Estimating the frequency of nonevents: The role of recollection failure in false recognition

NORMAN R. BROWN, LORI BUCHANAN, and ROBERTO CABEZA
University of Alberta, Edmonton, Alberta, Canada

Participants studied lists of multiply presented converging associates (e.g., *bed, dream, pillow*, etc.) and were timed as they estimated how often they saw list items, related foils (e.g., *blanket*), and nonpresented critical items (SLEEP). Average number of repetitions (few [3] vs. many [6]) and repetition variability (fixed vs. variable) were manipulated between subjects. Participants responded more slowly to critical items (3.18 sec) than to list items (2.45 sec) or foils (2.22 sec). In addition, critical-item judgments of frequency (JOFs) were about as large as list-item JOFs, and false recognition (i.e., nonzero JOFs) of critical items was most likely in the few–fixed condition (96%) and least likely in the many–fixed condition (74%). These findings suggest that people can use *recollection failure*—the absence of an anticipated recollective experience, coupled with strong familiarity—to distinguish critical items from list items and that recollection failure is weighted most heavily when people expect familiar probes to access episodic information.

In the past few years, the Deese/Roediger–McDermott paradigm (DRM) has received much attention because it reliably produces robust "false memories" in the laboratory. In a typical DRM experiment, participants first study a list of words (e.g., *dream, blanket, pillow*, etc.) associated with a critical target word (e.g., SLEEP) and are then tested on their memory for list items and the critical word. In such experiments, critical items often appear as intrusions in recall tests and are often considered "old" in old/new recognition tests (for reviews, see Roediger, McDermott, & Robinson, 1998, and Schacter, Norman, & Koustaal, 1998). Although these findings suggest that critical items and list items might be experienced in much the same way at test, there is a growing body of literature indicating that this is not the case. For example, Mather, Henkel, and Johnson (1997) and Norman and Schacter (1997) report that even *false-recognized* critical items produce less detailed memories than do list items, and Lampinen, Neuschatz, and Payne (1999) have found that participants are less confident when they assign a source to a nonpresented critical item than when they assign one to a list item.

Recently, Schacter and his colleagues have suggested that participants who have been exposed to DRM lists might use differences in the way that critical items and list items are experienced at test to suppress false memories (Israel & Schacter, 1997; Schacter, Israel, & Racine, 1999; also see Dobbins, Kroll, Yonelinas, & Liu, 1998; Strack & Bless, 1994). On this view, a test item, even a very familiar one, can be rejected if it fails to evoke a sufficiently detailed episodic memory. We use the term *recollection failure* to refer to this absence of an anticipated recollective experience. As stated, this notion applies to situations in which participants expect a very detailed memory but access a relative vague one and also applies to situations in which participants expect to retrieve some episodic information but retrieve none.

In this article, we argue that recollection failure is compatible with dual-process recognition models and demonstrate that it plays a predictable role in the rejection of nonpresented critical items. Our approach has three key elements. First, we adopt a dual-process model of recognition (Jacoby & Dallas, 1981; Mandler, 1980). Typically, such models assume that recognition judgments are based on familiarity and recollection: Familiarity refers to the feeling that an item or event occurred in the past, though other information is unavailable; recollection refers to the retrieval of specific information about a prior event. Under ordinary conditions, participants indicate that they have recognized a test item when it is very familiar and/or when they have succeeded in retrieving details about the encoding episode; test items are rejected when they are unfamiliar and fail to cue the retrieval of contextual details. The second element to this approach is the assumption that recollection failure is treated as a memory cue and can serve as the basis for the correct rejection of highly familiar critical items (Israel & Schacter, 1997; Schacter et al., 1999). Finally, we assume that recognition judgments, like other sorts of decisions, are based on a weighted combination of the available cues and that a cue's weight reflects its diagnostic value (Brown & Siegler, 1993; Johnson, Hashtroudi, & Lindsay, 1993).

Of course, a dearth of recoverable episodic information is diagnostic only when detailed recollection is common. As a result, manipulations that increase the probability that familiar words are recollected (e.g., longer

study times, increased repetition, increased vividness, decreased study–test delay) should also increase the weight given to recollection failure. Thus, other things being equal, factors that foster the encoding and retrieval of episodic context for list items should decrease the false recognition of nonpresented critical items. Consistent with this prediction, increased study time, repetition of list items, and vividness of context lead to decreased false-alarm rates, as do reduced study–test delays (Arndt & Hirshman, 1998; Hall & Kozloff, 1970; Israel & Schacter, 1997; Payne & Elie, 1998; Payne, Elie, Blackwell, & Neuschatz, 1996; Seamon, Luo, & Gallo, 1998; Tussing & Greene, 1997, 1999).

The present study was designed to test specific predictions derived from the proposed dual-process, recollection-failure approach and also to provide information about familiarity levels associated with list items and nonpresented critical items. In this experiment, participants viewed a list of multiply presented converging associates and were timed as they estimated how often they had seen list items, semantically related foils, and nonpresented critical items. Note that, although several prior DRM studies have manipulated presentation frequency (e.g., Payne & Elie, 1998; Tussing & Greene, 1997, 1999), this is the first to report frequency judgments and hence necessarily the first to provide data on the time required to make these judgments. In addition, the present study is the first to manipulate presentation frequency within lists as well as between them (see below).

There were several reasons why this task was well suited for studying false-memory phenomena. First, like old/new recognition judgments, judgments of frequency (JOFs) indicate whether the participants believe that the word was previously presented; a JOF of 0 is equivalent to a "new" response, and a JOF greater than 0 is equivalent to an "old" response. This allows us to determine how common false recognition is following exposure to a set of multiply presented converging associates.

Second, there is good evidence that nonzero JOFs mainly reflect familiarity, at least when the to-be-estimated items received uniform processing during study (Hintzman & Curran, 1994, 1995; Hintzman, Curran, & Oppy, 1992). A thorough review of this evidence is beyond the scope of this report. Nonetheless, it is important to note that JOFs are often very accurate (Hasher & Zacks, 1979) and that many findings in the frequency judgment literature have been successfully simulated by MINERVA2, a memory model that uses echo strength, a familiarity-like measure, as the sole indicator of event frequency (Hintzman, 1988). These facts suggest that JOFs can provide a valid index of familiarity and that the magnitude of the JOFs elicited by different types of stimuli can be used to assess claims concerning familiarity. In particular, if nonpresented critical items seem as familiar as list items, as is often assumed, then critical-item JOFs should be about as large as list-item JOFs.

A third reason for examining the DRM lists in the context of a frequency task is that response time (RT) patterns can be used to determine whether participants rely on a familiarity-based estimation strategy. Briefly, prior research has demonstrated that people sometimes retrieve and count relevant event instances prior to responding with a JOF and that they sometimes produce a JOF by mapping familiarity-based intuitions on to a response range (Brown, 1995, 1997; Conrad, Brown, & Cashman, 1998). When participants enumerate, RT increases steeply with presentation frequency. This is because instance retrieval requires a serial process, and it takes longer to retrieve many instances than it does to retrieve a few (Bousfield & Sedgewick, 1944). In contrast, reliance on a familiarity-based strategy produces a relative flat RT function.[1] This makes sense when one considers that people should be able to assess the familiarity of frequently presented items and rarely presented items in the same amount of time. As noted above, we have assumed that participants in this experiment would rely on familiarity to generate their JOFs. This assumption is based on prior research that has shown that familiarity-based estimates are common and enumeration-based estimates are rare when instances are presented in a uniform context (Brown, 1995). If this assumption is correct, the RT–frequency function should be very shallow.

Finally, it was possible to vary the composition of the study lists to determine whether conditions that promote recollection yield the predicted decrease in false-alarm rate for critical items. Specifically, we manipulated two variables: (1) the average number of repetitions (few vs. many), and (2) the variability of the presentation frequencies (varied vs. fixed). These variables were crossed, creating four types of study lists. In the *few–varied* condition, three associates were presented at Frequencies 1, 2, 4, and 5, and, in the *many–varied* condition, the same words were presented at Frequencies 2, 4, 8, and 10. In contrast, each converging associate was presented 3 times in the *few–fixed* lists and 6 times in the *many–fixed* lists.

We assumed that a durable or detailed episodic representation is not always stored in memory when an item is encountered but that each presentation affords the participant an additional opportunity to produce a memorable encoding. Thus, increasing the number of presentations should increase the likelihood of a detailed memory of at least one presentation (Dewhurst & Anderson, 1999; Jacoby, 1999; Jacoby, Jones, & Dolan, 1998). This implies that recollection should be more common in the many–fixed condition than in the few–fixed condition. If so, recollection failure should be given more weight following a many–fixed list than following a few–fixed list, and, as a result, participants in the many–fixed group should have more success in correctly rejecting nonpresented critical items than those in the few–fixed group.

To generate predictions for the varied-presentation conditions, we assumed that recollection increases with presentation frequency and that recollection failure is weighted more heavily when recollection is common than when it is not. We also assumed that the range of frequency employed in both varied conditions was large enough to en-

sure that recollection would be very likely for high-frequency items and very unlikely for low-frequency items. Under these assumptions, participants in the many–varied condition should encounter fewer contextually detailed memories than those in the many–fixed condition. In contrast, participants in the few–varied condition should encounter more of these detailed memories than those in the few–fixed condition. It follows that recollection failure should be given less weight in the many–varied condition than in the many–fixed condition and more weight in the few–varied condition than in the few–fixed condition. This pattern of weights would produce a repetitions × variability interaction; specifically, false-alarm rate for critical items should be lowest for the many–fixed group, highest for few–fixed group, and intermediate for the varied-presentation groups.

In summary, a frequency estimation task was used to study false memory for critical items. We expected that the RT function for list items would be flat, indicating that a familiarity-based process was used to generate frequency estimates. We also expected that critical-item JOFs and list-item JOFs would be about the same magnitude. This would provide evidence that participants experience a strong sense of familiarity when they encounter critical items. Finally, the dual-cue, recollection-failure account predicts that critical items should be falsely recognized most often in the few–fixed condition and least often in the many–fixed condition.

METHOD

Materials and Design

Study lists were created by crossing repetitions (few vs. many) and variability (fixed vs. varied) between subjects. Each participant studied 5 converging associates lists (COLD, ROUGH, SLEEP, CUP, WINDOW; Stadler, Roediger, & McDermott, 1999). Twelve words from each list were selected as list items, and 3 words (those listed 5th, 10th, and 15th) were selected as foils. Few-repetitions study lists consisted of 36 words, and many-repetitions lists consisted of 72. Each few–fixed list was divided into three 12-word blocks, and each many–fixed list was divided into six 12-word blocks. For each participant and list, list items were randomly assigned to position, with the constraints that each item appeared once per block and that no item appeared twice in a row.

For the varied lists, presentation frequency was determined by associative strength. Specifically, the 3 strongest associates appeared most often, the 3 weakest least often, and the 3 high-medium and 3 low-medium associates were assigned to intermediate levels of presentation frequency. Separate study lists were constructed for each participant. These lists were created such that repeated presentations were evenly distributed, and items assigned to different levels of presentation frequency were equally common in all parts of the list.

A unique 16-word (12 list items, 3 foils, and 1 critical item) test list was constructed for each study list. Each list was divided into three 5-word blocks. Each block contained 1 randomly selected foil and 1 randomly selected list item from each level of presentation frequency (associative strength). The nonpresented critical item was assigned to the middle block. Within a given block, position was random.

Procedure

During the study phase, items from one DRM list were visually presented for 2.0 sec each, followed by a 0.5-sec delay. Following the study phase, the participants worked on a series of maze problems for 2 min and were then instructed to initiate the test phase.

During the test phase, the participants initiated a trial by pressing their keyboard's "enter" key. This caused a test word to appear. The participants were instructed to decide how many times the word had appeared in the study list and to press the keyboard's space bar as soon as they had "a single numerical response in mind." When the space bar was pressed, a response field appeared, and the participants entered an estimate and then pressed the "enter" key. The test word, response field, and frequency estimate were then erased and replaced by a prompt to initiate the next trial. Each trial yielded two RT measures: The *decision-time* interval began with the presentation of the test word and ended when the participant pressed the space bar; the *total RT* interval was measured from the onset of the test word to the completion of the trial.

The WINDOW list was always presented first and served as practice. Latin squares were used to counterbalance the order of the remaining lists. Prior to receiving the first list, the participants in the few-repetitions groups were informed that each study list would be 36 words long, and, prior to the first test phase, they were informed that no test item had appeared more than 7 times on the previous list. The participants in the many-repetitions groups were informed that the study lists would be 72 words long and that no test word had appeared more than 14 times. All participants were informed that some of the test words had not been previously presented.

Participants

One hundred twenty-eight University of Alberta undergraduates received course credit for their participation in this experiment. Thirty-two were randomly assigned to each group. All were native speakers of English, and each was tested individually.

RESULTS

Response Time

In all analyses reported below, data from the first list were excluded. Because decision-time and total-time analyses produced the same pattern of results, only the former are reported.

Presentation frequency of list items was manipulated only in the few– and many–varied conditions. In these conditions, a mean RT was computed for each participant and each level of presentation frequency by averaging over the 12 items (3 items from 4 lists) that appeared at a given level. These means were submitted to a 2 (repetitions: few vs. many) × 4 (presentation frequency: low, 1 or 2; medium low, 2 or 4; medium high, 4 or 8; and high, 5 or 10) analysis of variance (ANOVA). RTs increased with presentation frequency [$F(3,186) = 5.10$, $MS_e = 0.20$, $p < .01$], but only slightly, but was unaffected by average number of repetitions ($F < 1.0$, for both the main effect of repetitions and the interaction). Averaging over groups, mean RTs for the 4 levels of presentation frequency were 2.42, 2.66, 2.57, and 2.70 sec. These small RT differences might reflect the fact that larger numbers are somewhat more difficult to access than small ones (Moyer & Dumais, 1978). However, the speed of the re-

Table 1
Mean Response Times With Standard Errors for Each Item Type and Group,
Averaged Over All Responses (Zero JOF and Nonzero JOF Responses),
Zero JOF Responses Only, and Nonzero JOF Responses Only

| | Few | | | | | | Many | | | | | |
| | Varied | | | Fixed | | | Varied | | | Fixed | | |
	List	Target	Foil	List	Target	Foil	List	Target	Foil	List	Target	Foil
All Responses												
M	2.55	3.54	2.25	2.28	2.70	2.15	2.63	3.24	2.23	2.34	3.25	2.23
SE	0.39	0.61	0.28	0.19	0.29	0.19	0.22	0.30	0.18	0.15	0.38	0.16
Zero JOF												
M	3.54	4.21	2.14	3.19	5.05	2.06	3.25	4.60	1.93	2.63	4.39	2.01
SE	0.61	0.99	0.12	0.59	0.97	0.10	0.39	0.87	0.07	0.39	0.70	0.09
Nonzero JOF												
M	2.25	3.40	2.74	2.25	2.55	2.40	2.61	3.00	3.49	2.33	2.89	3.15
SE	0.28	0.34	0.26	0.04	0.17	0.15	0.05	0.17	0.30	0.03	0.24	0.27

sponses and the shallowness of the function strongly suggest that the participants relied on a familiarity-based estimation strategy and rule out the possibility that a retrieve-and-count strategy played an important role in this task (Brown, 1995, 1997).

A second analysis was conducted to determine the relation between item type and RT. First, three mean RTs were computed for each participant: one for the 4 critical items, a second over the 12 foils, and a third for the 48 list items. These means were then submitted to a 3 (item type) $\times$ 2 (repetitions) $\times$ 2 (variability) ANOVA. Across conditions, the participants responded more slowly to nonpresented critical items ($M = 3.18$ sec) than to list items ($M = 2.45$ sec) or foils ($M = 2.22$ sec; see Table 1) [for item type, $F(2,248) = 32.72, MS_e = 1.77, p < .0001$]. All other main effects and interactions were nonsignificant. This analysis was followed by four one-way (item type) ANOVAs, one per group. In each, the item type effect was highly reliable ($p < .0001$), and least significant difference (LSD) tests indicated that critical-item RTs were slower than list-item and foil RTs.

The effect of item type on RT was not predicted. It is, however, consistent with the dual-cue, recollection-failure approach and with prior research (Tun, Wingfield, Rosen, & Blanchard, 1998). Specifically, we have assumed that critical items produce a strong sense of familiarity, which suggests that the target is "old," coupled with recollection failure, which suggests that it is "new." On this view, nonpresented critical items produce a "mnemonic Stroop effect," because participants must resolve a conflict produced by a contradictory configuration of cues. In contrast, participants can rapidly recognize list items (recollection and high familiarity) and reject foils (recollection failure and low familiarity), because, in both cases, the available cues converge on a single response.

Mean RTs were computed separately over all zero JOFs and nonzero JOFs. These data, which are presented in the bottom two rows of Table 1, are of theoretical interest because they provide the basis for rejecting a simple equal-variance signal-detection interpretation of the RT differences just reported (Hall, Sekuler, & Cushman,

1969; Miller & Wolford, 1999). This alternative account assumes that familiarity is the sole basis for an old/new recognition judgment, that RT is inversely related to distance from the old/new criterion, and that the familiarity distribution associated with critical items falls closer to this criterion than the distribution associated with list items. It follows that (1) list items should be faster than critical items when JOFs are greater than 0 and (2) critical items should be faster than list items when JOF equals 0. The data are inconsistent with the second prediction: Critical-item RTs were reliably slower than list-item RTs for both zero JOFs and nonzero JOFs. Collapsing over groups, mean RT for the zero JOFs was 4.46 sec for critical items and 3.20 sec for list items ($p < .01$, by a Wilcoxon test); comparable figures for the nonzero JOFs were 2.95 and 2.43 sec ($p < .0001$). Mean RTs were 2.03 sec when foils were correctly rejected and 2.91 sec when they were falsely recognized.

JOFs

Figure 1 displays response distributions for foils, critical items, and list items (for corresponding means, see Table 2). These distributions were formed by computing the percentage of items assigned to each possible JOF value for each item type in each condition.

The first thing to notice about these data is that the participants typically provided relatively large JOFs for critical items (for the few-presentations lists, $M = 2.51$; for the many-presentations lists, $M = 4.59$). On average, these critical-item JOFs were significantly smaller than list-item JOFs ($Ms = 3.00$ and 6.21, for the few- and many-repetitions conditions, respectively) and significantly larger than foil JOFs ($Ms = 0.35$ and 0.59, for the few- and many-repetitions conditions, respectively). These claims are based on two 3 (item type) $\times$ 2 (variability) ANOVAs, one for the few-repetitions groups and the other for the many-repetitions groups. In both cases, only the item type effect was reliable [for the few-repetitions group, $F(2,124) = 326.94, MS_e = 0.39, p < .0001$; for the many-repetitions groups, $F(2,124) = 255.52, MS_e = 2.10, p < .0001$]. Additional analyses
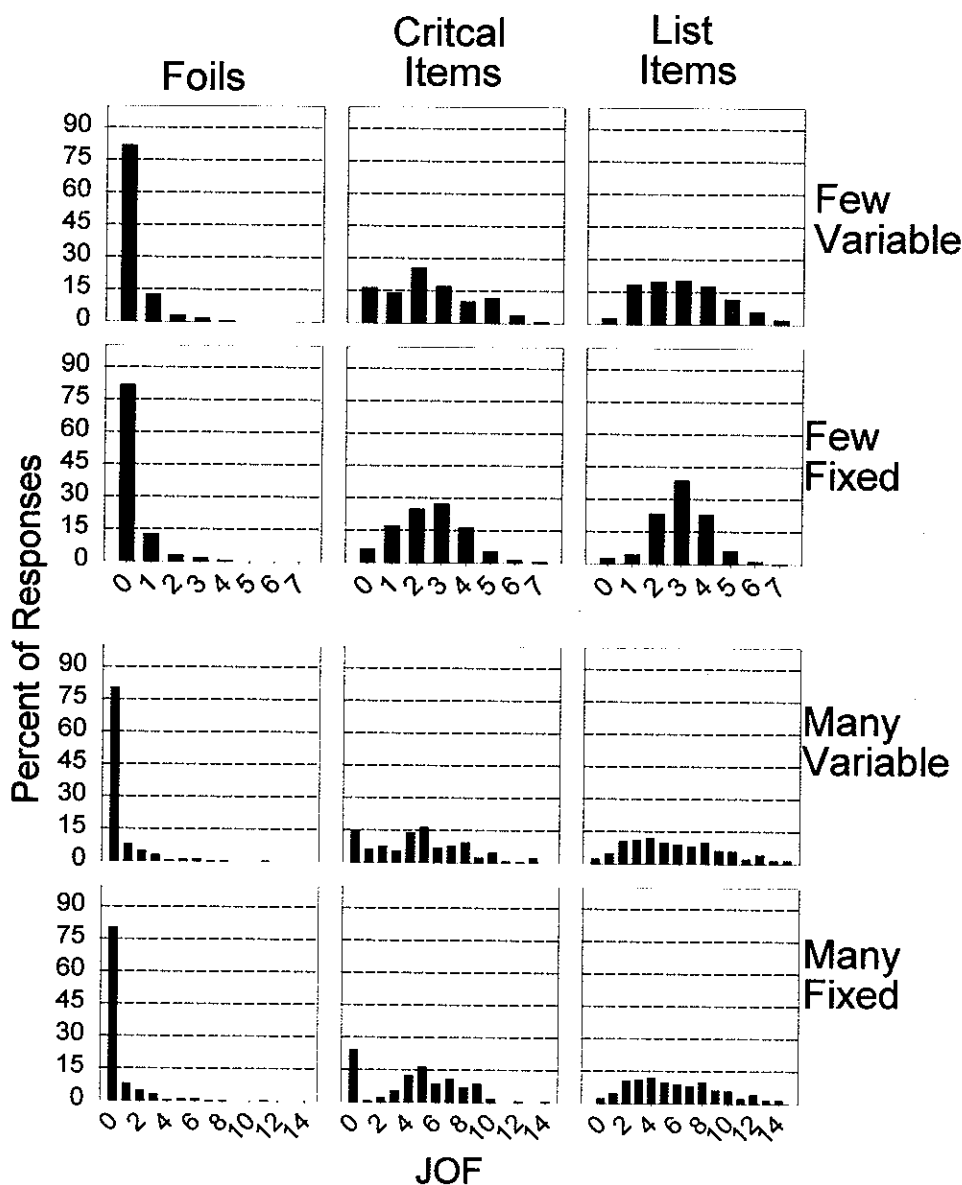
Figure 1. Percent of responses assigned to each JOF level for each item type and group.

were conducted for each group. In all cases, LSD tests indicated that the three item types produced reliably different JOFs.

When zero JOFs were excluded, estimates for critical items and list items were quite close. In the few–varied condition, mean JOFs were 2.93 for critical items and 3.12 for list items (n.s. by a Wilcoxon test). Comparable numbers were 2.74 and 3.07 in the few–fixed condition ($p <$ .001), 5.46 and 6.00 in the many–varied condition (n.s.), and 5.96 and 6.73 in the many–fixed condition ($p <$ .001).[2] In addition, response distributions produced by the two classes of items were quite similar (see Figure 1). This observation was confirmed by four correlations. In each, the 7 or 14 points representing nonzero JOFs in each critical-item panel were correlated with the corresponding points in the list-items panel; all correlations were statistically reliable and ranged from .81 in the many–varied condition to .87 in the many–fixed condition.

Assuming that nonzero JOFs provide an accurate index of familiarity, these data imply that nonpresented critical items were, on average, only slightly less familiar than the list items. Nonetheless, the participants were far less likely to indicate that they had studied the former (85% nonzero JOFs) than that they had studied the latter (98% nonzero JOFs); the participants failed to reject 21% of the foils (see Table 1). A 3 (item type) × 2 (repe-

Table 2
Mean JOFs and Percent Nonzero JOFs
With Standard Errors, for Each Item Type and Group

| | Few | | | | Many | | | | | |
| | Varied | | Fixed | | Varied | | Fixed | | Mean | SE |
	M	SE	M	SE	M	SE	M	SE		
					JOFs					
List items	3.03	0.08	2.99	0.07	5.85	0.23	6.57	0.24	4.61	0.17
Critical items	2.45	0.23	2.57	0.18	4.65	0.41	4.52	0.42	3.55	0.19
Foils	0.27	0.05	0.44	0.07	0.49	0.08	0.69	0.14	0.47	0.05
					% Nonzero JOFs					
List items	97.19	0.68	97.38	0.57	97.66	0.61	98.16	0.60	97.59	0.31
Critical items	83.59	3.99	93.75	2.51	85.16	4.33	75.78	5.78	84.57	2.20
Foils	18.19	3.26	26.03	4.10	19.44	2.48	19.47	3.28	20.78	1.67

titions) $\times$ 2 (variability) ANOVA carried out on percent of nonzero JOFs confirmed this observation. This analysis yielded a main effect for item type [$F(2,248)$ = 843.64, $MS_e = 0.02$, $p < .0001$]. More important, the three-way interaction was reliable [$F(2,248) = 3.11$, $MS_e = 0.02$, $p < .05$], as was the repetitions $\times$ variability interaction [$F(1,124) = 4.35$, $MS_e = 0.05$, $p < .05$]. Three follow-up 2 (repetitions) $\times$ 2 (variability) ANOVAs were performed, one for each item type. As data in Table 1 suggest, composition of the study list did not affect the percent of nonzero JOFs elicited by foils or list items; neither analysis produced a reliable effect. In contrast, the analysis performed on critical items yielded the predicted repetitions $\times$ variability interaction [$F(1,124) = 5.14$, $MS_e = 0.06$, $p < .05$]; the repetitions effect was also marginally significant [$F(1,124) = 3.62$, $MS_e = 0.06$, $p = .06$]. This interaction indicates that many–fixed participants were least likely to misrecognize critical items (76%), that few–fixed participants were most likely to misrecognize them (94%), and that the participants who studied the varied-frequency lists fell in between (84%).

Frequency Discrimination

For each participant and each test list, in the two varied-repetitions groups, the correlation between estimated and actual frequency was computed over the 12 list items. The mean of these correlations was .67 in the few–varied condition and .68 in the many–varied condition. This level of performance rules out the possibility that the participants randomly assigned nonzero JOFs to recognized words and, in so doing, provides additional support for the assumption that nonzero JOFs reflect different degrees of familiarity.

DISCUSSION

This study produced four main findings. First, in varied-frequency conditions, functions relating RT and presentation frequency were fast and relatively flat. This indicates that the participants relied on a familiarity-based process to generate their frequency estimates. Second, across the four conditions, list-item and critical-item JOFs resembled one another in magnitude and distribution. We take this to indicate that the former were only slightly more familiar than the latter. Third, percent of nonzero JOFs for critical items displayed the predicted repetitions $\times$ variability interaction, with false recognition most common in the few–fixed condition and least common in the many–fixed condition. This is consistent with the notion that people can use recollection failure to distinguish critical items from list items and that recollection failure is weighted most heavily when people expect familiar probes to access episodic information. Finally, the participants responded some 30% slower to critical items than to list items. The dual-cue, recollection-failure approach developed above provides a plausible explanation for this unexpected finding. On this account, critical items produce a mnemonic Stroop effect, because participants are confronted by a conflicting pair of cues, one (familiarity) implying that target is old and the other (recollection failure) implying that it is new. It is the effort required to resolve this conflict that slows responses to these items.

Like Hintzman and colleagues (Hintzman & Curran, 1994, 1995; Hintzman et al., 1992), we have assumed that the frequency judgments made by our participants involved two components: a recognition component and a component that maps familiarity-based intuitions onto a numerical response scale. The results of the present study, particularly the dissociative effect of list composition on RTs (no effect) and false-recognition rates (an interaction), suggest that the recognition component might be further decomposed into two subcomponents: one that assesses the match between familiarity and recollection and registers a problem when very familiar items fail to evoke episodic information, and a second that determines whether this mismatch provides sufficient grounds for rejecting a probe. Apparently, the participants in all conditions believed that familiarity and recollection should be related and were aware when they

were not. It also appears that their response to recollection failure was conditioned by the list-related tendency for familiar words to access episodic information.

This latter point raises an interesting issue. The present experiment differed from conventional DRM studies in a number of ways. Unlike most prior studies in the area, converging associates were present multiple times, participants made timed frequency judgments, lists were tested one at a time, and so on. It is possible that one of these differences was responsible for sensitizing people to recollection failure and facilitating its use as a negative recognition cue. However, it is also true that the dual-cue, recollection-failure approach provides a reasonable account for the effects of study time, test delay, and vividness on false recognition observed under more conventional conditions. This suggests that people may monitor recollection failure as a matter of course and rely on it when circumstances warrant. If so, the mnemonic Stroop effect may be a common feature of the DRM task. Clearly, future research is required to determine whether this is so.

REFERENCES

ARNDT, J., & HIRSHMAN, E. (1998). True and false recognition in MINERVA2: Explanations from a global matching perspective. *Journal of Memory & Language*, 39, 371-391.

BOUSFIELD, W. A., & SEDGEWICK, C. H. (1944). An analysis of sequences of restricted associative responses. *Journal of General Psychology*, 30, 149-165.

BROWN, N. R. (1995). Estimation strategies and the judgment of event frequency. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, 21, 1539-1553.

BROWN, N. R. (1997). Context memory and the selection of frequency estimation strategies. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, 23, 898-914.

BROWN, N. R., & SIEGLER, R. S. (1993). Metrics and mappings: A framework for understanding real-world quantitative estimation. *Psychological Review*, 100, 511-534.

CONRAD, F., BROWN, N. R., & CASHMAN, E. (1998). Strategies for answering behavioral frequency questions. *Memory*, 6, 339-366.

DEWHURST, S. A., & ANDERSON, S. J. (1999). Effects of exact and category repetition in true and false recognition memory. *Memory & Cognition*, 27, 665-673.

DOBBINS, I. G., KROLL, N. E. A., YONELINAS, A. P., & LIU, Q. (1998). Distinctiveness in recognition and free recall: The role of recollection in the rejection of the familiar. *Journal of Memory & Language*, 38, 381-400.

HALL, J. W., & KOZLOFF, E. E. (1970). False recognitions as a function of number of presentations. *American Journal of Psychology*, 83, 272-279.

HALL, J. W., SEKULER, R., & CUSHMAN, W. (1969). Effects of IAR occurrence during learning on response time during subsequent recognition. *Journal of Experimental Psychology*, 79, 39-42.

HASHER, L., & ZACKS, R. T. (1979). Automatic and effortful processes in memory. *Journal of Experimental Psychology: General*, 108, 356-388.

HINTZMAN, D. L. (1988). Judgments of frequency and recognition memory in a multi-trace memory model. *Psychological Review*, 95, 528-551.

HINTZMAN, D. L., & CURRAN, T. (1994). Retrieval dynamics of recognition and frequency judgments: Evidence for separate processes of retrieval and recall. *Journal of Memory & Language*, 33, 1-18.

HINTZMAN, D. L., & CURRAN, T. (1995).When encoding fails: Instructions, feedback, and registration without learning. *Memory & Cognition*, 23, 213-226.

HINTZMAN, D. L., CURRAN, T., & OPPY, B. (1992). Effects of similarity are repetition on memory: Registration without learning? *Journal of Experimental Psychology: Learning, Memory, & Cognition*, 18, 667-680.

ISRAEL, L., & SCHACTER, D. L. (1997). Pictorial encoding reduces false recognition of semantic associates. *Psychonomic Bulletin & Review*, 4, 577-581.

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., & DALLAS, M. (1981). On the relationship between autobiographical memory and perpetual learning. *Journal of Experimental Psychology: General*, 110, 306-340.

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.

JOHNSON, M. K., HASHTROUDI, S., & LINDSAY, D. S. (1993). Source monitoring. *Psychological Bulletin*, 114, 3-28.

LAMPINEN, J. M., NEUSCHATZ, J. S., & PAYNE, D. G. (1999). Source attributions and false memories: A test of the demand characteristics account. *Psychonomic Bulletin & Review*, 6, 130-135.

MANDLER, G. (1980). Recognizing: The judgement of previous occurrence. *Psychological Review*, 87, 252-271.

MATHER, M., HENKEL, L. A., & JOHNSON, M. K. (1997). Evaluating characteristics of false memories: Remember/know judgments and memory characteristics questionnaire compared. *Memory & Cognition*, 25, 826-837.

MILLER, M. B., & WOLFORD, G. L. (1999). Theoretical commentary: The role of criterion shift in false memory. *Psychological Review*, 106, 398-495.

MOYER, R. S., & DUMAIS, S. T. (1978). Mental comparison. In G. H. Bower (Ed.), *The psychology of learning and motivation* (Vol. 12, pp. 117-155). New York: Academic Press.

NORMAN, K. A., & SCHACTER, D. L. (1997). False recognition in younger and older adults: Exploring the characteristics of illusory memories. *Memory & Cognition*, 25, 838-848.

PAYNE, D. G., & ELIE, C. J. (1998, November). *Repeated list presentation reduces false memories for pictures and words*. Paper presented at the 39th Annual Meeting of the Psychonomic Society, Dallas.

PAYNE, D. G., ELIE, C. J., BLACKWELL, J. M., & NEUSCHATZ, J. S. (1996). Memory illusions: Recalling, recognizing, and recollecting events that never occurred. *Journal of Memory & Language*, 35, 261-285.

ROEDIGER, H. L., III, McDERMOTT, K. B., & ROBINSON, K. J. (1998). The role of associative processes in creating false memories. In M. A. Conway, S. E. Gathercole, & C. Cornoldi (Eds.), *Theories of memory II* (pp. 187-245). Hove, U.K.: Psychological Press.

SCHACTER, D. L., ISRAEL, L., & RACINE, C. (1999). Suppressing false recognition in younger and older adults: The distinctiveness heuristic. *Journal of Memory & Language*, 40, 1-24.

SCHACTER, D. L., NORMAN, K. A., & KOUSTAAL, W. (1998). The cognitive neuroscience of constructive memory. *Annual Review of Psychology*, 49, 289-318.

SEAMON, J. G., LUO, C. R., & GALLO, D. A. (1998). Creating false memories of words with or without recognition of list items—Evidence for nonconscious processes. *Psychological Science*, 9, 20-26.

STADLER, M. A., ROEDIGER, H. L., III, & McDERMOTT, K. B. (1999). Norms for word lists that create false memories. *Memory & Cognition*, 27, 494-500.

STRACK, F., & BLESS, H. (1994). Memory for nonoccurrence: Metacognitive and presuppositional strategies. *Journal of Memory & Language*, 33, 203-217.

TUN, P. A., WINGFIELD, A., ROSEN, M. I., & BLANCHARD, L. (1998). Response latencies for false memories: Gist-based processes in normal aging. *Psychology & Aging*, 13, 230-241.

TUSSING, A. A., & GREENE, R. L. (1997). False recognition of associates: How robust is the effect? *Psychonomic Bulletin & Review*, 4, 572-576.

TUSSING, A. A., & GREENE, R. L. (1999). Differential effects of repetition on true and false recognition. *Journal of Memory & Language*, 40, 520-533.

591   BROWN, BUCHANAN, AND CABEZA

NOTES

1. For example, in one experiment (Brown, 1995, Experiment 2), there was a condition known to foster enumeration, and there was a second condition that promoted the use of a familiarity-based estimation strategy. In the first condition, RTs increased monotonically from 3.5 to 10.5 sec, as presentation frequency increased from 2 to 16; in the second condition, RTs increased from 2.7 to 3.6 sec, over the same range of frequencies.

2. In all conditions, nonzero JOFs were significantly smaller for foils than for nonpresented critical items ($p < .0001$). Mean nonzero JOFs were 1.47, 1.68, 2.49, and 4.13 for the few–varied, few–fixed, many–varied, and many–fixed groups, respectively.

(Manuscript received December 16, 1998;
revision accepted for publication October, 25, 1999.)