MEMORY, 2010, 18 (4), 385–393

Psychology Press
Taylor & Francis Group

The effects of free recall testing on subsequent source memory

Gene A. Brewer, Richard L. Marsh, Joseph T. Meeks, and Arlo Clark-Foos
University of Georgia, Athens, GA, USA

Jason L. Hicks
Louisiana State University, Baton Rouge, LA, USA

The testing effect is the finding that prior retrieval of information from memory will result in better subsequent memory for that material. One explanation for these effects is that initial free recall testing increases the recollective details for tested information, which then becomes more available during a subsequent test phase. In three experiments we explored this hypothesis using a source-monitoring test phase after the initial free recall tests. We discovered that memory is differentially enhanced for certain recollective details depending on the nature of the free recall task. Thus further research needs to be conducted to specify how different kinds of memorial details are enhanced by free recall testing.

Keywords: Testing effect; Episodic memory.

Broadly construed, the testing effect is the empirical finding that information that has been tested earlier leads to more durable memory during a later testing phase (e.g., Roediger & Karpicke, 2006). Typically, the testing effect is shown on subsequent free and cued recall tests. Recently Chan and McDermott (2007) have proposed that initial free recall testing increases aspects of recollective experience. In one of Chan and McDermott's (2007) experiments, list discrimination was enhanced by intermediate free recall testing. According to the source-monitoring framework, temporal information and other recollective aspects of earlier experiences represent qualitative details that could be embellished in memory during an initial testing phase (Johnson, Hashtroudi, & Lindsey, 1993). The goal of the present work was to ascertain whether the testing effect would extend to other, less-salient characteristics that were present during the original encoding episode.

Free recall testing has been a standard memory examination technique for many years (for reviews see Crowder, 1976; Wixted & Rohrer, 1994). The free recall of an item from long-term memory acts as an additional encoding phase and potentially embellishes contextual information associated with the item (Bjork, 1975). Thus the act of recalling items from long-term memory modifies those items in some way, which leads to better subsequent memory for them as assessed by later memory tests. By this account, retrieval reactivates the recalled information in memory thereby making it more accessible. Consequently, later memory for that information should be better (i.e., the testing effect). At issue here is whether free recall processes establishes source-specifying characteristics in certain circumstances

Address correspondence to: Gene Brewer, Department of Psychology, University of Georgia, Athens, GA 30602-3013, USA. E-mail: gabrewer@uga.edu

We thank Michael Hopkins, Jason Kinney, and Hollis Johnson for their assistance in collecting the data.

http://www.psypress.com/memory DOI:10.1080/09658211003702163

that could subsequently lead to better source memory for the presentation modality of recalled information. But also at issue is whether there are other circumstances in which free recall may not benefit source memory for recalled information in the same manner. The role of free recall on memory for source information is poorly understood (but see Hicks & Marsh, 1999). Preliminary evidence for the beneficial role of free recall on source information for recalled items comes from dual-process accounts of the testing effect in recognition memory.

Our point of departure is that Chan and McDermott (2007) found an increase in list discrimination performance as well as an increase in Remember as compared with Know responses on final tests. They used a source-monitoring paradigm in which items from two lists were either tested or not tested by free recall (in between-participant conditions) prior to a final list-discrimination test. According to a dual-process account of intermediate testing on ultimate source discrimination, free recalling items from each list establishes additional information in memory that could be used later to specify which list that items had appeared on earlier (see Jang & Huber, 2008). On the one hand, additional temporal information from the recall test itself may be imbued in items as they are retrieved. However, other source characteristics that may be less relevant to free recall may not benefit as much from testing, and they may not be associated with temporal characteristics. By this account, temporal information may have a privileged status in the testing effect as compared with less-salient qualitative information. On the other hand, perhaps free recall testing increases all qualitative characteristics associated with the tested items in the same manner. If so, the testing effect should extend to non-temporal characteristics as well.

EXPERIMENT 1

To test these hypotheses we directly replicated Chan and McDermott's (2007) two-list discrimination paradigm, with two critical changes. First, the items experienced at encoding for both lists were spoken by either a male or female speaker (i.e., they had gender information). Second, we sought to replicate the list discrimination (temporal) advantage in one condition but, critically, we tested gender information in another condition. If the effect of testing extends to all manner of qualitative characteristics, then we should obtain comparable increases in both list and gender discrimination judgements subsequent to testing as compared with not having an earlier free recall task.

Method

Participants. A total of 120 University of Georgia undergraduates participated in exchange for credit in a research participation course. All participants were randomly assigned to one of four between-participants conditions and they were tested individually in sessions that lasted approximately 20 minutes.

Materials and procedure. All three experiments presented in this paper followed the same basic paradigm shown in Figure 1 and they used the same stimulus materials. A total of 60 medium-frequency words were chosen from the Kučera and Francis (1967) word norms. For each participant, all of these items were randomised anew to be spoken either by a male or a female. Additionally, the 60 items were further divided, randomly, such that half of the items spoken by a particular source would occur in a first presentation list of 30 items and the other half would occur in the second presentation list. During the study phases for each list, items were presented both visually on a computer monitor for 2 seconds apiece and were simultaneously spoken aloud

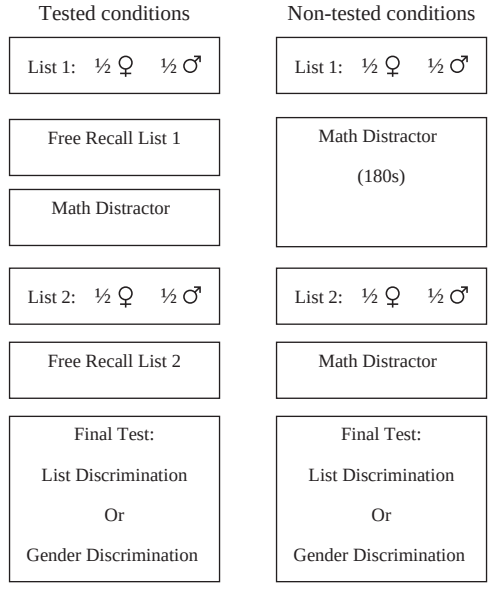

Figure 1. The general procedure for all three experiments.

over the computer speakers by either the male or female speaker. Participants were instructed to study the items for a later memory task with no indication given that their memory would be tested for gender, list, or word information. The general experimental paradigm resembled Chan and McDermott's (2007) experimental design as depicted in Figure 1.

The basic design followed a 2 (Recall vs No Recall) × 2 (List vs Gender discrimination) orthogonal cross of initial free recall testing and the type of final source-monitoring testing. More specifically, half of the participants recalled items from each list immediately after their presentation and the other half did a maths distractor task (i.e., adding two numbers together as quickly as possible) in order to equate the overall experimental time between the Recall versus the No Recall groups. After the first list presentation, participants in the Recall groups were instructed to type their responses into the computer and they were given 90 seconds to output as many items from the preceding list as possible. After typing each response, participants pressed the enter key and then typed their next response below their previous response. Half of the participants in the Recall group and half of the participants in the No Recall group completed a final list discrimination test immediately following their second free recall test or maths distractor respectively. The remaining participants took a gender discrimination test.

The discrimination tests used a two-alternative forced choice (2AFC) procedure in which participants were given one item at a time and were asked to judge on which list the item occurred, or whether a male versus a female had originally spoken the item. Participants in the list discrimination group never made decisions about the gender of the speaker of the item, and vice versa. All 2AFC responses were made by pressing one of two keys on the keyboard, then the next item was presented. All 60 items from the two lists were tested. The proportion of items correctly discriminated pooling over list (either for source or gender depending on the discrimination condition) was the primary dependent variable in the following analyses but we also recorded participants' free recall responses, and supplemental analyses will be reported. There was no significant difference between performances for list 1 versus list 2 items, and thus pooled estimates of discrimination and recall are provided in Table 1 and reported in the results section. Consequently, these results suggest that participant's encoding strategies did not differ between lists.

Result and discussion

Unless explicitly stated, all effects reported in this manuscript are statistically significant at the conventional level ($p < .05$). The means and standard errors for this and subsequent experiments are presented in Tables 1 and 2, which depict accurate discrimination of list membership or gender of the speaker as well as the mean number of items recalled, respectively. For overall discrimination performance, intermediate free recall testing enhanced list discrimination but did not enhance gender discrimination, $F(1, 116) = 10.98$, $\eta_p^2 = .09$. List discrimination judgements wholly replicated Chan and McDermott's (2007) finding that free recall testing after each list improved performance on a subsequent list discrimination test, $F(1, 116) = 11.66$, $\eta_p^2 = .09$. By contrast, there was no statistical difference in gender discrimination in the recall versus the no-recall conditions despite performance being statistically above chance, smallest $t(29) = 3.35$, $d = 1.24$. This result indicates that participants sufficiently encoded the gender dimension but that intermediate free recall testing did not provide a general enhancement to subsequent discriminability for that particular dimension. Somewhat at odds with the general enhancement from intermediate testing that is typically reported in the literature, free recall testing did not uniformly confer an advantage to all manner of source characteristics that were presented at encoding.

To further investigate these effects we analysed source discriminability separately for those items that were correctly recalled and not recalled in the list and gender discrimination conditions. As can be seen in Table 2, there was no difference in the overall proportion of items recalled in the list and gender discrimination conditions. Retrieving items during intermediate free recall testing provided a general benefit to both recalled ($M = .77$) and non-recalled ($M = .75$) items in the list discrimination condition as compared with the comparable control condition ($M = .65$ in Table 1). Neither the recalled ($M = .59$) nor non-recalled ($M = .58$) items in the gender discrimination condition were discriminated above levels in the comparable no-recall condition ($M = .62$ from Table 1). Thus the enhancement in

TABLE 1
Mean source discrimination for the three experiments, standard errors in parentheses

		Discrimination					
		List			*Gender*		
Experiment and measure	*Condition*	*List 1*	*List 2*	*Average*	*Male*	*Female*	*Average*
Experiment 1							
	Free Recall	.75 (.02)	.78 (.02)	.76 (.01)	.59 (.02)	.57 (.03)	.58 (.02)
	No Recall	.63 (.03)	.67 (.03)	.65 (.02)	.66 (.03)	.59 (.03)	.62 (.03)
Experiment 2							
	Gender Recall	.62 (.03)	.67 (.03)	.65 (.02)	.69 (.02)	.63 (.02)	.66 (.02)
	No Recall	.64 (.02)	.66 (.03)	.65 (.01)	.60 (.02)	.58 (.02)	.59 (.02)
Experiment 3							
	Free Recall			.76 (.02)			
	Temporal Recall			.71 (.02)			

There was no gender discrimination task in Experiment 3. Discrimination refers to the proportion of items correctly attributed to either the list or gender source.

memory arising from intermediate free recall testing extended to both the recalled and the non-recalled items list memory but it did not extend to gender memory in either case. This general enhancement to both recalled and non-recalled items in the list discrimination condition suggests that testing may serve to contextually isolate lists of items leading to improved temporal discriminability (Chan & McDermott, 2007; Jang & Huber, 2008).

One alternative explanation for the current results is that list 2 items are stronger than list 1 items due to their more recent presentation and subsequent recall testing. This explanation suggests that participants may have improved on their source memory performance by attributing forgotten items to the weaker source (e.g., list 1). Therefore participants in the free recall testing condition may have had a bias to guess list 1 for items when they did not have enough source information for making a correct judgement. This bias would be accompanied by a higher misattribution rate for forgotten list 2 items. As can be seen in Table 1, participants had equivalent source memory performance for list 1 and list 2 items. Therefore it seems as if testing embellished memory for list, rather than creating a bias in favour of attributing forgotten items to the weaker source. Of course, this result does not unequivocally rule out a differential strengthening plus bias account of this form of the testing effect. In the general discussion we will return to this issue and suggest future research to fully disentangle these two competing accounts.

These results of Experiment 1 inform us in important ways about the role of free recall

TABLE 2
Mean proportion of words recalled for the three experiments, standard errors in parentheses

		Recall					
		List			*Gender*		
Experiment and measure	*Condition*	*List 1*	*List 2*	*Average*	*List 1*	*List 2*	*Average*
Experiment 1							
	Free Recall	.40 (.02)	.40 (.02)	.40 (.02)	.42 (.02)	.40 (.02)	.41 (.02)
Experiment 2							
	Gender Recall	.29 (.02)	.26 (.02)	.28 (.02)	.32 (.02)	.30 (.02)	0.31 (.02)
Experiment 3							
	Free Recall	.32 (.03)	.33 (.03)	.32 (.03)			
	Temporal Recall	.20 (.01)	.18 (.02)	.19 (.02)			

Recall refers to the total number of items (out of 30 in each list and out of 60 averaged over both lists) that were recalled in each condition.

testing on later source discrimination. First, when items from two lists are free recalled immediately after presentation, subsequent list discrimination is significantly increased as compared with a no-recall control group who only performed a maths distractor task (Chan & McDermott, 2007). Second, and most important, free recall testing did not confer as strong an advantage for gender discrimination as it did for list discrimination, despite otherwise identical study and test conditions. The positive effects of testing on source memory performance may not hold in the same fashion for all qualitative characteristics experienced during the original encoding episode. Even for items that were explicitly output on intermediate free recall tests, subsequent discrimination was better for temporal information than for gender information. This outcome strongly suggests that the dynamics of the testing effect need further scrutiny, which we begin to do in the next two experiments.

EXPERIMENT 2

Our working hypothesis from the preceding results is that free recall testing imbues items with extra temporal details at the time of recall that subsequently aids list discrimination judgements on the final test. By this account, the gender of the speaker is not necessary during free recall and this information may not be accessed or embellished as described earlier. Consequently, memory for gender information is not as affected as temporal information by immediate free recall testing. If this hypothesis is true, then we reasoned that requiring gender information during the immediate free recall test would highlight this particular qualitative characteristic. By asking people to specify gender at the time of recall, we predicted that such information would therefore elicit the same advantage to the gender discrimination as was found in Experiment 1 for list discrimination. Thus, in the next experiment when we asked for recall, participants were additionally asked to specify the gender of the speaker of the word after outputting each item. We had no a priori prediction of whether such a requirement during free recall would change the effect of free recall testing in terms of list discrimination; rather, we treated this issue as an empirical question.

Method

Participants. A total of 120 University of Georgia undergraduates participated in exchange for credit in a research participation course. All participants were randomly assigned to one of four between-participants conditions and they were tested individually in sessions that lasted approximately 20 minutes.

Materials and procedure. All materials and procedural details of this experiment followed those from Experiment 1, with only one exception. During the free recall phases for each of the testing groups participants were instructed to indicate the gender of the speaker for each word that they recalled by pressing one of two keys to indicate whether a male versus a female spoke the item. Other than this change, the experiment was conducted identically to the procedure reported earlier. As in Experiment 1, performance for list 1 versus list 2 items did not differ in terms of free recall performance or source discrimination. This result suggests that having participants make a post-retrieval gender judgement during recall of the first list did not change encoding strategies for the second list.

Result and discussion

The outcomes from this experiment are summarised in the middle rows of Table 1. Gender discrimination was improved following immediate free recall testing that involved requiring post-retrieval gender judgements as compared with list discrimination performance, $F(1, 116) = 3.91$, $\eta_p^2 = .03$. This interaction can be decomposed into two effects. When participants were focused on gender during the free recall phase they were better at discriminating gender information on the later source test as compared with not free recalling at all, $t(58) = 2.85$, $d = .75$. By contrast, when participants were focused on gender during the initial free recall, subsequent list discrimination judgements did not show the same testing effect found in Experiment 1 or in Chan and McDermott (2007); $t(58) = .022$. This failure to replicate prior findings is consistent with the hypothesis that free recall dynamics influence source characteristics in specific ways. We will return to this issue in the general discussion after we report the outcomes from our last experiment.

When examining only the recalled items, list discrimination ($M = .64$ of the recalled items were correctly discriminated) was not different from gender discrimination ($M = .66$ of the recalled items were correctly discriminated), $F(1, 58) < 1$. In terms of the specific benefits to discrimination for items that were free recalled compared with those that were not recalled, the pattern of results was similar to the findings from Experiment 1. Retrieving items during intermediate free recall testing failed to provide a benefit to either the recalled ($M = .64$) or non-recalled ($M = .65$) items in the list discrimination condition when compared with the comparable control condition ($M = .65$ in Table 1). Conversely, the benefit to gender discrimination now extended to both the recalled ($M = .67$) and non-recalled ($M = .66$) items in the gender discrimination above levels in the comparable no-recall condition ($M = .59$ from Table 1). Thus the enhancement in memory arising from intermediate free recall testing was general, extending to gender discrimination for both recalled and non-recalled items when the free recall procedure focused participants on gender characteristics. The general enhancement to both recalled and non-recalled items is consistent with recent research demonstrating a testing-induced facilitation for non-recalled items (Chan, 2009, 2010).

The results from Experiment 2 therefore clearly demonstrate that a testing effect can be obtained on a non-temporal dimension as long as the immediate test is designed to reinforce the specific attribute that will be later evaluated. By contrast, this gender-recall manipulation reduced the testing advantage for list membership that was observed by Chan and McDermott (2007) and that was obtained in Experiment 1. By way of cross-experimental comparison, list discrimination performance fell to much lower levels and was equated across the previously tested versus non-tested conditions. We contend that focusing people on gender information during the immediate recall tests reduced the encoding of temporal characteristics that otherwise would have been stored during these initial tests (Jang & Huber, 2008). The important point is that the testing effect is sensitive to the nature of the immediate recall test beyond its general improvement across recalled items.

EXPERIMENT 3

In Experiment 3 we explored the notion of whether free recall testing truly confers a special advantage on temporal information as can be inferred from Experiment 1. The very act of recalling a list may solidify contextual details regarding when items were studied (Jang & Huber, 2008). There are, however, other temporal characteristics that could be associated with items on a list; and we were curious as to whether a requirement to evaluate such details would negatively affect source memory on a subsequent list discrimination test. As is well known, the similarity of to-be-discriminated source attributes can lead to confusion during source monitoring (for reviews see Johnson et al., 1993; Johnson & Raye, 1981). Thus we hypothesised that if the temporal characteristics on which participants are focused during intermediate recall serve to make the lists more similar, then they should have a more difficult time being subsequently discriminated. Consequently we tested only two conditions, which both required participants to free recall after each studied list. In one condition the free recall test was used just as in Experiment 1 and as in the Chan and McDermott (2007) paper. In the other condition we asked people to specify the location of an item in terms of which third of the list it had appeared. Similar to requiring gender information in Experiment 2, we reasoned that requiring such within-list information during the initial recall might disrupt the normal enhancement of information about between-list membership that leads to an effect of free recall testing on list discrimination found in Experiment 1.

Method

Participants. A total of 60 University of Georgia undergraduates participated in exchange for credit in a research participation course. All participants were randomly assigned to one of two between-participants conditions and they were tested individually in sessions that lasted approximately 20 minutes.

Materials and procedure. Only two conditions were tested in which all participants had initial free recall testing and then a subsequent list discrimination task. The only difference between these two conditions was whether or not we focused the participants on a within-list temporal distinction during the initial free recall testing.

Participants in the temporal recall group were asked to indicate whether each item that they recalled was presented at the beginning, middle, or end of the list by pressing one of three keys after typing in their recalled word. The free recall group was simply told to recall items in the same manner as was described for Experiments 1 and 2. As in Experiments 1 and 2, performance for items from both lists did not differ. This result suggests that having participants make a post-retrieval judgement during recall of the first list did not change encoding strategies for the second list.

Result and discussion

Requiring within-list temporal judgements on the immediate recall test attenuated the effect of free recall testing on final list discrimination judgements, $t(58) = 2.27$, $d = .59$ (see Table 1). As discussed previously, one concern is that more items were recalled in the standard condition as compared with the temporal recall condition (see Table 2). To alleviate these concerns we compared discrimination for only the recalled items in the standard recall ($M = .76$ of the recalled items were correctly discriminated) and the temporal recall ($M = .68$ of the recalled items were correctly discriminated) conditions, $F(1, 58) = 6.08$, $\eta_p^2 = .10$. Furthermore, comparing discrimination performance for recalled and non-recalled items replicated the findings from both Experiments 1 and 2. Making temporal judgements after recalling items at recall resulted in a decrement to both the recalled ($M = .69$) and non-recalled ($M = .71$) items in the list discrimination condition. Similar to Experiment 1, the benefit to list discrimination extended to both the recalled ($M = .76$) and non-recalled ($M = .76$) items in the standard recall condition. Thus the decrement to list discrimination from making post-retrieval temporal judgements extended to both the recalled and non-recalled items.

Of course, the magnitude of these effects is being judged relative to the comparable control results observed in Experiments 1 and 2. If we are permitted to make this comparison, then the improvement from testing to list discrimination holds across both the free recall and temporal recall conditions in Experiment 3, $t(29) = 3.89$, $d = 1.00$ and $t(29) = 2.12$, $d = .55$ respectively. In summary, the point of conducting Experiment 3 was to demonstrate that within-list similarity could impair ultimate list discrimination by rendering the evidence necessary for between-list judgements (i.e., temporal characteristics) more difficult to differentiate.

GENERAL DISCUSSION

These three experiments demonstrate that the act of free recalling leads to general improvements in memory for the items that were originally studied by increasing recollective details (Chan & McDermott, 2007). Importantly, the nature of the intermediate recall tasks influenced subsequent source memory for those items in subtle ways. Experiment 1 replicated the beneficial effects of free recall testing on list discrimination that was found by Chan and McDermott (2007). Additionally, this experiment demonstrated that free recall testing did not improve performance in a subsequent source memory task that asked participants about a less-salient qualitative characteristic (i.e., free recall did not contextualise gender information for non-recalled items but it did provide a boost in gender discrimination for only the recalled items). In Experiment 2 people who were concentrating on the gender of the items that they were recalling eliminated the effect of free recall testing effect on ultimate list discrimination, but those tested on gender discrimination showed a distinct advantage. Together, the results of Experiments 1 and 2 argue against a strict temporal-segmenting account of the testing effect on list discrimination. If the recall periods acted as a temporal break point between lists then the effect of gender-focused recall in Experiment 2 should not have nullified the list discrimination effect that was found in Chan and McDermott (2007) and replicated in Experiment 1. Rather, the dynamics of the free recall period influenced list discrimination by adding memorial information that would either subsequently be criterial to the source decision or non-criterial to the source decision. Finally, the results from Experiment 3 demonstrated a general enhancement from free recall testing where participants considered list characteristics for recalled items but at a diminished level as compared with simply free recalling.

Together, the preceding results converge on the notion that testing does not convey a unitary influence on recollective details that will be needed for subsequent source memory performance. Researchers may believe that free recall testing enhances recollective details, or

qualitative characteristics as specified in the source-monitoring framework. We do not disagree with this general hypothesis, but we do believe that the exact memorial details of recalled information will be a function of the nature of the free recall test *and* what is ultimately tested on a later occasion. The diagnosticity of qualitative information from previously recalling an item will depend on the manner in which that item was recalled as well as what focus the participant had during the recall period. These intermediate recall phases are essentially periods in which people are in what Tulving (1983) has termed a retrieval mode. During this period, people constrain their memory searches on certain attributes that will presumably aid in retrieval (Shimizu & Jacoby, 2005). They also evaluate the products of such retrieval attempts using late-correction processes. The effects of free recall on source discrimination that we have reported here are quite consistent with the notion of a post-retrieval evaluation process occurring at initial testing, which potentially enhances different characteristics in memory that in turn become easier to differentiate at a later point in time. For example, when participants in Experiment 1 engaged standard free recall strategies they presumably relied on temporal context cues to interrogate memory (Howard & Kahana, 1999). Thus, using temporal information as a context cue served to isolate lists of words and later made list discriminations easier (Jang & Huber, 2008; Szpunar, McDermott, & Roediger, 2008).

In our second experiment we redirected participants to use post-retrieval monitoring strategies that required gender information. This manipulation removed the positive increase to list discrimination found in Experiment 1, nevertheless it increased the gender information on the final source test as compared with a no-testing control condition. Focusing participants on contextual features (i.e., gender) that would not be criterial to a later source judgement had the effect of diminishing between-list contextual features. In the third experiment, asking participants to monitor their output using within-list temporal cues diminished final list discrimination as compared with only free recalling. The source-monitoring framework readily explains this finding. Competing information in memory needs to be differentiated, and the degree of similarity between those variously competing traces will influence their source discriminability (Johnson et al., 1993; Johnson & Raye, 1981). In Experiment 3, focusing participants on within-list characteristics made the information necessary for between-list judgements more similar in terms of their qualitative characteristics. The similarity in memory attributes reinforced by post-retrieval monitoring made it more difficult for participants to resolve the source of those recalled items. In Experiments 2 and 3, requiring gender information and within-list judgements tended to reduce between-list discriminations respectively. Therefore the results from this study both qualify the effects of immediate free recall testing and draw a strong parallel to the transfer-appropriate processing literature (Morris, Bransford, & Franks, 1977; Tulving & Thompson, 1973). Concerning the latter, the most readily available information on a given memory test will be a consequence of the manner in which it was studied or reinforced prior to that test.

This research highlights several avenues for future research. One important direction for researchers interested in the effects of testing on subsequent source memory will be to investigate additional within- and between-list source dimensions. One limitation of the current study is that memory may be better for list information than gender information. Future studies should attempt to equate discriminability for the two source dimensions as well as manipulating both source dimensions either within or between lists. Also, future studies could add new items to the final discrimination task in order to get a better handle on the differential strengthening account described earlier. Clearer evidence in favour of a bias towards list 1 items may be found in a design that includes new items into the source memory test (i.e., more new items would incorrectly be attributed to list 1 than list 2). Finally, the results from Experiment 3 suggest that intermediate testing can have detrimental consequences for later source memory if it serves to make memory for the sources more similar. Future research may profitably expand on this notion by looking for mechanisms underlying the testing effect that may cause subsequent source memory errors. Researchers interested in flashbulb memory have suggested that people will rehearse and ruminate on the contextual details surrounding a salient event and inadvertently intermix details from their current context leading to erroneous memory (i.e., wrong time slice errors; Brewer, 1992). Thus there are many questions yet to be answered with regards to how free recall leads to improved source memory.

The overarching point from the current study is that post-retrieval monitoring processes changed the effect of free recall testing on subsequent discrimination tasks. A future direction for researchers will be to more thoroughly investigate the effects of both source-constrained search and post-retrieval monitoring processes on subsequent source memory. For example, when participants are constraining their search of long-term memory on certain attributes or characteristics, then these may be the only characteristics that are reinforced and/or embellished by testing (Jacoby, Shimizu, Daniels, & Rhodes, 2005). Later, when participants come to a point at which they must discriminate information in long-term memory on a dimension that was relevant to previous recall attempts, transfer will occur and the effects of testing will become apparent (McDaniel, 2007). The goal of the present work was to delineate the role of post-retrieval monitoring on increasing specific recollective details that might later improve source discriminability. Obviously, other dimensions and procedural variations on our experiments need to be tested, but we believe that these results can be used profitably to advance both theoretical and applied educational principles.

Manuscript received 31 July 2009
Manuscript accepted 11 February 2010
First published online 19 April 2010

REFERENCES

Bjork, R. A. (1975). Retrieval as a memory modifier. In R. Solso (Ed.), *Information processing and cognition: The Loyola Symposium* (pp. 123–144). Hillsdale, NJ: Lawrence Erlbaum Associates Inc.

Brewer, W. F. (1992). The theoretical and empirical status of the flashbulb memory hypothesis. In E. Winograd & U. Neisser (Eds.), *Affect and accuracy in recall: Studies of "flashbulb" memories* (Vol. 4, pp. 274–305). New York: Cambridge University Press.

Chan, J. C. K. (2009). When does retrieval induce forgetting and when does it induce facilitation? Implications for retrieval inhibition, testing effect, and text processing. *Journal of Memory and Language*, *61*, 153–170.

Chan, J. C. K. (2010). Long-term effects of testing on the recall of nontested materials. *Memory*, *10*(1), 49–57.

Chan, J. C. K., & McDermott, K. B. (2007). The testing effect in recognition memory: A dual process account. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *33*, 431–437.

Crowder, R. G. (1976). *Principles of learning and memory*. Hillsdale, NJ: Lawrence Erlbaum Associates Inc.

Hicks, J. L., & Marsh, R. L. (1999). Attempts to reduce the incidence of false recall with source monitoring. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *25*, 1195–1209.

Howard, M. W., & Kahana, M. J. (1999). Contextual variability and serial position effects in free recall. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *25*, 923–941.

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

Jang, Y., & Huber, D. E. (2008). Context retrieval and context change in free recall: Recalling from long-term memory drives list isolation. *Journal of Experimental Psychology: Learning, Memory and Cognition*, *34*, 112–127.

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

Johnson, M. K., & Raye, C. L. (1981). Reality monitoring. *Psychological Review*, *88*, 67–85.

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

McDaniel, M. A. (2007). Transfer. In H. L. Roediger III, Y. Dudai, & S. M. Fitzpatrick (Eds.), *The science of learning and memory: Concepts*. Oxford, UK: Oxford University Press.

Morris, C. D., Bransford, J. D., & Franks, J. J. (1977). Levels of processing versus transfer-appropriate processing. *Journal of Verbal Learning and Verbal Behavior*, *16*, 519–533.

Roediger, H. L., & Karpicke, J. D. (2006). The power of testing memory: Basic research and implications for educational practice. *Perspectives on Psychological Science*, *1*, 181–210.

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

Szpunar, K. K., McDermott, K. B., & Roediger, H. L. (2008). Testing during study insulates against the buildup of proactive interference. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *6*, 1392–1399.

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

Tulving, E., & Thompson, D. M. (1973). Encoding specificity and retrieval processes in episodic memory. *Psychological Review*, *80*, 352–373.

Wixted, J. T., & Rohrer, D. (1994). Analysing the dynamics of free recall: An integrative review of the empirical literature. *Psychonomic Bulletin & Review*, *1*, 89–106.