallow first test. Thus, a primary goal of the present study was to ascertain which of these two scenarios better matched the empirical data.

To further evaluate the difference in processing between a deep and shallow retrieval mode, we administered a Remember–Know recognition test as the final test in Experiment 1 (Gardiner, Gregg, Mashru, & Thaman, 2001; Gardiner, Ramponi, & Richardson-Klavehn, 2002). If the consequences of being in a deep retrieval mode yield qualitative differences in recollection, then more remember responses should be issued on the final test for rejected distractors from a deep retrieval mode as compared with the amount of recollection for rejected distractors tested in a shallow retrieval mode. By contrast, if the consequences of being in different retrieval modes are a function of differences in familiarity levels of the originally rejected distractors, then the know responses to those distractors should differ on the final test.[1] The distinction between a difference in familiarity and recollection is important because it could help elucidate the type of processing that differentiates a deep from a shallow retrieval mode. Another goal of the current study was to determine the degree of flexibility with which retrieval modes can be used. Based on previous findings, it is not clear whether or not people are willing to change their evaluative criteria during a single test, even using quantitative information (e.g., Morrell, Gaitan, & Wixted, 2002; Rhodes & Jacoby, 2007). In addition to using a Remember–Know procedure in Experiment 1, we manipulated the frequency with which the participants would have to switch their use of deep and shallow retrieval mode.

In Experiment 2, we asked participants to tell us how they were rejecting the distractors during the first tests in order to gain further insight into the nature of being in these two different retrieval modes. In Experiment 3, we changed the manner in which differential memory strength was achieved by replacing the orienting task with the number of times that study trials were repeated. That experiment assesses whether the semantic, qualitative orienting manipulation is crucial to establishing different retrieval modes or whether retrieval modes can be strength-based. In Experiment 4 we administered two-alternative forced-choice tests during the first tests in order to understand if those retrieval modes affect performance in the same way as an initial item recognition test. As a package, these experiments were designed to collectively understand how different retrieval modes affect the early-selection control mechanisms on those tests and to examine the memory performance that results from applying them.

Experiment 1

This first experiment was a conceptual replication of Jacoby et al.'s (2005a) first experiment. We made two major changes. First, as discussed earlier, the final test was not an item Old–New test, but rather was a Remember–Know–New test. We expected to replicate the result that retrieval processes operating during a deep retrieval mode yields better memory performance for distractors on subsequent tests of memory. We also wanted to know whether better performance on the subsequent test was a result of more recollection or more familiarity. Second, in a different condition we did not block testing of deep and shallow items during the first test, but rather, asked on an item by item basis whether or not the item was judged under the deep versus shallow orienting instruction. Our purpose for including that condition was to understand whether cognitive control and the early-selection processes were flexible enough to be initiated on a trial by trial basis or whether it would be too costly to switch constantly between the two retrieval modes. In this latter case, participants may give up and adopt some middling retrieval mode that is a compromise of the two that they established separately in Jacoby and his colleagues' past work.

Method

Participants

Undergraduate students from the University of Georgia volunteered in exchange for partial credit toward a research appreciation requirement. Each participant was tested individually in sessions that lasted approximately 30 min. Two participants who failed to comply with the experimental instructions were not retained in the final sample, yielding a sample size of 31 people in each condition.

Materials and procedure

We chose 180 medium frequency items from the Kucera and Francis (1967) word norms. The software controlling the experiment randomly chose (anew for each participant tested) 30 of these to serve as the deeply studied items and 30 more to serve as the shallowly studied items. Of the remaining 120 items, half were used as distractors during the first test and half were used as distractors on the final test. During encoding the two orienting tasks were randomly intermixed among the trials throughout the study phase. On the deep trials, a query appeared asking participants to rate the pleasantness of words on a seven-point scale. On shallow trials, the query asked them to count the number of letters in the word. Participants responded using the number keys across the top row of the keyboard. To maintain maximum attention, a warning tone and fixation point occurred for 250 ms prior to the word and the query appearing (this also occurred during both tests).

After the encoding phase, participants were engaged in a distractor task for 3 min. Following this they read instructions for the first-test phase. In the blocked condition, participants were informed that new items and deep items were intermixed or that new and shallow items were intermixed. The software counterbalanced which test occurred first based on whether the subject identification number was odd or even; however, all participants took both the shallow and the deep test. After one test was completed they received instructions for the next test. Any time instructions were delivered in any experiment in this report, participants first read the instructions and then the experimenter verbally reiterated them. Each of two tests consisted of the 30 items rated during encoding and 30 new distractor items. These 60 distractors would later become the target items in the final test. In the item by item condition, the first tests were not blocked. Rather, deep and shallow items were intermixed with the distractors and participants were queried on each trial as to whether or not they rated the item for pleasantness or counted the number of letters. When the items were new, half required a deep rating and half required a shallow rating. Participant's responses served to classify those items as

[1] Our use of the Remember-Know procedure should not be construed as our strict belief that a Remember response wholly reflects recollection and that a Know response wholly reflects familiarity as in a strict dual-process model of memory. However, we do believe that these responses can be used as a subjective guide in tracking relative amounts of recollection and familiarity and that these responses are useful indicants of rough changes in the underlying processes.

deep and shallow on the final test. Participants responded by using two labeled keys (i.e., the F and J keys).

When the first test had concluded, instructions for the final test were given. These instructions explained that participants should declare any item old that was experienced anywhere in the experiment even if it was a new item in the previous test phase. We gave detailed Remember–Know instructions as in all of our other uses of the Remember–Know procedure (e.g., Cook, Marsh, & Hicks, 2006). These instructions took about 5 min to deliver and reexplain; and the experimenter did not advance to the test phase unless he or she was sure that the participant understood the task. Participants again used the home keys for remember and know judgments, and used the spacebar to declare items new. Thus, it was a simultaneous, three-response option judgment during this final test. This final test consisted of 60 old items (the deep and shallow tested distractors from the first test) plus 60 brand new distractors that had not been experienced in the course of the experiment.

Results and discussion

Unless explicitly labeled with *p* value, no statistical test in this experiment and those that follow had a probability of a Type I error greater than 5%. Performance on the first test is summarized in Table 1 and performance on the final test is summarized in Table 2. For the sake of expedience, we do not report all statistical analyses on the data from the first test. As the reader may anticipate from over 30 years of work on levels of processing manipulations, we obtained statistically reliable mirror effects on the first test: relative to the shallow manipulation, deep encoding increased the hit rate and decreased the false alarm rate to distractors. However, two potentially surprising results should be addressed. The reaction time to reject new items was slower in the item by item condition as compared with the blocked condition for both shallow items, $t(59) = 5.62$, as well as deep items, $t(60) = 7.71$ This result likely reflects two components that are unique to the item by item condition. First, participants had to read a new query on every test trial. Second, the type of information used to evaluate each word (i.e., the retrieval mode) changed throughout the test. Both of these components could have resulted in longer response times in the item by item condition. Another interesting finding was that the false alarm rate was higher for the shallowly studied items in the item by item condition as compared with the blocked condition, $t(60) = 2.67$. This result could have been a product of shallow items being tested with the deep items on the same test. This unique feature of the item by item condition might have changed the manner in which participants inspected the new items with a shallow query. Nevertheless, this result could have been a function of other factors.

Of more interest were the consequences of being placed in the two different retrieval modes during the first test as expressed by performance on the final test summarized in Table 2. We analyzed the hit rates from the final test using a 2 (condition) $\times$ 2 (depth of processing) mixed model Analysis of Variance (ANOVA). The hit rate was calculated as the sum of the items claimed remember or know. The distractors tested with the deep items were learned better than those tested with the shallow items, $F(1, 60) = 27.49$, $\eta^2 = .313$. Each of the simple effects was statistically significant as well, $t(30) = 4.18$ for the blocked condition, and $t(30) = 3.27$ for the item by item condition. There was neither a condition nor an interaction effect which suggests that the effect of retrieval mode was just as large in the item by item condition as in the blocked condition. The false alarm rate was calculated as the sum of the new items on the final recognition task claimed remember or know. Because the false alarm rates were the same across the two between-subjects conditions, the analyses on the hit rates will approximate a model conducted on two-high threshold corrected recognition. The interested reader may wish to know that of the false alarm rates reported in Table 2, the proportion labeled remembered was .08 and .05 for the blocked and item by item condition, respectively; and thus, participants seem to have understood the Remember–Know procedure accurately. Clearly, we have replicated the effects that Jacoby and his colleagues reported on the consequences of processing information under qualitatively different retrieval modes (e.g., Shimizu & Jacoby, 2005).

Of the hit rate, we report the proportion that was labeled remember thereby making the complement of this measure the proportion of the hit rate labeled know. In an identical ANOVA model on the proportion of hits given a remember response, more recollection was obtained for the distractors studied with the deep retrieval mode as

Table 1
Hits and false alarm rates with latencies to correct rejections in the first test of Experiments 1–4.

Experiment and condition	Hit rates				False alarm rates				Latencies to correct rejections			
	Deep		Shallow		Deep		Shallow		Deep		Shallow	
	M	SE	M	SE	M	SE	M	SE	M	SE	M	SE
Experiment 1												
Blocked	.87	.02	.67	.02	.11	.02	.18	.02	1019	48	1062	50
Item by item	.87	.02	.70	.03	.09	.02	.32	.05	1775	85	1914	145
Experiment 2	.90	.02	.69	.02	.11	.01	.22	.02	1695	78	1710	72
Experiment 3	.87	.02	.70	.03	.09	.02	.17	.02	901	42	961	42
Experiment 4												
Query	.93	.02	.82	.02	Are complements to the hit rate in 2AFC recognition				2091	103	2638	136
No query	.94	.02	.83	.02					1568	75	2135	125

Table 2
Hits and false alarm rates with the proportion of hits labeled remember in the final test of Experiments 1–4.

Experiment and condition	Hit rates				Proportion hit labeled remember					
	Deep		Shallow		Deep		Shallow		False alarm rates	
	M	SE	M	SE	M	SE	M	SE	M	SE
Experiment 1										
Blocked	.82	.02	.76	.02	.53	.03	.42	.03	.28	.02
Item by item	.81	.02	.76	.03	.51	.04	.43	.04	.27	.03
Experiment 2	.87	.01	.83	.02	.55	.03	.48	.03	.27	.02
Experiment 3	.77	.02	.78	.03	.37	.04	.37	.04	.23	.03
Experiment 4										
Query	.63	.03	.68	.03	.35	.03	.39	.04	.27	.04
No query	.60	.03	.73	.02	.40	.04	.45	.04	.28	.02

compared with the shallow retrieval mode, $F(1, 60) = 26.11$, $\eta^2 = .301$. No other terms in this analysis were statistically significant. The effect of prior retrieval mode was found in each of the simple effects as well, for the blocked condition, $t(30) = 3.77$, for the item by item condition, $t(30) = 3.47$. Therefore, when distractors were examined with the deep retrieval mode they were imbued with more recollective details as compared with when they were examined with the shallower retrieval mode. We would argue from these data that a deeper retrieval mode confers its advantage in information subjectively expressed as recollection as opposed to familiarity.[2] We now draw the reader's attention back to Table 1 where the latencies to reject distractors as new on the first test are reported. There was no statistical difference to reject the distractors as new between the two retrieval modes, and if anything there is a nominal difference favoring the deep retrieval mode being faster (see Jacoby et al., 2005b for a consistent finding). Consequently, the reader cannot attribute the difference in final test performance to time on task during the first test. Rather, the differences must arise from the attributes of memory that participants are inspecting under the two different modes of cognitive control.

Experiment 2

To gain further insight on how the early-selection retrieval processes operate to achieve the greater recollection on the final test, in Experiment 2, we asked participants to provide us with a rationale for rejecting distractors on the first test. There are three basic ways to reject distractors as new (Gallo, Weiss, & Schacter, 2004; Ghetti, 2003; Starns, Cook, Hicks, & Marsh, 2006). First, a test candidate can lack sufficient familiarity to be called old (i.e., it does not feel old). Second, one can have a negative memory insofar as one feels that the item would be too personally or idiosyncratically salient to have been studied and not remembered (e.g., a pet's name would have been remembered; Brown, Lewis, & Monk, 1977). Third, and finally, one could use a recall to reject strategy if there was a lawful relation between studied items and distractors. However, only the first two strategies are relevant to the current manipulations. As a consequence, when participants declared an item new during the first test, they had to answer a binary question about why they were calling the item new. They responded that either it lacked familiarity or was an item they would have remembered. Given the finding that a deep retrieval mode imbues items with more recollective details, we hypothesized that more often participants would say that an item would have been remembered under a deep retrieval mode as compared with a shallow one. In all other respects, this experiment was identical to the blocked condition in Experiment 1.

Method

Participants

Undergraduates from the University of Georgia volunteered in exchange for partial credit toward a research appreciation requirement. Each participant was tested individually in sessions that lasted approximately 30 min. A total of 35 participants were tested but one was removed for not complying with the experimental instructions.

Materials and procedure

The procedure was identical in all respects to the blocked condition of Experiment 1 except for the introspective judgment made during the first test. In that phase, when participants declared that they had not studied an item, they were required to inform us why they felt that the item was new. The instructions explained the distinction between an item lacking familiarity and an item that would have been so salient that participants probably would not forget having studied it. Examples of each judgment were given, and like all other instructions, the experimenter required that the participant be able to verbally repeat back in their own words when each response option should be used.

[2] The same conclusions would be drawn from the raw and unconditionalized remember and know claims. We also analyzed familiarity under the assumption that recollection and familiarity are independent processes. Under this assumption, familiarity is calculated by dividing the amount of know claims by 1 minus the amount of remember claims (know/[1-remember]). These results demonstrated nominal increases in familiarity for the deep as compared with the shallow foils, but neither the main effect of depth nor the simple effect in the blocked and item by item condition ($ps > .10$) were statistically significant. This outcome is consistent with the notion that a deep retrieval mode primarily exerts its benefit through increased recollection and not increased familiarity of the foils. These assertions are true of the subsequent experiment as well.

Table 3
Rejection strategies applied to the distractor items correctly labeled new during the first test in Experiment 2.

	Would have remembered				Lacks familiarity			
	Deep		Shallow		Deep		Shallow	
	M	SE	M	SE	M	SE	M	SE
Raw claims	.37	.03	.28	.03	.52	.03	.50	.03
Subsequent hits	.93	.03	.88	.02	.87	.01	.82	.02
Subsequent Rs	.53	.06	.47	.04	.48	.03	.39	.03

Results and discussion

As is clear from the results in Table 1, we replicated the mirror effect in first-test performance. Regarding the final test data, we first tested the hit rates (see Table 2). Indeed, the deep retrieval constraint again lead to better recognition memory on the final test as compared with the shallow retrieval constraint, $t(33) = 2.58$. The proportion of the hit rate that was labeled remembered was again greater under the deep retrieval constraint as compared with the shallow constraint, $t(33) = 2.65$. Finally, the false alarm rate in this experiment is comparable to that obtained in the first experiment (with .06 being labeled remembered). Therefore, in every respect this experiment replicated the blocked condition of Experiment 1. The novel aspect concerned participant's claims when they correctly declared a distractor item new during the first test. These data are summarized in Table 3 in which the compliment of the false alarm rates from Table 1 will be the sum of the two claims (lacks familiarity and would have remembered). We conducted a 2 (constraint: deep versus shallow) × 2 (claim: would have remembered versus lacks familiarity) repeated measures ANOVA on the proportion of correctly rejected distractors as a function of which rejection strategy was chosen for each item. Not surprisingly, participants more often claimed to base their rejection strategies on lacking familiarity, $F(1, 33) = 11.96$, $\eta^2 = .272$. However, the interaction was statistically significant, $F(1, 33) = 4.58$, $\eta^2 = .125$ indicating that under the deep retrieval mode, participants more often claimed that they would have remembered the item as compared with the shallow retrieval mode, simple effect: $t(33) = 3.43$.

We also examined final recognition performance as a function of the rejection claims given to items in the first test (i.e., conditionalized on claims) as well as the type of constraint for items in the initial testing session.[3] First we conducted a 2 (constraint: deep versus shallow) × 2 (claim: would have remembered versus lacks familiarity) repeated measures ANOVA on the foil hit rate in the final test. As summarized in the second row of Table 3 (labeled subsequent hits), the deep constraint on the first test led to more hits on the final test, replicating the analysis reported previously, $F(1, 29) = 6.13$, $\eta^2 = .175$. However, claims of would have remembered also led to more hits in the final test, $F(1, 29) = 12.29$, $\eta^2 = .298$. We also conducted an identical ANOVA on the proportion of tested foils given a Remember response in the final test (summarized in the third row of Table 3). Examining the proportion of each type of judgment rendered during the first test, would have remembered judgments led to more later reports of recollection during the final test, $F(1, 24) = 5.29$, $\eta^2 = .181$. Thus, for distractor items receiving a would have remembered judgment, better memory on the final test was found. The reader should note that a previous analysis demonstrated that being in a deep retrieval mode led to more reports of would have remembered judgments as compared with a shallow retrieval mode. As in Experiment 1, the latencies on the first test reported in Table 1 are not statistically different, but being in the deep retrieval mode again had people responding nominally faster; thus, time on task does not explain the results (Jacoby et al., 2005b).

The outcomes of this experiment do illuminate how the retrieval constraints are operating. Under the deep retrieval mode, participants must be examining their memories for evidence of the pleasantness rating task and their idiosyncratic reactions to the concepts studied. Not finding memory for those details, they more often evaluate that they would have remembered studying those items had they been studied with a pleasantness judgment. Consequently, the difference in the two retrieval modes appears to be more than a quantitative difference in the memorial evidence required to perform under the two different retrieval modes; rather, there appears to be a qualitative difference in the type of memorial evidence being evaluated that impacts performance in the final test.

Experiment 3

One reason that Jacoby and his colleagues (and us in Experiments 1 and 2) may have obtained these results is owing to the very nature of the depth of processing task. The orienting tasks focus people on very different aspects of the words: semantic meaning in one case and orthographic structure in the other. The same is true in the neuropsychological literature on retrieval orientations where different modalities (also see Jacoby et al., 2005a) and different classes of items are studied (e.g., pictures versus auditory words). The different retrieval modes can then capitalize on these different features to better focus memorial investigation of the candidates for just those features that are likely to be diagnostic. In Experiments 1 and 2 herein that there were likely to be idiosyncratic characteristics laid down when processing distractor items in

[3] The reader should note that the degrees of freedom are reduced when a participant did not contribute to all cells of interest. Also, the nature of these results are potentially limited in interpretation given the temporal distance between the two recognition judgments and also by the fact that the processing on the second judgment of a particular item cannot be completely separated from the processing on the same item on the first test. Regardless, we believe that these data are still informative.

the deep retrieval mode, and more familiarity-based criteria being used to reject distractors in the shallow retrieval mode. If this is true, then we should be able to remove the significant differences in recollection on the final test by changing the nature of the memory strength manipulation. To achieve this result, we replaced the deep and shallow orienting tasks in Experiments 1 and 2 with a repetition manipulation (Jacoby, 1999; Jacoby, Jones, & Dolan, 1998). As such our goal was to produce a mirror effect on the first test that would not produce different retrieval modes during the first tests. To do so, we presented some items one time during study and other items were presented three times. We reasoned that there should not a single, salient difference between the once and thrice presented items. Thus, there should be no qualitative feature to distinguish the type of information used in each of the separate initial testing sessions. We therefore predicted no differential memory for the foils despite testing all of the thrice presented items together and all the once presented items together thereby strongly encouraging that the retrieval mode be changed for each of the first tests.

Method

Participants

University of Georgia undergraduates volunteered in exchange for partial credit toward a research appreciation requirement. Each of the 30 participants was tested individually in sessions that lasted approximately 30 min.

Materials and procedure

Once again we used the blocked procedure of Experiment 1 for the first test. During encoding, 30 items were randomly seen once (weak) and another 30 items were seen three times (strong). The first tests were counterbalanced for weak and strong items depending on the participant identification number. The other type of test then followed. Thus, participants knew whether each test was testing the thrice presented versus the weak once presented items. The composition and procedure for the final test was identical to that described in Experiment 1 and also used in Experiment 2.

Results and discussion

In Tables 1 and 2 the thrice presented items have been labeled *deep* and the once presented items have been labeled *shallow*. This has been done for simplicity's sake even though this manipulation is not truly one of deep and shallow processing. As the reader can see in Table 1 we replicated Experiments 1 and 2 by obtaining the expected mirror effect in hits and false alarms on the first test. On the final test, however, we found no evidence of retrieval mode differences. The hit rates for the distractors tested with thrice presented items were recognized no more often than those tested with the once presented items, $t(29) < 1$, ns. The proportion of the hit rate that was given a remember response was the same for both retrieval modes, $t(29) < 1$, ns. We do not need to appeal to issues concerning statistical power because the exact same design was used in the two previous experiments with the exception of how items were made stronger versus weaker.

Our goal in this experiment was to demonstrate that retrieval modes are not a consequence of obtaining memory strength differences and simply testing strong items and weak items separately. The cognitive control mechanisms that lead to differences in retrieval modes need a diagnostic feature or quality to inspect (Hornberger et al., 2006). Manipulating numbers of presentations is likely to have given participants different quantitative amounts of whatever qualitative dimension they were looking for in old and new items (i.e., strength or familiarity), but it does not provide a qualitative difference that would serve to change the retrieval modes for different tests (i.e., recollection). This result is consistent with Jacoby et al.'s (2005b) finding that foil memory is better when it is associated with a deep, semantic retrieval mode even when the initial strength of the items (as measured by first-test performance) is equated for deep and shallow orienting tasks. Both results strongly suggest that source-constrained retrieval effects are the result of examining memory for a dimension or qualitative characteristic that differs across the modes of retrieval.

Experiment 4

The previous three experiments have tested the consequences of placing participants in an episodic retrieval mode for item recognition. In this final experiment our goal was to reverse the effect of a deep retrieval mode usually leaving better memory for distractors. We hypothesized that using a two-alternative forced-choice test during the first test would change the retrieval mode by making the recognition of a target and a distractor relative to one another. Because the deep items should be identified more easily in a two-alternative test, the distractor items paired with the studied items should get shorter shrift in terms of encoding as compared with distractors paired with shallower items. Distractors paired with shallower items should be considered more deeply during the first test because they differ in familiarity and recollection less than do the distractors that get paired with deep items. Reversing the retrieval mode differences demonstrated in the first two experiments extends eliminating the effect in the third experiment by showing that participants flexibly adopt retrieval strategies that optimize performance.

Because a two-alternative judgment requires participants to choose between a new and old item, we predicted that the relative judgment would reverse the effect even when participants knew they were choosing between a deep and new item versus between a shallow and a new item (i.e., the item by item condition in Experiment 1). Therefore, we tested two conditions in this experiment, one in which we queried participants to choose the deep item from a deep-new pair or queried them to choose the shallow item from a shallow-new pair. Performance in that condition was compared to performance in a condition in which participants were simply asked to choose the old item in all pairs during the initial test. In the query condition we established conditions for a retrieval mode that allowed participants to look for specific details in both alternatives that may or may not impact performance when the recognition judgment is relative to the new item paired with it.

Method

Participants

University of Georgia undergraduates volunteered in exchange for partial credit toward a research appreciation requirement. We tested 31 participants in each of the query and no query conditions. One participant in the query condition was replaced due to failure to comply with the experimental instructions.

Materials and procedure

The essential details of this experiment followed the procedure of the item by item condition in Experiment 1. Participants were randomly assigned to either the query or no query condition. All participants studied a random sequence of words for which they judged the pleasantness or counted the number of letters in each word. Words were chosen randomly anew for each participant tested from the master pool of 180 words and there were 30 each in the deep and the shallow conditions. The first test was modified such that an old and a new item were paired with one another. Shallow and deep trials were intermixed and in the query condition where participants were correctly cued to pick the item for which they judged for pleasantness earlier or to pick the item for which they counted the number of letters earlier. In the no query condition, participants were simply asked to pick the studied item. The final test was identical to that reported in the previous three experiments insofar as we asked participants to identify the earlier experienced distractors during the first test from new distractors drawn from the remainder of the master word pool. They did so by reporting whether they remembered or just knew the item was experienced earlier using the response alternatives of remember, know, and new.

Results and discussion

With a two-alternative recognition test, the false alarm rate is the complement of the hit rate, so the reader should note that only the hit rates are available for report in Table 1.[4] As summarized there, participants were able to choose the deep item from a distractor more easily than they could choose a shallow item from a distractor, $F(1, 60) = 68.06$, $\eta^2 = .531$. In the 2 (constraint: deep versus shallow) $\times$ 2 (condition: query versus no query) the effect of depth was statistically significant during the final test, $F(1, 60) = 40.79$, $\eta^2 = .405$. Unlike in Experiments 1 and 2, the hit rate during the final test was higher for the distractors paired with shallow items thereby showing a reversal of the effect demonstrated by Jacoby and his colleagues. However, there was also a significant interaction between the two variables, $F(1, 60) = 7.98$, $\eta^2 = .117$. The retrieval mode instantiated by telling participants whether the alternative was studied under deep or shallow instructions minimized the difference between the deep and shallow constraints. This difference of 5% points in the query condition was still significant favoring distractors paired with shallow items, $t(30) = 2.59$, but the effect was smaller than the difference of 13% points in the no query condition, $t(30) = 6.35$. Because the false alarm rates during the final test were equivalent between groups, these analyses would be identical to those conducted on two-high threshold corrected recognition measures. That portion of the false alarm rate attributed to recollection was .07 and .05 for the query and no query conditions, respectively. Not surprisingly, the proportion of the hit rate claimed to be remembered was higher with distractors paired with the shallow items, $F(1, 60) = 4.01$, $\eta^2 = .063$. However, the effect is weak because neither of the simple effects for each condition reached conventional levels of significance, $1 < t(30) < 1.96$, ns. No other terms in the model were significant for claims of recollection.

The purpose of this final experiment was to reverse the effects of being placed in a deep versus shallow retrieval mode demonstrated in the first two experiments. We were successful in that regard because distractor items paired with deeply studied items were recognized less frequently and had fewer associated recollective details. The discrimination between a shallow item and a distractor resulted in better memory for the distractor. Unlike the previous three experiments, the difficulty in discrimination resulted in latency differences during the first test that were statistically significant (see Table 1). Participants were slower in rejecting the distractor when it was paired with a shallow item, $F(1, 60) = 68{,}09$, $\eta^2 = .532$, and they were slower when a query instantiated a retrieval mode, $F(1, 60) = 12.49$, $\eta^2 = 172$. Thus, discriminating the distractor from a shallowly studied item took longer which probably resulted in the better memory performance for these items. However, our goal in conducting this final experiment was to demonstrate that being in a deep retrieval mode does not necessarily increase subsequent mnemonic accuracy for information processed in that orientation. Whether or not a retrieval mode results in better or worse performance on a subsequent memory test is going to be a function of how cognitive control of memory directs attention toward or away from the diagnostic features of an encoding phase.

General discussion

The goal of this study was to understand better how cognitive control of recognition memory operates. We took as our starting point Jacoby et al.'s (2005a, 2005b) paradigm of first constraining retrieval during a memory test and then examined its consequences for unstudied items in a later Remember–Know recognition test. We discovered that when people inspect memory for more diagnostic information (i.e., a deep retrieval mode) that retrieval

[4] Technically, if we had been using signal detection measures, we would have calculated the hit rate as choosing the left alternative when it was old and the false alarm rate as choosing the left alternative when it was new. In this way, the hits and false alarms would not be linearly dependent as in this report. However, for the current purposes, our approach should be adequate.

processes and stimulus evaluation leaves more recollective details in the unstudied items. This finding suggests that the deep retrieval mode in that experiment led to a qualitative evaluation of the distractor items. We also discovered that retrieval modes can be used flexibly and dynamically on a trial by trial basis. Heretofore, a retrieval mode has been characterized as a more or less constant mental agenda that one is either in or not in at particular point in time (Lepage, Ghaffar, Nyberg, & Tulving, 2000). Craik, Govoni, Naveh-Benjamin, and Anderson (1996) have even gone as far as to claim that being in a retrieval mode protects recognition memory from the deleterious effects of divided attention (but see, Hicks & Marsh, 2000; Lozito & Mulligan, 2006). In contrast to such a monolithic view of what constitutes a retrieval mode, participants switched back and forth easily between two different kinds of source-constrained retrievals on a trial by trial basis. Thus, one nontrivial contribution of this study is to alert memory theorists that retrieval modes are dynamic and flexible states of cognitive control (see Rhodes & Jacoby, 2007 for a related finding).

We were also able to extend Jacoby et al.'s (2005a, 2005b) work by examining people's impressions at the moment that they came to the conclusion that a distractor was new during the first recognition test. We found that when placed in a retrieval mode for deeply studied items, people more often claimed that the distractor was new because they otherwise would have remembered it. In essence, the participants are indicating (albeit introspectively) that they are using different evaluative criteria in the deep mode as compared with the shallow mode. That outcome dovetails nicely with the neuropsychological work on retrieval orientations cited in the introduction. In that work, the neuropsychological signature of rejecting a new item during what is essentially the first test in the current paradigm is different depending on the class of items being tested. So, there is convergence insofar as a different pattern of brain activity maps onto a difference in behavioral responses as tested here. Together the two sorts of results argue that recognition memory tests are not uniform tests of memory during which old items are merely discriminated from new items. Rather, cognitive control mechanisms place people in a variety of different retrieval modes depending on the nature of the diagnostic evidence available. Experiment 4 demonstrated that not all types of memory tests are conducive to finding retrieval mode effects, at least with the foil paradigm employed in the current study.

We have argued throughout this article that our conception of different retrieval modes is based on a qualitative characteristic (or perhaps sets of them) or specific attributes of information that maximally discriminates old from new items. Of course, this depiction of a retrieval mode borrows heavily from Johnson and her colleagues work on the mental agendas that are used to perform source monitoring (e.g., Johnson et al., 1993). In that work, a mental agenda is a heuristic or filter that sets the standard of evaluation in examining the evidence in a memory trace. Such standards should never be labeled as correct or incorrect, but rather are probably subject to bias from both what is retrieved from memory and any other concurrent cognitive activities that are taking place contemporaneously. For example, schemas will bias source monitoring judgments (Bayen, Nakamura, Dupuis, & Yang, 2000; Hicks & Cockman, 2003; Sherman & Bessenoff, 1999). When retrieving information that is semantically consistent with a source, there is a high probability of attributing that information to the schema-consistent source.

There is also the possibility that retrieval modes can be layered and/or staged. By this we mean that failing to find evidence of the most diagnostic characteristic, the search passes to looking for the next-best piece of evidence. Our use of the Remember–Know procedure in the final test raises the possibility of whether that paradigm fosters such a dual-staged retrieval mode. In a Remember–Know procedure, a premium is placed on recollected information over feelings of familiarity by the very nature of the way the instructions are delivered. For this reason, some authors are now reporting that fast recollection occurs with a slower familiarity process that follows (Dewhurst, Holmes, Brandt, & Dean, 2006; Hicks & Marsh, 2000, Footnote 2). By contrast, there is evidence that familiarity is available more quickly than recollection in a speeded test (e.g., Toth, 1996). The current results suggest that this may be a normal state of affairs (fast F, slow R), but that a different retrieval mode such as the Remember–Know procedure or Hicks and Marsh's (2000) simultaneous use of anagrams and seen items in a standard recognition test can induce a retrieval mode that favors first consulting recollection, especially in a nonspeeded test. From this perspective it may be pointless to debate which process is "faster" or which comes "first" because the answer to that question is determined by cognitive control mechanisms establishing a particular retrieval mode that is optimized to the circumstances of the test. According to our argument, recollection can be selected early or used for late correction. The particular use of these two strategies will depend on the nature of the retrieval mode selected to evaluate test items. If qualitative detail is available for early retrieval constraint, then recollection may be used as early-selection criteria. Otherwise, recollection may be utilized later on in the recognition decision process to evaluate the test item. Of course, recollection may serve both functions or neither function on a given test trial.

The field of recognition memory seems to have gotten itself stuck on the debate about single versus dual-process models. As Jacoby and his colleagues (2005a, 2005b) have demonstrated, there are still interesting questions to be asked of basic recognition processes other than debating whether or not one process is enough. More generally, the cognitive control of memory establishes the very criteria used to evaluate memorial information. But, very little work has appeared to date on exactly how criteria are set, and even less has appeared on how participants are setting early selection versus late correction modes of evaluation. The procedure we have adapted from Jacoby and his colleague's work has proven itself to be immensely amenable to answering some of these questions. For example, Experiments 3 and 4 clearly show that strength-based and relative strength judgments are not amenable to early retrieval constraint. Instead, a qualitative feature must be present in order to guide the recognition decision. It is possible, however, that some qualitative characteristics may be more useful as diagnostic criteria for a retrieval mode. For example, pleasantness ratings may result in more qualitative information being available to utilize in a retrieval mode as compared with another semantic orienting task, such as animacy judgments. In closing, we hope that others find this approach as appealing as we do because it can be used to profitably illuminate how decision criteria (both early selection and late correction) are established during a memory test.

Acknowledgments

We thank Tilden Repp, Ellie Kenemer, Brannen McDougald, Jing Feng, Melanie Woodward, and Candace Miller for their dedicated help in collecting the data.

References

Bayen, U. J., Nakamura, G. V., Dupuis, S. E., & Yang, C. (2000). The use of schematic knowledge about sources in source monitoring. *Memory & Cognition, 28*, 480–500.

Bodner, G. E., & Lindsay, D. S. (2003). Remembering and knowing in context. *Journal of Memory and Language, 48*, 563–580.

Brown, J., Lewis, V. J., & Monk, A. F. (1977). Memorability, word frequency and negative recognition. *Quarterly Journal of Experimental Psychology, 29*, 461–473.

Cook, G. I., Marsh, R. L., & Hicks, J. L. (2006). The role of recollection and familiarity in the context variability mirror effect. *Memory & Cognition, 34*, 240–250.

Craik, F. I. M., Govoni, R., Naveh-Benjamin, M., & Anderson, N. D. (1996). The effects of divided attention on encoding and retrieval processes in human memory. *Journal of Experimental Psychology: General, 125*, 159–180.

Dewhurst, S. A., Holmes, S. J., Brandt, K. R., & Dean, G. M. (2006). Measuring the speed of the conscious components of recognition memory: Remembering is faster than knowing. *Consciousness and Cognition, 15*, 147–162.

Gallo, D. A., Weiss, J. A., & Schacter, D. L. (2004). Reducing false recognition with criteria recollection tests: Distinctiveness heuristic versus criterion shifts. *Journal of Memory and Language, 51*, 473–493.

Gardiner, J. M., Gregg, V. H., Mashru, R., & Thaman, M. (2001). Impact of encoding depth on awareness of perceptual effects in recognition memory. *Memory & Cognition, 29*, 433–440.

Gardiner, J. M., Ramponi, C., & Richardson-Klavehn, A. (2002). Recognition memory and decision processes: A meta-analysis of remember, know, and guess responses. *Memory, 10*, 83–98.

Ghetti, S. (2003). Memory for nonoccurrences: The role of metacognition. *Journal of Memory and Language, 48*, 722–739.

Herron, J. E., & Rugg, M. D. (2003). Retrieval orientation and the control of recollection. *Journal of Cognitive Neuroscience, 15*, 843–854.

Hicks, J. L., & Cockman, D. W. (2003). The effect of general knowledge on source memory and decision processes. *Journal of Memory and Language, 48*, 489–501.

Hicks, J. L., & Marsh, R. L. (2000). Toward specifying the attentional demands of recognition memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 26*, 1483–1498.

Hornberger, M., Rugg, M. D., & Henson, R. N. (2006). fMRI correlates of retrieval orientation. *Neuropsychologia*.

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

Jacoby, L. L., Jones, T. C., & Dolan, P. O. (1998). Two effects of repetition: Support for a dual-process model of Know judgments and exclusion errors. *Psychonomic Bulletin & Review, 5*, 705–709.

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

Jacoby, L. L., Shimizu, Y., Daniels, K. A., & Rhodes, M. G. (2005a). Modes of cognitive control in recognition and source memory: Depth of retrieval. *Psychonomic Bulletin & Review, 12*, 852–857.

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

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

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

Lepage, M., Ghaffar, O., Nyberg, L., & Tulving, E. (2000). Prefrontal cortex and episodic memory retrieval mode. *Proceedings of the National Academies of Sciences, 97*, 506–511.

Lozito, J. P., & Mulligan, N. W. (2006). Exploring the role of attention during memory retrieval: Effects of semantic encoding and divided attention. *Memory & Cognition, 34*, 986–998.

McCabe, D. P., & Balota, D. A. (2007). Contexts effects on remembering and knowing: The expectancy heuristic. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 33*, 536–549.

Morrell, H. E. R., Gaitan, S., & Wixted, J. T. (2002). On the nature of the decision axis in signal-detection-based models of recognition memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 28*, 1095–1110.

Rajaram, S., Srinivas, K., & Roediger, H. L. III, (1998). A transfer-appropriate account of context effects in word fragment completion. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 24*, 993–1004.

Rhodes, M. G., & Jacoby, L. J. (2007). On the dynamic nature of response criterion in recognition memory: Effects of base rate, awareness, and feedback. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 33*, 305–320.

Rugg, M. D., & Wilding, E. L. (2000). Retrieval processing and episodic memory. *Trends in Cognitive Science, 4*, 108–115.

Sherman, J. W., & Bessenoff, G. R. (1999). Stereotypes as source-monitoring cues: On the interaction between episodic and semantic memory. *Psychological Science, 10*, 106–110.

Shimizu, Y., & Jacoby, L. L. (2005). Similarity-guided depth of retrieval: Constraining at the front end. *Canadian Journal of Experimental Psychology, 59*, 17–21.

Starns, J. J., Cook, G. I., Hicks, J. L., & Marsh, R. L. (2006). On rejecting emotional lures created by phonological neighborhood activation. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 32*, 847–853.

Toth, J. P. (1996). Conceptual automaticity in recognition memory: Levels of processing effects on familiarity. *Canadian Journal of Experimental Psychology, 50*, 123–138.

Tulving, E. (1976). Ecphoric processes in recall and recognition. In J. Brown (Ed.), *Recall and recognition* (pp. 37–73). London: Wiley.

Tulving, E. (1983). *Elements of episodic memory*. New York: Oxford University Press.