



# Practicing more retrieval routes leads to greater memory retention



Jun Zheng<sup>a,b</sup>, Wei Zhang<sup>b</sup>, Tongtong Li<sup>b</sup>, Zhaomin Liu<sup>c,\*</sup>, Liang Luo<sup>a,b,\*\*</sup>

<sup>a</sup> Collaborative Innovation Center of Assessment Toward Basic Education Quality, Beijing Normal University, Beijing, China

<sup>b</sup> State Key Laboratory of Cognitive Neuroscience and Learning, IDG/McGovern Institute for Brain Research, Beijing Normal University, Beijing, China

<sup>c</sup> School of Sociology, China University of Political Science and Law, Beijing, China

## ARTICLE INFO

### Article history:

Received 9 June 2015

Received in revised form 20 May 2016

Accepted 27 May 2016

Available online 4 June 2016

### Keywords:

Retrieval variability

Retrieval routes

Memory retention

## ABSTRACT

A wealth of research has shown that retrieval practice plays a significant role in improving memory retention. The current study focused on one simple yet rarely examined question: would repeated retrieval using two different retrieval routes or using the same retrieval route twice lead to greater long-term memory retention? Participants elaborately learned 22 Japanese-Chinese translation word pairs using two different mediators. Half an hour after the initial study phase, the participants completed two retrieval sessions using either one mediator ( $T_{m1}T_{m1}$ ) or two different mediators ( $T_{m1}T_{m2}$ ). On the final test, which was performed 1 week after the retrieval practice phase, the participants received only the cue with a request to report the mediator (M1 or M2) followed by the target (Experiment 1) or only the mediator (M1 or M2) with a request to report the target (Experiment 2). The results of Experiment 1 indicated that the participants who practiced under the  $T_{m1}T_{m2}$  condition exhibited greater target retention than those who practiced under the  $T_{m1}T_{m1}$  condition. This difference in performance was due to the significant disadvantage in mediator retrieval and decoding of the unpracticed mediator under the  $T_{m1}T_{m1}$  condition. Although mediators were provided to participants on the final test in Experiment 2, decoding of the unpracticed mediators remained less effective than decoding of the practiced mediators. We conclude that practicing multiple retrieval routes leads to greater memory retention than focusing on a single retrieval route. Thus, increasing retrieval variability during repeated retrieval practice indeed significantly improves long-term retention in a delay test.

© 2016 Elsevier B.V. All rights reserved.

## 1. Introduction

A great deal of research suggests that retrieval practice not only serves to assess what was learned but also plays a significant role in promoting learning and memory (Abott, 1909; Karpicke & Roediger, 2008; Roediger & Karpicke, 2006; for a review, see Roediger & Butler, 2011). Researchers have performed many experiments to establish the most effective modes of retrieval practice, including the ideal intervals (Karpicke & Roediger, 2007; Landauer & Bjork, 1978) and criterion (Rawson & Dunlosky, 2011) of repeated retrieval, the amount of elaboration (Carpenter, 2009; Carpenter & DeLosh, 2006), and the recollection of the initial episodic context (Karpicke, Lehman, & Gallo, 2014; Karpicke & Zaromb, 2010) during the retrieval phase. Surprisingly, few studies have explored one simple but rather important question: whether consolidating multiple retrieval routes or consolidating one retrieval route multiple times would achieve greater memory retention?

In the current study, two experiments were performed to examine which mode of retrieval is more beneficial to memory retention: retrieval practice using two different mediators or using the same mediator twice?

Although there has been a lack of research directly comparing the effectiveness of repeated retrieval using multiple retrieval routes with that using a single retrieval route multiple times, prior research focused on encoding variability (e.g. Gartman & Johnson, 1972; Glanzer & Duarte, 1971; Greenberg & Verfaellie, 2010; Smith, Glenberg, & Bjork, 1978), retrieval variability (e.g., Finley, 2012) and retrieval-based learning (e.g. Carpenter, 2009; Lehman, Smith, & Karpicke, 2014) has provided some indirect evidence related to this issue. Prior studies have demonstrated the benefit of encoding variability at the semantic level (Glanzer & Duarte, 1971; Greenberg & Verfaellie, 2010; Hintzman & Stern, 1978) and the physical/context level (Smith et al., 1978) during the initial study phase. For example, in an experiment by Greenberg and Verfaellie (2010), healthy elderly people studied word pairs three times either under a fixed condition (the mediator remained constant with each presentation, e.g., ARMY invades CITY) or under a varied condition (the mediator changed with each presentation, e.g., ARMY invades CITY, ARMY flees CITY, or ARMY patrols CITY). The participants who studied under the varied condition performed better on a

\* Correspondence to: Z. Liu, School of Sociology, China University of Political Science and Law, Beijing 102249, China.

\*\* Correspondence to: L. Luo, Collaborative Innovation Center of Assessment Toward Basic Education Quality, Beijing Normal University, Beijing 100875, China.

E-mail addresses: [zhaominl@cupl.edu.cn](mailto:zhaominl@cupl.edu.cn) (Z. Liu), [luoliang@bnu.edu.cn](mailto:luoliang@bnu.edu.cn) (L. Luo).

subsequent associative-recognition task than those who studied under the fixed condition. Similarly, Smith et al. (1978, Experiment 1) assigned the participants to study the to-be-remembered list either twice in exactly the same room or once each in two very different rooms. Then, all participants underwent a surprise free recall test in a neutral room. The superior performance of the group that studied in different rooms was consistent with the concept of mnemonic benefit for encoding variability. To our knowledge, Finley (2012) was the first to consider the issue of retrieval variability during the retrieval practice phase. In Finley's experiment, English homographic target words were presented together with four associated cue words (two cues for each of the two meanings) during the initial study phase. For example, one target word was "bat", and its four cue words were "swing, hit, fangs, and cave". Five minutes later, the participants were shown two cue words for each target word to complete the cued recall test. Double-meaning retrieval cues (e.g., swing and fangs) indeed yielded higher recall than single-meaning retrieval cues (e.g., swing and hit) (Experiment 3). This result demonstrated that increasing the variability of retrieval cues indeed greatly improves the present retrieval performance. Based on these results, would increasing the variability of repeated retrieval significantly improve long-term retention on a delay test? On the basis of the literature related to encoding variability and retrieval variability, repeated retrieval using two different retrieval routes is more likely to construct retrieval variability than repeated retrieval using the same retrieval route twice. Thus, we inferred that repeated retrieval using two different retrieval routes would be more effective than repeated retrieval using the same retrieval route twice, but this hypothesis required the support of direct evidence.

Recently, based on the accounts of spacing effects according to the contextual variability and study-phase retrieval theories, Karpicke and colleagues proposed an episodic context account of retrieval-based learning (Karpicke, Lehman, & Aue, 2014; Lehman et al., 2014), which emphasizes the roles of context reinstatement and updating during successful retrieval in increasing the retention of retrieved information. The context representation of successfully retrieved items may be updated into a composite of temporal contextual features during the initial study and retrieval practice phases. When participants subsequently attempt to retrieve the items, they can restrict their search to those items which are associated with the updated context (Karpicke, Lehman & Aue, 2014). However, the effect of contextual variability evoked by retrieval variability during the repeated retrieval practice phase on memory performance has not been explored. According to the episodic context account of retrieval-based learning, we inferred that repeated retrieval using two different retrieval routes would increase contextual variability in encoded traces and that the combination of retrieval routes would be more unique and specific to the target. Because the difference in memory retention between distinct modes of repeated retrieval was not directly examined in prior studies, it is necessary to directly compare these two retrieval conditions to gain a deeper understanding of the episodic context account.

Additionally, Bjork (1975) proposed that the concept of retrieval routes is also a concept of elaboration. In accordance with this proposal, we consider that two properties define the degree of "elaboration": the number of mediators (cascade form) included in one retrieval route and the number of retrieval routes provided by the mediators (parallel form). Multiple studies by Carpenter and colleagues have explored elaborative retrieval as a retrieval route that is encoded by highly relevant information (e.g. Carpenter, 2009; Carpenter & DeLosh, 2006). Carpenter (2009) stated that when people attempt to retrieve a target from memory, they activate several semantically related words while searching for the target and that this semantic elaboration during initial retrieval enhances retention on future tests (Carpenter, 2009, 2011). For example, after initial study of a word pair (e.g., Basket: Bread), recalling a target from a cue (e.g., Basket: \_\_\_\_\_) is more likely to activate elaborative information that is related to the cue (e.g., Eggs or Wicker). However, the effect of manipulating semantic elaboration by changing the

number of retrieval routes provided by the mediators has rarely been examined. The present study was performed to resolve this deficiency.

The current study directly examined whether retrieval practice using two different mediators or the same mediator twice would achieve greater memory retention. As depicted in Fig. 1, the paradigm in the current experiments was modelled after the report by Pyc and Rawson (2010). Both experiments included three phases: an initial study phase, a retrieval practice phase, and a final test phase. The initial study and retrieval practice phases were identical between the two experiments. In the initial study phase, participants were instructed to utilize mediators to study a list of Japanese-Chinese translation pairs. They completed two initial study sessions (S1 & S2, see Fig. 1), each involving one mediator for each pair. To better control the establishment of the retrieval route, mediators were assigned by the experimenter instead of the participants. Thus, two mediators were introduced to ensure that all participants would be provided an opportunity to establish two retrieval routes for each word pair. After the initial study phase, the participants completed two retrieval practice sessions (T1 & T2, see Fig. 1), either using the same mediator ( $T_{m1}T_{m1}$  condition) or different mediators ( $T_{m1}T_{m2}$  condition). Under the  $T_{m1}T_{m1}$  condition, each participant was provided with two opportunities to consolidate the retrieval route for one mediator but no opportunities for the other mediator. Under the  $T_{m1}T_{m2}$  condition, each participant was provided with one opportunity to consolidate the retrieval route for each of the two mediators. In Experiment 1, a final cued recall test, which involved cue recall and mediator recall (CMR test), was adopted to explore the differences in performance of target recall, mediator retrieval (i.e., the mediator is recalled when prompted with the cue) and mediator decoding (i.e., the mediator elicits the target from memory) between the two retrieval practice conditions. In Experiment 2, a final test using only one mediator (M test) was adopted to further explore mediator decoding.

## 2. Experiment 1

In Experiment 1, after the initial study phase, participants were randomly assigned to two groups, the  $T_{m1}T_{m1}$  condition or the  $T_{m1}T_{m2}$  condition. Under the  $T_{m1}T_{m1}$  condition, only one mediator was used during the two sessions of retrieval practice. Under the  $T_{m1}T_{m2}$  condition, two mediators were used during the two sessions of retrieval practice. All participants received the final cued test to recall the mediator and the target (see Fig. 1). We assessed whether consolidating more retrieval routes would lead to greater memory retention. When the target was successfully retrieved twice during the retrieval practice phase, this item was considered to have been successfully retrieved either twice via a single retrieval route ( $T_{m1}T_{m1}$  condition) or once each via two different retrieval routes ( $T_{m1}T_{m2}$  condition). Thus, how items which were successfully retrieved twice during the retrieval practice phase retained on the final test was the primary focus of the current study. We hypothesized that target retention under the  $T_{m1}T_{m2}$  condition would be greater than that under the  $T_{m1}T_{m1}$  condition one week after retrieval practice phase. To further explain the mechanism underlying this hypothesized difference in target retention, we specifically analysed mediator retrieval and mediator decoding performance.

### 2.1. Method

#### 2.1.1. Participants

Sixty-four students (30 males, 34 females) from Beijing Normal University participated in Experiment 1 for a reward. Thirty participants (16 females; mean age = 22.37;  $SD = 2.08$ ) practiced under the  $T_{m1}T_{m1}$  condition, and thirty-four participants (18 females; mean age = 21.59;  $SD = 2.46$ ) practiced under the  $T_{m1}T_{m2}$  condition. All participants had normal or corrected vision, were native speakers of Chinese, and had no previous experience with Japanese. The participants completed the experiment in individual booths on desktop PCs.

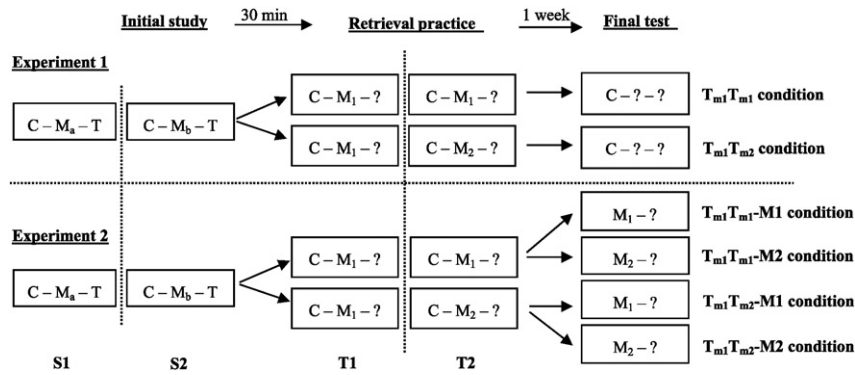


Fig. 1. Illustration of the general methodology used in the current study.

2.1.2. Materials

Twenty-two Japanese-Chinese translation pairs were adopted as learning material. Japanese words were used because they are from a foreign language but are orthographically similar to Chinese words. Therefore, these materials meet the requirements of the experiment very well. The preparation process included three steps: selecting Japanese words, generating mediators, and evaluating these words.

First, Japanese words were selected to form Japanese-Chinese translation pairs from the Chinese-Japanese lexicon (Liu, 1984; Wang, 2000). Japanese words were strictly screened using the following criteria: 1) the Japanese words should consist of exactly the same characters as Chinese characters, but without any meaning under Chinese language circumstances; 2) the Japanese words should include two characters, identical to the number of characters in the corresponding Chinese words; and 3) the corresponding Chinese words should be common, concrete and imaginable nouns. Using these selection criteria, 51 Japanese words were selected for the material pool.

Second, 15 students were instructed to generate two mediators (Chinese words) for each word pair; each mediator shared one character with the Japanese words and was semantically associated with a Chinese target word. Mediators were selected using the following criteria: 1) two mediators are generated for each word pair; 2) both mediators are common, concrete, and imaginable nouns; and 3) there is no obvious difference in the strength of association with the target between the two mediators. Using these criteria, three word pairs were excluded because of a failure to generate two mediators; six word pairs were excluded because all mediators generated were too abstract; twelve word pairs were excluded because the strength of association with the target was clearly greater for one mediator than the other (as judged by the experimenter); and three word pairs were excluded because we found that the Japanese word had a meaning in ancient Chinese. Finally, 27 word pairs, each matched with two specific mediators, were selected for further evaluation. Then, the experimenter generated a sentence corresponding to each mediator to describe its semantic association with the target. For example, “泥棒” was the Japanese word, and its Chinese translation was “小偷” (thief). For the pair “泥棒—小偷”, the mediators were “泥巴” (mud) and “木棒” (stick), with each mediator sharing one of the two characters of the Japanese word. The generated sentences were “小偷把泥巴糊脸上以防被认出” (mud: the thief plastered mud on the face) and “用木棒打小偷” (stick: beat the thief with the stick).

Then, 105 students completed concrete assessment scales (1: very abstract; 6: very concrete) for the 54 generated mediators. Another 115 students completed semantic correlation scales (1: no correlation; 4: strongly correlated) for the 54 “mediator-target” pairs and the 27 “mediator 1–mediator 2” pairs. Another 117 students completed a scale to evaluate the helpfulness of the sentence for remembering the word pairs. Ultimately, 5 Japanese–Chinese translation pairs were excluded because of a large difference in concrete value, semantic correlation with the target or sentence helpfulness between the two mediators. Finally, the materials used in the two experiments consisted of 22

Japanese-Chinese translation pairs and their corresponding two mediators and sentences. We divided the mediators into two groups and analysed their ratings. A paired-sample *t*-test showed no significant difference in concrete value ( $M = 4.58, SD = 0.93$  vs.  $M = 4.52, SD = 1.07$ ),  $t(21) = 0.235, SE = 0.265, p = 0.817$ , correlation with the target ( $M = 1.85, SD = 0.52$  vs.  $M = 1.73, SD = 0.53$ ),  $t(21) = 0.754, SE = 0.163, p = 0.459$ , or sentence helpfulness ( $M = 2.52, SD = 0.15$  vs.  $M = 2.48, SD = 0.17$ ),  $t(21) = 0.848, SE = 0.051, p = 0.406$ , between the two mediator groups.

2.1.3. Procedure

The procedure involved three stages: an initial study phase, a retrieval practice phase, and a final test phase (see Fig. 1). The experiment was programmed using E-Prime 1.0.

In the initial study phase, all participants were informed that they would learn a list of 22 Japanese-Chinese translation pairs, which were accompanied by two mediators and their corresponding sentences, in two sessions (S1 and S2 in Fig. 1). The presentation order of the word pairs was randomized. During the study trial, participants saw a Japanese word and its Chinese translation word in the centre of the computer screen, with the Japanese word (cue) on the left and the Chinese translation word (target) on the right. The mediator was in the middle between the cue and the target, and the sentence was in the centre below the word pairs. At the bottom of the screen, a 4-point rating scale appeared and requested a participant response. The participants were asked to type in a number indicating the helpfulness of the presented mediator for learning the word pair when they attempted to use the mediator to recall the word pair in the subsequent recall test. A response of 1 indicated “not helpful”, and 4 indicated “very helpful”. Each study trial was self-paced by the participants, and as soon as participants pressed the “Enter” key, the next trial appeared. This process continued until the end of the session. Then, all participants learned the word pair list a second time using the other mediator. The sequence in which the two mediator groups were presented was balanced between participants. For convenience, the mediator group that was presented in the first study session (S1) was referred to as Ma, and the mediator group that was presented in the second study session (S2) was referred to as Mb. After completing the initial study phase, the participants were asked to play the game “Angry Birds” for 30 min.

Table 1 Mean helpfulness rating (on a 1–4 scale) of the mediator for remembering word pairs and learning time of the two initial study lists (min) in Experiment 1. Standard deviations are presented in parentheses.

Condition	Mediator helpfulness		Learning time	
	S1	S2	S1	S2
T <sub>m1</sub> T <sub>m1</sub>	2.31 (0.40)	2.48 (0.48)	6.18 (3.30)	5.96 (3.09)
T <sub>m1</sub> T <sub>m2</sub>	2.31 (0.44)	2.36 (0.38)	6.53 (2.60)	6.45 (3.07)

**Table 2**

Retrieval performance in the  $T_{m1}T_{m1}$  condition and  $T_{m1}T_{m2}$  condition in Experiment 1. Standard deviations are presented in parentheses.

Condition	$T1_C T2_C$	$T1_C T2_N$	$T1_N T2_C$	$T1_N T2_N$
$T_{m1}T_{m1}$	0.79 (0.19)	0.01 (0.02)	0.02 (0.02)	0.19 (0.17)
$T_{m1}T_{m2}$	0.81 (0.18)	0.04 (0.06)	0.04 (0.06)	0.10 (0.12)

Note:  $T1_C T2_C$  items were successfully retrieved during both the first retrieval session and the second retrieval session.  $T1_C T2_N$  items were successfully retrieved during the first retrieval session but were not successfully retrieved during the second retrieval session.  $T1_N T2_C$  items were not successfully retrieved during the first retrieval session but successfully retrieved during the second retrieval session.  $T1_N T2_N$  items were not successfully retrieved during either retrieval sessions. All totals do not add up to 100% because the numbers were rounded.

After this thirty-minute break, the retrieval practice phase was initiated. During the retrieval practice phase, Japanese words and their mediators were presented to the participants, and the participants were assigned to recall the Chinese responses. The participants were then instructed to type the Chinese word and its corresponding sentence from the initial study of the word pairs. Participants needed only to type the main points of the sentence rather than repeat the intact expression. This operation was employed to ensure that the participants indeed retrieved the target via the mediator assigned to them, as this step is critical in consolidating the retrieval route to which the participants were previously exposed. The target response and its corresponding sentence were separated by a space. All participants were randomly distributed into two groups to complete two sessions of retrieval practice (see Fig. 1). For the  $T_{m1}T_{m1}$  condition, the participants completed two sessions of retrieval practice using the same mediator. For the  $T_{m1}T_{m2}$  condition, the participants completed two sessions of retrieval practice using a different mediator in each session. For convenience, the mediator group that was presented in the first retrieval session ( $T1$ ) was labelled M1 under both two conditions. The same mediator group (M1) was presented in the second retrieval session ( $T2$ ) under the  $T_{m1}T_{m1}$  condition. However, the other mediator group, which was named M2, was presented in the second retrieval session ( $T2$ ) under the  $T_{m1}T_{m2}$  condition. To control for any effect of order in encountering the mediators during the initial study phase, the two mediator groups that were presented during the retrieval practice phase were reconstructed. The M1 group consisted of half of groups Ma and Mb, and M2 group consisted of the remaining half of groups Ma and Mb. The construction of M1 and M2 was balanced between participants. Upon completion of the retrieval task, the participants were dismissed and asked to return 1 week later.

For the final test, the participants were shown Japanese words and were instructed to report their mediator words and the corresponding Chinese words as soon as possible. As a note, the participants had an opportunity to recall only one mediator. If they were able to recall both mediators, then they needed to choose one. If a participant failed to recall any mediator word or the Chinese word, the participant would say, "I forgot." For example, if a person successfully remembered the mediator word but failed to recall the Chinese word, then the response would be the mediator word followed by "I forgot." Once the participants had

finished each recall trial, they pressed the Enter button to begin the next trial. A digital voice recorder was used to record their responses.

#### 2.1.4. Analysis

For the initial study phase, to compare the effectiveness of the mediators and learning difficulty between the two conditions, a mediator helpfulness rating and learning time were analysed.

For the retrieval practice phase, four possible outcomes may occur: items were successfully retrieved during both the first retrieval session and the second retrieval session ( $T1_C T2_C$  items); items were successfully retrieved during the first retrieval session but not during the second retrieval session ( $T1_C T2_N$  items); items were not successfully retrieved during the first retrieval session but were successfully retrieved during the second retrieval session ( $T1_N T2_C$  items); or items were not successfully retrieved during either retrieval sessions ( $T1_N T2_N$  items). Thus, two indexes can be separately calculated to represent retrieval performance: the proportion of items successfully retrieved at least once ( $(T1_C T2_C \text{ items} + T1_C T2_N \text{ items} + T1_N T2_C \text{ items}) / 22$ ) and the proportion of items successfully retrieved twice ( $T1_C T2_C \text{ items} / 22$ ). As mentioned above, successful consolidation of the retrieval route "cue-mediator-target" was considered to have occurred when the target was successfully recalled from the cue via the mediator. Under the  $T_{m1}T_{m1}$  condition, when participants successfully retrieved the target twice via a single mediator (i.e.,  $T1_C T2_C$  items), the participants were considered to have successfully consolidated this retrieval route twice. Alternatively, under the  $T_{m1}T_{m2}$  condition, when participants successfully retrieved the target via each of the two different mediators (i.e.,  $T1_C T2_C$  items), they were considered to have successfully consolidated each retrieval route once. Therefore, the latter index would be more meaningful for the purpose of the present study.

For the final test phase, an analysis for  $T1_C T2_C$  items on their final recall performance was conducted on the following three parameters: total target performance, mediator retrieval, and mediator decoding. For  $T1_C T2_C$  items, total target performance was calculated as the proportion of those targets which were correctly recalled on the final test, regardless of whether the participants correctly reported the mediators. Total target performance could be divided into two subsets: target performance with mediator retrieval and target performance without mediator retrieval. Target performance with mediator retrieval was determined as the probability that those targets would be successfully recalled via the mediator (M1 or M2). Target performance without mediator retrieval was the probability that those targets would be correctly recalled without reporting the correct mediators. Target performance with mediator retrieval could be further divided into two subsets, targets recalled via M1 and targets recalled via M2, which we analysed separately. According to Pyc and Rawson (2010), mediator retrieval and mediator decoding are two important factors related to mediator effectiveness. We further analysed the participants' performance in mediator retrieval and decoding. Mediator retrieval was calculated as the proportion of mediators that were correctly recalled, regardless of whether the target was correctly provided. Mediator decoding was the probability that those mediators which were correctly reported would successfully

**Table 3**

Final test performance for  $T1_C T2_C$  items in Experiment 1. Standard deviations are presented in parentheses.

Condition	Total target performance	Target performance with mediator retrieval	Target recalled via		Mediator retrieval		Mediator decoding	
			M1	M2	M1	M2	M1	M2
$T_{m1}T_{m1}$	0.81 (0.12)	0.68 (0.17)	0.56 (0.16)	0.12 (0.07)	0.59 (0.17)	0.15 (0.09)	0.95 (0.08)	0.84 (0.20)
$T_{m1}T_{m2}$	0.87 (0.09)	0.79 (0.13)	0.41 (0.10)	0.39 (0.13)	0.44 (0.10)	0.42 (0.12)	0.91 (0.12)	0.92 (0.13)

Note: For those items which were successfully retrieved twice during the retrieval practice phase ( $T1_C T2_C$  items), indexes of the final test and their computational formulas were defined as follows: total target performance: the number of targets that were correctly recalled in the final test for  $T1_C T2_C$  items divided by the number of  $T1_C T2_C$  items; target performance with mediator retrieval: the number of targets that were correctly recalled via M1 or M2 in the final test for  $T1_C T2_C$  items divided by the number of  $T1_C T2_C$  items; M1 (or M2) retrieval: the number of mediators in M1 (or M2) that were correctly recalled in the final test for  $T1_C T2_C$  items divided by the number of  $T1_C T2_C$  items; and M1 (or M2) decoding: the number of mediators in M1 (or M2) that correctly elicited the target in the final test for  $T1_C T2_C$  items divided by the number of mediators in M1 (or M2) that were correctly recalled in the final test for  $T1_C T2_C$  items.

elicit their target. An analysis of final recall performance for all 22 items that were initially studied can be found in the supplementary materials.

## 2.2. Results

### 2.2.1. Initial study phase

The average mediator helpfulness rating and learning time are presented in Table 1. For the initial mediator helpfulness rating, a 2 (retrieval practice:  $T_{m1}T_{m1}$  or  $T_{m1}T_{m2}$ )  $\times$  2 (study list: S1 or S2) mixed analysis of variance (ANOVA) revealed a main effect of study list,  $F(1, 62) = 4.521$ ,  $MSE = 0.366$ ,  $p = 0.037$ ,  $\eta_p^2 = 0.068$ , with the mediator helpfulness rating higher for S2 than for S1. There was no significant effect of retrieval practice,  $F < 1$ , or a significant interaction effect,  $F(1, 62) = 1.521$ ,  $MSE = 0.123$ ,  $p = 0.222$ ,  $\eta_p^2 = 0.024$ . For learning time, no effect of retrieval practice, study list, or their interaction was found,  $F_s < 1$ .

### 2.2.2. Retrieval practice phase

The retrieval performance<sup>1</sup> of each possible outcome is listed in Table 2. One-way ANOVA revealed that the proportion of items that were successfully retrieved at least once ( $(T_{1c}T_{2c}$  items +  $T_{1c}T_{2n}$  items +  $T_{1n}T_{2c}$  items) / 22) under the  $T_{m1}T_{m2}$  condition ( $M = 0.90$ ,  $SD = 0.12$ ) was significantly greater than that under the  $T_{m1}T_{m1}$  condition ( $M = 0.81$ ,  $SD = 0.17$ ),  $F(1, 62) = 5.076$ ,  $MSE = 55.417$ ,  $p = 0.028$ ,  $\eta_p^2 = 0.076$ . More importantly, there was no significant difference in the proportion of items that were successfully retrieved twice ( $T_{1c}T_{2c}$  items / 22) across conditions,  $F < 1$ .

### 2.2.3. Final test phase

As noted above, when items were successfully retrieved twice during the retrieval practice phase, they were considered to have been successfully retrieved via one retrieval route twice ( $T_{m1}T_{m1}$  condition) or once each via two different retrieval routes ( $T_{m1}T_{m2}$  condition). Thus, an analysis for  $T_{1c}T_{2c}$  items on the final test was conducted. The results of the final recall test for all 22 items are presented in Appendix 1.<sup>2</sup>

Final recall test performance for items that were successfully retrieved twice during the retrieval practice phase ( $T_{1c}T_{2c}$  items) is shown in Table 3. One-way ANOVA showed that total target performance under the  $T_{m1}T_{m2}$  condition was significantly higher than that under the  $T_{m1}T_{m1}$  condition,  $F(1, 62) = 5.045$ ,  $MSE = 0.057$ ,  $p = 0.028$ ,  $\eta_p^2 = 0.075$ . Target performance with mediator retrieval under the  $T_{m1}T_{m2}$  condition was also significantly higher than that under the  $T_{m1}T_{m1}$  condition,  $F(1, 62) = 9.560$ ,  $MSE = 0.209$ ,  $p = 0.003$ ,  $\eta_p^2 = 0.134$ . Interestingly, the participants exposed to the  $T_{m1}T_{m1}$  condition ( $M = 0.13$ ,  $SD = 0.12$ ) recalled more targets without mediator retrieval than those exposed to the  $T_{m1}T_{m2}$  condition ( $M = 0.07$ ,  $SD = 0.07$ ),  $F(1, 62) = 5.408$ ,  $MSE = 0.047$ ,  $p = 0.023$ ,  $\eta_p^2 = 0.080$ . For those targets which were successfully recalled via mediator retrieval, target performance with M1 retrieval under the  $T_{m1}T_{m1}$  condition was significantly higher than that under the  $T_{m1}T_{m2}$  condition,  $F(1, 62) = 19.431$ ,  $MSE = 0.357$ ,  $p < 0.001$ ,  $\eta_p^2 = 0.239$ , but the opposite results were observed for target performance with M2 retrieval,  $F(1, 62) = 100.512$ ,  $MSE = 1.113$ ,  $p < 0.001$ ,  $\eta_p^2 = 0.618$ .

Concerning mediator retrieval, recollection of mediators under the  $T_{m1}T_{m2}$  condition ( $M = 0.86$ ,  $SD = 0.10$ ) was significantly greater than that under the  $T_{m1}T_{m1}$  condition ( $M = 0.74$ ,  $SD = 0.17$ ),  $F(1, 62) = 12.656$ ,  $MSE = 0.231$ ,  $p = 0.001$ ,  $\eta_p^2 = 0.170$ . Additionally, M1

<sup>1</sup> Target retrieval was regarded as a correct answer with a correct sentence. The probability of reporting a correct target but no sentence under the  $T_{m1}T_{m1}$  condition was 2.88% and 1.97% for sessions 1 and 2 of retrieval practice, respectively; under the  $T_{m1}T_{m2}$  condition, these values were 2.41% and 2.81%, respectively.

<sup>2</sup> When all 22 items were included in the analysis, the pattern of results was the same as the results from analysis of only items that were successfully retrieved twice during the retrieval practice phase, except that the difference in targets recalled without mediator retrieval was no longer significant,  $F < 1.83$ . Further details can be found in the supplementary materials.

**Table 4**

Mean helpfulness rating (on a 1–4 scale) of the mediator for remembering word pairs and learning time of the two initial study lists (min) in Experiment 2. Standard deviations are presented in parentheses.

Condition	Mediator helpfulness		Learning time	
	S1	S2	S1	S2
$T_{m1}T_{m1}$ -M1	2.45 (0.51)	2.55 (0.52)	8.70 (5.58)	6.63 (4.62)
$T_{m1}T_{m1}$ -M2	2.43 (0.37)	2.62 (0.46)	7.09 (3.60)	6.16 (3.25)
$T_{m1}T_{m2}$ -M1	2.28 (0.47)	2.26 (0.48)	7.08 (2.99)	7.21 (4.66)
$T_{m1}T_{m2}$ -M2	2.38 (0.42)	2.37 (0.48)	6.85 (4.62)	6.59 (3.12)

retrieval under the  $T_{m1}T_{m1}$  condition was significantly greater than that under the  $T_{m1}T_{m2}$  condition,  $F(1, 62) = 18.106$ ,  $MSE = 0.328$ ,  $p < 0.001$ ,  $\eta_p^2 = 0.226$ , but the opposite results were observed for M2 retrieval,  $F(1, 62) = 95.069$ ,  $MSE = 1.110$ ,  $p < 0.001$ ,  $\eta_p^2 = 0.625$ .

In contrast, no significant difference in mediator decoding was observed between the  $T_{m1}T_{m1}$  condition ( $M = 0.92$ ,  $SD = 0.08$ ) and the  $T_{m1}T_{m2}$  condition ( $M = 0.92$ ,  $SD = 0.09$ ),  $F < 1$ . M1 decoding under the  $T_{m1}T_{m2}$  condition was not significantly different from that under the  $T_{m1}T_{m1}$  condition,  $F(1, 62) = 1.778$ ,  $MSE = 0.019$ ,  $p = 0.187$ ,  $\eta_p^2 = 0.028$ . However, M2 decoding under the  $T_{m1}T_{m2}$  condition was marginally significantly greater than that under the  $T_{m1}T_{m1}$  condition,  $F(1, 60)^3 = 3.666$ ,  $MSE = 0.097$ ,  $p = 0.060$ ,  $\eta_p^2 = 0.058$ .

## 2.3. Discussion

During the initial study phase, there was only a main effect of study list on the mediator helpfulness rating. Even so, the study list might not have influenced the results for each condition during the retrieval practice phase. As stated above, the reconstruction of M1 and M2 was balanced between participants. For half of the participants, M1 was reconstructed to contain half of the contents of Ma and Mb; the remaining mediators constituted M2. For the other half of the participants, M1 was reconstructed to contain the opposite half of the contents of Ma and Mb; the remaining mediators constituted M2. Thus, the difference in the mediator helpfulness rating across study lists during the initial study phase may not have influenced the results during the retrieval practice and final test phases. Additionally, there was no significant difference in learning time between the conditions, and this result suggests that learning difficulty was not significantly different between conditions.

During the retrieval practice phase, the proportion of items that were successfully retrieved at least once under the  $T_{m1}T_{m2}$  condition was significantly greater than that under the  $T_{m1}T_{m1}$  condition. One possible explanation for this result was that the participants were more likely to succeed in target retrieval when relying on two mediators rather than one mediator. If participants exposed to the  $T_{m1}T_{m1}$  condition failed to utilize M1 during the first retrieval session, they were unlikely to utilize the same mediator (M1) during the second retrieval session. However, if participants exposed to the  $T_{m1}T_{m2}$  condition failed to utilize M1 in the first retrieval session, they were provided with another opportunity to utilize the different M2 during the second retrieval session. To verify this inference, the proportion of items that were successfully retrieved only once was computed by calculating the sum of  $T_{1c}T_{2n}$  items and  $T_{1n}T_{2c}$  items. Consistent with the above inference, the proportion of items that were successfully retrieved only once under the  $T_{m1}T_{m2}$  condition ( $M = 0.09$ ;  $SD = 0.10$ ) was indeed significantly greater than that under the  $T_{m1}T_{m1}$  condition ( $M = 0.02$ ;  $SD = 0.04$ ),  $F(1, 62) = 10.768$ ,  $MSE = 0.060$ ,  $p = 0.002$ ,  $\eta_p^2 = 0.148$ . Additional comparisons were conducted to examine any possible differences during each retrieval session between the  $T_{m1}T_{m1}$  and  $T_{m1}T_{m2}$  conditions. In fact, there was no significant difference in the proportion of items that were successfully retrieved during the first retrieval

<sup>3</sup> Two participants did not recall any M2 in final test.

session,  $F(1, 62) = 2.0$ ,  $MSE = 0.054$ ,  $p = 0.162$ ,  $\eta_p^2 = 0.031$ , in which M1 was presented under both conditions, between conditions (see Table 2,  $T_{1c}T_{2c}$  items +  $T_{1c}T_{2n}$  items). Furthermore, there was no significant difference in the proportion of items that were successfully retrieved during the second retrieval session,  $F(1, 62) = 1.405$ ,  $MSE = 0.039$ ,  $p = 0.240$ ,  $\eta_p^2 = 0.022$ , in which different mediators were used for each condition, between the two conditions (see Table 2,  $T_{1c}T_{2c}$  items +  $T_{1n}T_{2c}$  items). In particular, there was no significant difference in the proportion of items that were successfully retrieved twice between the two conditions.

During the final test phase, considering either only those items which were successfully retrieved twice during the retrieval practice phase ( $T_{1c}T_{2c}$  items) or all 22 items that were initially studied (see Appendix 1), total target performance was significantly higher under the  $T_{m1}T_{m2}$  condition than under the  $T_{m1}T_{m1}$  condition. This result indicates that consolidating two different retrieval routes was more beneficial for memory retention than consolidating a single retrieval route twice. The result of target performance with mediator retrieval supported this conclusion. Further analysis revealed that the participants exposed to the  $T_{m1}T_{m1}$  condition relied more heavily on M1 to complete target recall but were indeed incompetent at utilizing M2 to complete target recall. From the analysis above, it could be inferred that the participants exposed to the  $T_{m1}T_{m2}$  condition outperformed those exposed to the  $T_{m1}T_{m1}$  condition because of the significant disadvantage of the latter participants in retrieval and decoding of M2 (i.e., the unpracticed group of mediators).

The results from Experiment 1 clearly reveal how and why total target performance was significantly higher under the  $T_{m1}T_{m2}$  condition than under the  $T_{m1}T_{m1}$  condition. However, in the current experiment, each cue included two mediators; thus, these two mediators competed with each other when the participants were instructed to report one mediator corresponding to the cue. Therefore, the design of the CMR test may induce unfair competition for M2 under the  $T_{m1}T_{m1}$  condition and for M1 under the  $T_{m1}T_{m2}$  condition, and this bias may influence the participants' mediator retrieval performance, thereby influencing mediator decoding performance. First, in contrast to the participants exposed to the  $T_{m1}T_{m2}$  condition, the participants exposed to the  $T_{m1}T_{m1}$  condition never practiced M2. Therefore, the probability that M2 would be retrieved was likely lower for the latter participants. However, this did not mean that M2 was not remembered. At a certain level of decoding, more targets would be recalled if more constituents of M2 were remembered. Second, in contrast to the  $T_{m1}T_{m1}$  condition, the  $T_{m1}T_{m2}$  condition involved practicing M1 only once; as a consequence, there was more intense competition between M1 and M2, which was also practiced once under the  $T_{m1}T_{m2}$  condition. Because of this increased competition under the  $T_{m1}T_{m2}$  condition, the retrieval and decoding probabilities of M1 decreased. However, this did not mean that M1 was not remembered or could not be used to elicit the target, as indicated by the result that there was no difference in M1 decoding between the two conditions in Experiment 1. As noted above, to overcome this limitation, Experiment 2 was designed.

### 3. Experiment 2

In Experiment 2, participants were instructed to recall the target from the given mediator during the final test phase. Once one could remember the target using the mediator, that mediator was considered to have been successfully decoded to the target. The participants were divided into four groups: two groups separately used M1 or M2 to recall the target under the  $T_{m1}T_{m1}$  condition, and the other two groups separately used M1 or M2 to recall the target under the  $T_{m1}T_{m2}$  condition (see Fig. 1). It was expected that the proportion of targets recalled from M1 under the  $T_{m1}T_{m1}$  condition would not be different from the proportion of targets recalled from M1 or M2 under the  $T_{m1}T_{m2}$  condition. However, the proportion of targets recalled from M2 under the  $T_{m1}T_{m1}$  condition was hypothesized to be significantly lower than the

proportion of targets recalled from M1 or M2 under the  $T_{m1}T_{m2}$  condition.

#### 3.1. Method

##### 3.1.1. Participants

Overall, 103 students (33 males, 70 females) from Beijing Normal University participated in Experiment 2 for a reward. The participants were randomly assigned to the four practice groups. Twenty-five participants (18 females; mean age = 20.75;  $SD = 1.76$ ) were exposed to the  $T_{m1}T_{m1}$ -M1 condition; twenty-four participants (17 females; mean age = 22.29;  $SD = 2.61$ ) were exposed to the  $T_{m1}T_{m1}$ -M2 condition; twenty-eight participants (18 females; mean age = 21.86;  $SD = 2.66$ ) were exposed to the  $T_{m1}T_{m2}$ -M1 condition; and twenty-six participants (17 females; mean age = 22.69;  $SD = 1.78$ ) were exposed to the  $T_{m1}T_{m2}$ -M2 condition. The results from the initial study phase for one  $T_{m1}T_{m2}$ -M2 participant were missing due to computer error, as were the results from the final test phase for one  $T_{m1}T_{m1}$ -M1 participant. All participants had normal or corrected vision, were native speakers of Chinese, and had no previous experience with Japanese.

##### 3.1.2. Materials

All materials were the same as those used in Experiment 1.

##### 3.1.3. Procedure

The initial study and retrieval practice phases were identical to those used in Experiment 1. On the final test, the participants in Experiment 2 received only one of the mediators to recall the target, without receiving a prompt of the Japanese cue word.

##### 3.1.4. Analysis

The data analyses for the first two phases were identical to those used in Experiment 1. Additionally, mediator decoding and the response time of mediator decoding during the final test phase were analysed. The time from the appearance of the mediator to the reporting of the target was defined as the response time of mediator decoding and was recorded using SRBox.

#### 3.2. Results

##### 3.2.1. Initial study phase

The average mediator helpfulness rating and learning time are presented in Table 4. The data were analysed via  $2$  (retrieval practice:  $T_{m1}T_{m1}$  or  $T_{m1}T_{m2}$ )  $\times$   $2$  (mediator used in the final test: M1 or M2)  $\times$   $2$  (study list: S1 or S2) mixed ANOVA. For the mediator helpfulness rating, there were main effects of retrieval practice,  $F(1, 98) = 4.820$ ,  $MSE = 1.844$ ,  $p = 0.030$ ,  $\eta_p^2 = 0.047$ , and study list,  $F(1, 98) = 4.264$ ,  $MSE = 0.218$ ,  $p = 0.042$ ,  $\eta_p^2 = 0.042$ , and their interaction was significant,  $F(1, 98) = 6.044$ ,  $MSE = 0.308$ ,  $p = 0.016$ ,  $\eta_p^2 = 0.058$ . None of these significant effects interacted with effect of the mediator used in the final test ( $F_s < 1$ ), and there was no main effect of the mediator used in the final test ( $F < 1$ ). Further analysis showed that the mediator helpfulness rating was significantly higher for S2 than for S1 under the  $T_{m1}T_{m1}$  condition,  $F(1, 98) = 9.858$ ,  $MSE = 0.50$ ,  $p = 0.002$ ,  $\eta_p^2 = 0.091$ . However, there was no significant difference in the mediator helpfulness rating between S1 and S2 under the  $T_{m1}T_{m2}$  condition,  $F < 1$ . Moreover, there was no significant difference in the mediator helpfulness rating for S1 between the  $T_{m1}T_{m1}$  and  $T_{m1}T_{m2}$  conditions,  $F(1, 98) = 1.631$ ,  $MSE = 0.322$ ,  $p = 0.205$ ,  $\eta_p^2 = 0.016$ . However, the mediator helpfulness rating for S2 was significantly higher under the  $T_{m1}T_{m1}$  condition than under the  $T_{m1}T_{m2}$  condition,  $F(1, 98) = 7.751$ ,  $MSE = 1.830$ ,  $p = 0.006$ ,  $\eta_p^2 = 0.073$ . For learning time, the only significant main effect was study list,  $F(1, 98) = 7.163$ ,  $MSE = 31.036$ ,  $p = 0.009$ ,  $\eta_p^2 = 0.068$ , and the interaction of study list with retrieval practice was significant,  $F(1, 98) = 5.940$ ,  $MSE = 25.735$ ,  $p = 0.017$ ,  $\eta_p^2 = 0.057$ . Further analysis showed that learning

**Table 5**

Retrieval performance in the four conditions in Experiment 2. Standard deviations are presented in parentheses.

Condition	T1 <sub>c</sub> T2 <sub>c</sub>	T1 <sub>c</sub> T2 <sub>N</sub>	T1 <sub>N</sub> T2 <sub>c</sub>	T1 <sub>N</sub> T2 <sub>N</sub>
T <sub>m1</sub> T <sub>m1</sub> -M1	0.77 (0.20)	0.02 (0.03)	0.03 (0.04)	0.19 (0.18)
T <sub>m1</sub> T <sub>m1</sub> -M2	0.74 (0.26)	0.01 (0.03)	0.03 (0.04)	0.22 (0.22)
T <sub>m1</sub> T <sub>m2</sub> -M1	0.79 (0.19)	0.05 (0.07)	0.04 (0.05)	0.12 (0.14)
T <sub>m1</sub> T <sub>m2</sub> -M2	0.73 (0.24)	0.06 (0.07)	0.06 (0.07)	0.15 (0.19)

Note: All totals do not add up to 100% because the numbers were rounded.

time was significantly longer for S1 than for S2 under the T<sub>m1</sub>T<sub>m1</sub> condition,  $F(1, 98) = 12.598$ ,  $MSE = 0.421$ ,  $p = 0.001$ ,  $\eta_p^2 = 0.114$ . However, there was no significant difference in learning time between S1 and S2 under the T<sub>m1</sub>T<sub>m2</sub> condition,  $F < 1$ . Furthermore, there was no significant difference in learning time for either S1 or S2 between the T<sub>m1</sub>T<sub>m1</sub> and T<sub>m1</sub>T<sub>m2</sub> conditions,  $F_s < 1.188$ .

### 3.2.2. Retrieval practice phase

The retrieval performance<sup>4</sup> of each possible outcome is listed in Table 5. For the items that were successfully retrieved at least once (T1<sub>c</sub>T2<sub>c</sub> items + T1<sub>c</sub>T2<sub>N</sub> items + T1<sub>N</sub>T2<sub>c</sub> items), a 2 (retrieval practice: T<sub>m1</sub>T<sub>m1</sub> or T<sub>m1</sub>T<sub>m2</sub>) × 2 (mediator used in the final test: M1 or M2) ANOVA revealed only a marginally significant effect of retrieval practice,  $F(1, 99) = 3.517$ ,  $MSE = 56.431$ ,  $p = 0.064$ ,  $\eta_p^2 = 0.034$ . There was no significant effect of mediator used or any interaction,  $F_s < 1$ . For the items that were successfully retrieved twice (T1<sub>c</sub>T2<sub>c</sub> items), there was no significant main effect of the mediator used,  $F(1, 99) = 1.146$ ,  $MSE = 27.346$ ,  $p = 0.287$ ,  $\eta_p^2 = 0.011$ , or the retrieval practice or any interaction between these two factors,  $F_s < 1$ .

### 3.2.3. Final test phase

Final test performance is shown in Table 6. First, mediator decoding of those items which were successfully retrieved twice during the retrieval practice phase (T1<sub>c</sub>T2<sub>c</sub> items) was analysed. A 2 (retrieval practice: T<sub>m1</sub>T<sub>m1</sub> or T<sub>m1</sub>T<sub>m2</sub>) × 2 (mediator used in the final test: M1 or M2) ANOVA yielded a significant main effect of retrieval practice,  $F(1, 98) = 13.498$ ,  $MSE = 0.483$ ,  $p < 0.001$ ,  $\eta_p^2 = 0.121$ , with greater mediator decoding under the T<sub>m1</sub>T<sub>m2</sub> condition than under the T<sub>m1</sub>T<sub>m1</sub> condition. This analysis also revealed a main effect of the mediator used in the final test,  $F(1, 98) = 33.809$ ,  $MSE = 1.210$ ,  $p < 0.001$ ,  $\eta_p^2 = 0.257$ , with greater mediator decoding using M1 than using M2. More importantly, the interaction between these factors was significant,  $F(1, 98) = 21.766$ ,  $MSE = 0.779$ ,  $p < 0.001$ ,  $\eta_p^2 = 0.182$ . Further tests to examine differences in final test performance according to mediator used between the T<sub>m1</sub>T<sub>m1</sub> and T<sub>m1</sub>T<sub>m2</sub> conditions were conducted. This analysis showed a significant difference in test performance using M2,  $F(1, 98) = 34.177$ ,  $MSE = 1.223$ ,  $p < 0.001$ ,  $\eta_p^2 = 0.259$ , but not M1,  $F < 1$ , between the T<sub>m1</sub>T<sub>m1</sub> and T<sub>m1</sub>T<sub>m2</sub> conditions. Decoding performance under the T<sub>m1</sub>T<sub>m2</sub>-M2 condition was significantly higher than that under the T<sub>m1</sub>T<sub>m1</sub>-M2 condition. Additionally, a significant difference in performance under the T<sub>m1</sub>T<sub>m1</sub> condition,  $F(1, 98) = 51.897$ ,  $MSE = 1.857$ ,  $p < 0.001$ ,  $\eta_p^2 = 0.346$ , but not the T<sub>m1</sub>T<sub>m2</sub> condition,  $F < 1$ , was observed between M1 and M2 conditions. Decoding performance under the T<sub>m1</sub>T<sub>m1</sub>-M1 condition was significantly higher than that under the T<sub>m1</sub>T<sub>m1</sub>-M2 condition.

In addition, the response time of mediator decoding was recorded in Experiment 2. A 2 (retrieval practice: T<sub>m1</sub>T<sub>m1</sub> or T<sub>m1</sub>T<sub>m2</sub>) × 2 (mediator used in the final test: M1 or M2) ANOVA showed no main effect of the mediator used in the final test,  $F(1, 95) = 1.738$ ,  $MSE = 6.716$ ,  $p = 0.191$ ,  $\eta_p^2 = 0.018$ . The main effect of retrieval practice was marginally

<sup>4</sup> Target retrieval was regarded as a correct answer with a correct sentence. The probabilities of providing a correct target but no sentence during sessions 1 and 2 were 4.17% and 4.17%, respectively, under the T<sub>m1</sub>T<sub>m1</sub>-M1 condition; 8.14% and 7.39%, respectively, under the T<sub>m1</sub>T<sub>m1</sub>-M2 condition; 1.95% and 3.73%, respectively, under the T<sub>m1</sub>T<sub>m2</sub>-M1 condition; and 2.10% and 2.97%, respectively, under the T<sub>m1</sub>T<sub>m2</sub>-M2 condition.

**Table 6**

Final test performance for T1<sub>c</sub>T2<sub>c</sub> items in Experiment 2. Standard deviations are presented in parentheses.

Condition	Mediator decoding	Response time of mediator decoding (s)
T <sub>m1</sub> T <sub>m1</sub> -M1	0.83 (0.11)	3.44 (1.15)
T <sub>m1</sub> T <sub>m1</sub> -M2	0.44 (0.26)	5.10 (2.86)
T <sub>m1</sub> T <sub>m2</sub> -M1	0.79 (0.14)	3.84 (1.84)
T <sub>m1</sub> T <sub>m2</sub> -M2	0.75 (0.21)	3.22 (1.72)

Note: For those items which were successfully retrieved twice during retrieval practice phase (T1<sub>c</sub>T2<sub>c</sub> items), index of the final test and its computational formula were defined as follows. Mediator decoding: the number of mediator that correctly elicited the target in the final test for T1<sub>c</sub>T2<sub>c</sub> items divided by the number of T1<sub>c</sub>T2<sub>c</sub> items.

significant,  $F(1, 95) = 3.497$ ,  $MSE = 13.509$ ,  $p = 0.065$ ,  $\eta_p^2 = 0.036$ , and the interaction between these two factors was significant,  $F(1, 95) = 8.187$ ,  $MSE = 31.628$ ,  $p = 0.005$ ,  $\eta_p^2 = 0.079$ . Further analysis showed a significant difference in performance using M2,  $F(1, 95) = 10.878$ ,  $MSE = 42.022$ ,  $p = 0.001$ ,  $\eta_p^2 = 0.103$ , but not M1,  $F < 1$ , between the T<sub>m1</sub>T<sub>m1</sub> and T<sub>m1</sub>T<sub>m2</sub> conditions. The mediator decoding response time was significantly longer under the T<sub>m1</sub>T<sub>m1</sub>-M2 condition than under the T<sub>m1</sub>T<sub>m2</sub>-M2 condition. Additionally, a significant difference in performance under the T<sub>m1</sub>T<sub>m1</sub> condition,  $F(1, 95) = 8.011$ ,  $MSE = 30.947$ ,  $p = 0.006$ ,  $\eta_p^2 = 0.078$ , but not the T<sub>m1</sub>T<sub>m2</sub> condition,  $F(1, 95) = 1.309$ ,  $MSE = 5.055$ ,  $p = 0.256$ ,  $\eta_p^2 = 0.014$ , was observed between the M1 and M2 conditions. The decoding response time was significantly longer under the T<sub>m1</sub>T<sub>m1</sub>-M2 condition than under the T<sub>m1</sub>T<sub>m1</sub>-M1 condition. The results of all 22 items that were initially studied are shown in Appendix 2.<sup>5</sup>

### 3.3. Discussion

During the initial study phase, the interaction of study list with retrieval practice was significant for both mediator helpfulness rating and learning time. A simple effect analysis that was conducted to examine any possible differences between the study lists indicated that the participants exposed to the T<sub>m1</sub>T<sub>m1</sub> condition tended to spend more time on S1 than on S2 and to provide a lower helpfulness rating for S1 than for S2. Considering the reconstruction and counterbalancing of the mediator groups used during the retrieval practice phase, the differences across study lists during the initial study phase may not have influenced the results during the retrieval practice and final test phases. Additionally, a simple effect analysis that was conducted to examine any possible differences between retrieval practice conditions indicated that the participants exposed to the T<sub>m1</sub>T<sub>m1</sub> condition tended to report a higher helpfulness rating for S2 than the participants exposed to the T<sub>m1</sub>T<sub>m2</sub> condition. However, there was no significant difference in the learning time of S2 between the T<sub>m1</sub>T<sub>m1</sub> and T<sub>m1</sub>T<sub>m2</sub> conditions. If the difference in mediator helpfulness rating for S2 between retrieval practice conditions influenced subsequent memory performance, then this difference would be more likely to improve performance under the T<sub>m1</sub>T<sub>m1</sub> condition than under the T<sub>m1</sub>T<sub>m2</sub> condition.

During the retrieval practice phase, the proportion of items that were successfully retrieved at least once was marginally significantly higher under the T<sub>m1</sub>T<sub>m2</sub> condition than under the T<sub>m1</sub>T<sub>m1</sub> condition. To examine the possible reason for this result, analyses of the proportion of items that were successfully retrieved only once, the proportion of items that were successfully retrieved during the first retrieval session and the proportion of items that were successfully retrieved during the second retrieval session were conducted to compare the T<sub>m1</sub>T<sub>m1</sub> and T<sub>m1</sub>T<sub>m2</sub> conditions. The results were consistent with those from

<sup>5</sup> When all 22 items were included in the analysis (see the Appendix 2), the pattern of results was the same as the results from analysis of only items that were successfully retrieved twice during the retrieval practice phase, except that the main effect of retrieval practice on the response time of mediator decoding was no longer significant,  $F < 1.96$ . Further details can be found in the supplementary materials.

Experiment 1.<sup>6</sup> More importantly, the proportion of items that were successfully retrieved twice did not differ between the conditions, and this parameter served as the basis of the final analysis.

For the final test phase, the results suggested that mediator decoding was significantly greater under the  $T_{m1}T_{m2}$ -M2 condition than under the  $T_{m1}T_{m1}$ -M2 condition; and the results of mediator decoding response time were reversed. Therefore, although M2 was presented in the final test to avoid potential bias, decoding performance using M2 was indeed impaired under the  $T_{m1}T_{m1}$  condition compared with the  $T_{m1}T_{m2}$  condition due to a lack of retrieval practice using M2. However, there was no significant difference in mediator decoding performance or response time between the  $T_{m1}T_{m2}$ -M1 and  $T_{m1}T_{m1}$ -M1 conditions. Thus, when M1 was presented in the final test under all conditions to avoid potential bias, decoding performance using M1 did not significantly differ between the  $T_{m1}T_{m2}$  and  $T_{m1}T_{m1}$  conditions. These results of Experiment 2 were consistent with the results of Experiment 1.

#### 4. General discussion

The current study compared the mnemonic benefit of repeated retrieval by consolidating two retrieval routes with consolidating a single retrieval route twice. In Experiment 1, both for only those items which were successfully retrieved twice during the retrieval practice phase ( $T1_cT2_c$  items) and for all 22 items that were initially studied, the increase in total target performance under the  $T_{m1}T_{m2}$  condition demonstrated that practicing more retrieval routes leads to greater memory retention. The results of Experiment 1 revealed that the deficiency of  $T_{m1}T_{m1}$  condition in promoting memory was due to a significant disadvantage in mediator retrieval and decoding using the unpracticed mediator (M2). Even when those mediators were presented in the final test in Experiment 2, mediator decoding of M2 under the  $T_{m1}T_{m1}$  condition remained the lowest, and the corresponding response time was the longest. Thus, compared with consolidating a single retrieval route twice, consolidating two retrieval routes enhanced memory retention.

In the current study, in comparison with repeated retrieval by consolidating a single retrieval route twice, repeated retrieval by consolidating two retrieval routes increased the retrieval variability during the retrieval practice phase. These results demonstrated that increasing the variability of retrieval practice by adding retrieval routes highly efficiently promoted long-term memory retention; this finding was consistent with the report by Finley (2012). In Finley (2012, Experiment 3), participants were instructed to recall targets simultaneously using two retrieval cues, either with identical meanings or distinct meanings, that they had learned during the initial study phase. That result showed that double-meaning retrieval cues indeed yielded higher recall than single-meaning retrieval cues. Therefore, increasing the variability of retrieval practice indeed greatly improved the likelihood of the present successful retrieval. In addition, the current study used two separate groups of mediators in the two sessions of retrieval practice, providing further evidence for the mnemonic benefit of retrieval variability to performance on a delay test.

Why would repeated retrieval by consolidating two retrieval routes be more beneficial for long-term memory retention than consolidating a single retrieval route twice? Two possible reasons might explain these results. First, the episodic context account of retrieval-based

learning proposed by Karpicke, Lehman and Aue (2014) and Lehman et al. (2014) provides a perspective on the present results. According to the episodic context account (Karpicke, Lehman & Aue, 2014), during the retrieval practice phase, participants attempt to reinstate the temporal context associated with an item and then use reinstated temporal context features to guide their retrieval search. Thus, past context (context A) and present context (context B) should be integrated via successful retrieval, and the subsequent context representation related to the item should be updated. This process allows people to restrict their memory search when subsequently retrieving the target. It can be inferred that retrieval variability increased long-term memory retention by providing an additional opportunity to produce a more unique temporal context for subsequent recall of an item. However, this is a theoretical inference that requires evidence from empirical research. In the current study, participants exposed to the  $T_{m1}T_{m1}$  condition practiced a single mediator twice, while participants exposed to the  $T_{m1}T_{m2}$  condition practiced two mediators once. For those items which were successfully retrieved twice during the retrieval practice, the context representation was reinstated and updated for only M1 under the  $T_{m1}T_{m1}$  condition but for both M1 and M2 under the  $T_{m1}T_{m2}$  condition. As hypothesized, in comparison with repeated practice consolidating one retrieval route twice, repeated practice consolidating two retrieval routes produced a more unique temporal context cue by increasing retrieval variability, thereby improving long-term memory retention. Thus, the results of the current study support and extend the episodic context account of retrieval-based learning.

Another possible explanation involves the elaborative retrieval hypothesis proposed by Carpenter (2009) and Carpenter and DeLosh (2006). The number of retrieval routes labelled by mediators, which may be the same as the number of mediators labelled to a retrieval route, is an important form of elaboration. Retrieval practice using different mediators enabled participants to access multiple retrieval routes; subsequently, any recovered route would lead to successful memory recall (Carpenter, 2009). In the present study, the  $T_{m1}T_{m2}$  condition constructed two retrieval routes, both of which were used once to successfully access the target. Therefore, either pathway, when recovered, would facilitate subsequent recall. Under the  $T_{m1}T_{m1}$  condition, although one retrieval route was consolidated via two sessions of retrieval practice, indeed further enhancing its effectiveness, if the mediator was unavailable, there was a higher probability of recall failure on the final test. The finding of a deficiency in M2 retrieval and decoding in both experiments specifically emphasized the importance of multiple forms of elaborative information.

However, these two hypotheses are not mutually exclusive (Rawson, Vaughn, & Carpenter, 2015). Semantic elaboration could have two possible effects. In some cases, semantic elaboration during the retrieval practice phase may produce an effect of cue overload (Lehman et al., 2014; Surprenant & Neath, 2009). The mediator generated by the participants may be linked to the cue or to the target, which itself may become a competitor that floods the search set and ultimately reduces access to the target. For instance, Lehman et al.'s (2014) elaboration task involved typing the first two words that came to mind when a target word was presented. The observed decrease in target recall and increase in prior-list intrusions demonstrated that this type of elaboration indeed produced cue overload by increasing the number of candidates in the search set (Lehman et al., 2014). In contrast, when mediator retrieval is practiced to successfully establish the relationship of the cue to the target, semantic elaboration improves memory retention. As proposed in the episodic context account, elaborative information may be integrated as a unique contextual cue via successful retrieval, and the integration of information that is specific to the target may further restrict the size of the search set, ultimately increasing access to the target. In the current study, the manipulation of semantic elaboration may induce the latter effect.

In fact, these two possible consequences of semantic elaboration may simultaneously occur in one experiment, although with varying

<sup>6</sup> The proportion of items successfully retrieved only once was indeed significantly higher under the  $T_{m1}T_{m2}$  condition ( $M = 0.10$ ;  $SD = 0.10$ ) than under the  $T_{m1}T_{m1}$  condition, ( $M = 0.04$ ;  $SD = 0.05$ ),  $F(1, 101) = 15.224$ ,  $MSE = 0.098$ ,  $p < 0.001$ ,  $\eta_p^2 = 0.131$ . There was no significant difference in the proportion of items successfully retrieved during the first retrieval session,  $F(1, 101) = 1.420$ ,  $MSE = 0.058$ ,  $p = 0.235$ ,  $\eta_p^2 = 0.014$ , in which M1 was presented under both conditions, between the  $T_{m1}T_{m1}$  and  $T_{m1}T_{m2}$  conditions (see Table 5,  $T1_cT2_c$  items +  $T1_cT2_N$  items). Furthermore, no significant difference in the proportion of items successfully retrieved during the second retrieval session,  $F(1, 101) = 0.427$ ,  $MSE = 0.017$ ,  $p = 0.515$ ,  $\eta_p^2 = 0.004$ , in which different mediators were presented for each condition, between the two conditions (see Table 5,  $T1_cT2_c$  items +  $T1_NT2_c$  items).



degrees. As demonstrated in both experiments of this study, mediator decoding of M1 under the  $T_{m1}T_{m1}$  condition ( $M_{Exp1} = 0.95$ ;  $M_{Exp2} = 0.83$ ) was slightly greater than mediator decoding of M1 ( $M_{Exp1} = 0.91$ ;  $M_{Exp2} = 0.79$ ) or M2 ( $M_{Exp1} = 0.92$ ;  $M_{Exp2} = 0.75$ ) under the  $T_{m1}T_{m2}$  condition. Considering only the effectiveness of the mediator as a cue on the final test, we inferred that there may be mild cue overload under the  $T_{m1}T_{m2}$  condition. That is to say, combining the results of M1 retrieval from Experiment 1 and the results of M1 decoding from Experiments 1 and 2, the  $T_{m1}T_{m1}$  condition indeed increased the diagnostic value of the retrieval route using M1 (the effectiveness of M1 as a retrieval cue on the final test) (Karpicke & Smith, 2012; Nairne, 2002; Raaijmakers & Shiffrin, 1981). Otherwise, repeated consolidation of a single retrieval route may produce a decontextualization process in which the target becomes more retrievable and is no longer exclusively associated with the specific retrieval route (Karpicke, Lehman & Aue, 2014), although this effect represented a small portion of the total performance. The data supported this hypothesis. Concerning target performance without successfully retrieving a mediator, the participants exposed to the  $T_{m1}T_{m1}$  condition ( $M = 0.13$ ,  $SD = 0.12$ ) were more likely to recall targets without mediator retrieval than those exposed to the  $T_{m1}T_{m2}$  condition ( $M = 0.07$ ,  $SD = 0.07$ ),  $F(1, 62) = 5.408$ ,  $MSE = 0.047$ ,  $p = 0.023$ ,  $\eta_p^2 = 0.080$ . Specifically, this aspect of target performance includes successful recall via other mediators that participants themselves generated or via no mediator. The  $T_{m1}T_{m1}$  condition was more likely to induce target recall via other mediators (0.09 vs. 0.06,  $F(1, 62) = 3.141$ ,  $MSE = 0.019$ ,  $p = 0.081$ ,  $\eta_p^2 = 0.048$ ) or no mediator (0.04 vs. 0.02,  $F(1, 62) = 2.538$ ,  $MSE = 0.007$ ,  $p = 0.116$ ,  $\eta_p^2 = 0.039$ ) than the  $T_{m1}T_{m2}$  condition. This result supported that consolidating one retrieval route may induce decontextualization (Karpicke, Lehman & Aue, 2014).

However, the response time of mediator decoding under the  $T_{m1}T_{m1}$ -M1 condition did not appear to differ from that of the  $T_{m1}T_{m2}$ -

M1 or the  $T_{m1}T_{m2}$ -M2 condition. The response time results demonstrated that no inhibitory effect between the two mediators was present under the  $T_{m1}T_{m2}$  condition. Therefore, considering efficient memory retention, using multiple retrieval routes significantly masked any potential cue overload and provided significantly greater mnemonic value than using a single retrieval route. Moreover, consolidating two retrieval routes allowed participants to establish a more unique context for target retrieval. In other words, this type of elaboration facilitated the formation to form a more unique contextual cue with higher diagnostic value during retrieval practice. The present results indicate a clear conclusion: consolidating multiple retrieval routes is more beneficial in improving long-term retention than consolidating a single retrieval route. Retrieval variability during the retrieval practice phase play a significant role in memory retention. When items are repeatedly retrieved in multiple temporal contexts, they become associated with a variety of contextual features that serve as effective retrieval cues on subsequent tests (Karpicke, Lehman & Aue, 2014). Furthermore, elaborative information may facilitate the construction of this contextual variability. Under the proper operation, different forms of elaborative information may not restrain each other. However, the current study only examined the circumstance of two mediators. Further research should investigate retrieval-based learning using additional retrieval routes and should determine whether inferences appear upon increasing the number of practiced mediators.

**Funding**

This study was supported by the National Natural Science Foundation of China (31100817, 31000506), Humanity and Social Science Foundation of China University of Political Science and Law (15ZFG18001) and the Training Program Foundation for the Excellent Youth Scholars of China University of Political Science and Law.

**Appendix 1**

Final test performance of all items initial studied in Experiment 1. Standard deviations are presented in parentheses.

Condition	Total target performance	Target performance with mediator retrieval	Target recalled via		Mediator retrieval		Mediator decoding	
			M1	M2	M1	M2	M1	M2
$T_{m1}T_{m1}$	0.68 (0.19)	0.58 (0.22)	0.48 (0.20)	0.10 (0.07)	0.54 (0.18)	0.15 (0.09)	0.87 (0.16)	0.71 (0.29)
$T_{m1}T_{m2}$	0.78 (0.14)	0.71 (0.18)	0.36 (0.12)	0.35 (0.13)	0.42 (0.11)	0.41 (0.14)	0.86 (0.14)	0.86 (0.15)

Note: For all 22 items that were initially studied, indexes of the final test and their computational formulas were defined as follows: target performance: the number of targets that were correctly recalled in the final test divided by 22; target performance with mediator retrieval: the number of targets that were correctly recalled via M1 or M2 in the final test divided by 22; M1 (or M2) retrieval: the number of mediators in M1 (or M2) that were correctly recalled in the final test divided by 22; M1 (or M2) decoding: the number of mediators in M1 (or M2) that correctly elicited the targets in the final test divided by the number of mediators in M1 (or M2) that were correctly recalled in the final test.

**Appendix 2**

Final test performance of all items in Experiment 2. Standard deviations are presented in parentheses.

Condition	Mediator decoding	Response time of mediator decoding (s)
$T_{m1}T_{m1}$ -M1	0.70 (0.18)	3.40 (1.05)
$T_{m1}T_{m1}$ -M2	0.38 (0.25)	4.74 (2.65)
$T_{m1}T_{m2}$ -M1	0.70 (0.19)	3.78 (1.67)
$T_{m1}T_{m2}$ -M2	0.66 (0.27)	3.31 (1.83)

Note: For all 22 items that were initial studied, the index of the final test and its computational formulas were defined as follows: mediator decoding: the number of mediator that correctly elicited the target in the final test divided by 22.

**Appendix 3. Supplementary data**

Supplementary data to this article can be found online at <http://dx.doi.org/10.1016/j.actpsy.2016.05.014>.

## References

- Abott, E. E. (1909). On the analysis of the factor of recall in the learning process. *Psychological Review*, 11(1), 159–177. <http://dx.doi.org/10.1037/h0093018>.
- Bjork, R. A. (1975). Retrieval as a memory modifier: An interpretation of negative recency and related phenomena. In R. L. Solso (Ed.), *Information processing and cognition: The Loyola symposium* (pp. 123–144). Hillsdale, NJ: Erlbaum.
- Carpenter, S. K. (2009). Cue strength as a moderator of the testing effect: The benefits of elaborative retrieval. *Journal of Experimental Psychology*, 35(6), 1563–1569. <http://dx.doi.org/10.1037/a0017021>.
- Carpenter, S. K. (2011). Semantic information activated during retrieval contributes to later retention: Support for the mediator effectiveness hypothesis of the testing effect. *Journal of Experimental Psychology*, 37(6), 1547–1552. <http://dx.doi.org/10.1037/a0024140>.
- Carpenter, S. K., & DeLosh, E. L. (2006). Impoverished cue support enhances subsequent retention: Support for the elaborative retrieval explanation of the testing effect. *Memory & Cognition*, 34(2), 268–276. <http://dx.doi.org/10.3758/bf03193405>.
- Finley, J. R. (2012). *Retrieval cue variability: When and why are two meanings better than one?* (Doctoral dissertation) University of Illinois at Urbana-Champaign.
- Gartman, L. M., & Johnson, N. F. (1972). Massed versus distributed repetition of homographs: A test of the differential-encoding hypothesis. *Journal of Verbal Learning and Verbal Behavior*, 11(6), 801–808. [http://dx.doi.org/10.1016/S0022-5371\(72\)80016-1](http://dx.doi.org/10.1016/S0022-5371(72)80016-1).
- Glanzer, M., & Duarte, A. (1971). Repetition between and within languages in free recall. *Journal of Verbal Learning and Verbal Behavior*, 10(6), 625–630. [http://dx.doi.org/10.1016/S0022-5371\(71\)80069-5](http://dx.doi.org/10.1016/S0022-5371(71)80069-5).
- Greenberg, D. L., & Verfaellie, M. (2010). Effects of fixed- and varied-context repetition on associative recognition in amnesia. *Journal of the International Neuropsychological Society*, 16(4), 596–602. <http://dx.doi.org/10.1017/S1355617710000287>.
- Hintzman, D. L., & Stern, L. D. (1978). Contextual variability and memory for frequency. *Journal of Experimental Psychology: Human Learning and Memory*, 4(5), 539–549. <http://dx.doi.org/10.1037/0278-7393.4.5.539>.
- Karpicke, J. D., & Roediger, H. L., III (2007). Expanding retrieval practice promotes short-term retention, but equally spaced retrieval enhances long-term retention. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 33(4), 704–719. <http://dx.doi.org/10.1037/0278-7393.33.4.704>.
- Karpicke, J. D., & Roediger, H. L., III (2008). The critical importance of retrieval for learning. *Science*, 319(5865), 966–968. <http://dx.doi.org/10.1126/science.1152408>.
- Karpicke, J. D., & Smith, M. A. (2012). Separate mnemonic effects of retrieval practice and elaborative encoding. *Journal of Memory and Language*, 67(1), 17–29. <http://dx.doi.org/10.1016/j.jml.2012.02.004>.
- Karpicke, J. D., & Zaromb, F. M. (2010). Retrieval mode distinguishes the testing effect from the generation effect. *Journal of Memory and Language*, 62(3), 227–239. <http://dx.doi.org/10.1016/j.jml.2009.11.010>.
- Karpicke, J. D., Lehman, M., & Aue, W. R. (2014). Retrieval-based learning: An episodic context account. In B. H. Ross (Ed.), *Psychology of learning and motivation*, vol. 61, vol. 61. (pp. 237–284). <http://dx.doi.org/10.1016/B978-0-12-800283-4.00007-1>.
- Karpicke, J. D., Lehman, M., & Gallo, D. A. (2014). Retrieval-Based Learning: The role of initial retrieval orientation. Unpublished manuscript.
- Landauer, T. K., & Bjork, R. A. (1978). Optimum rehearsal patterns and name learning. In M. Fruneberg, P. E. Morris, & R. N. Sykes (Eds.), *Practical aspects of memory*, 1 (pp. 625–632). London: Academic Press.
- Lehman, M., Smith, M. A., & Karpicke, J. D. (2014). Toward an episodic context account of retrieval-based learning: Dissociating retrieval practice and elaboration. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 40(6), 1787–1794. <http://dx.doi.org/10.1037/xlm0000012>.
- Liu, W. (1984). *Concise Japanese-Chinese dictionary*. Beijing: The Commercial Press.
- Nairne, J. S. (2002). The myth of the encoding-retrieval match. *Memory*, 10(5–6), 389–395. <http://dx.doi.org/10.1080/09658210244000216>.
- Pyc, M. A., & Rawson, K. A. (2010). Why testing improves memory: Mediator effectiveness hypothesis. *Science*, 330(6002), 335. <http://dx.doi.org/10.1126/science.1191465>.
- Raaijmakers, J. G., & Shiffrin, R. M. (1981). Search of associative memory. *Psychological Review*, 88(2), 93–134. <http://dx.doi.org/10.1037//0033-295x.88.2.93>.
- Rawson, K. A., & Dunlosky, J. (2011). Optimizing schedules of retrieval practice for durable and efficient learning: How much is enough? *Journal of Experimental Psychology: General*, 140(3), 283–302. <http://dx.doi.org/10.1037/a0023956>.
- Rawson, K. A., Vaughn, K. E., & Carpenter, S. K. (2015). Does the benefit of testing depend on lag, and if so, why? Evaluating the elaborative retrieval hypothesis. *Memory & Cognition*, 43(4), 619–633. <http://dx.doi.org/10.3758/s13421-014-0477-z>.
- Roediger, H. L., & Butler, A. C. (2011). The critical role of retrieval practice in long-term retention. *Trends in Cognitive Sciences*, 15(1), 20–27. <http://dx.doi.org/10.1016/j.tics.2010.09.003>.
- Roediger, H. L., & Karpicke, J. D. (2006). Test-enhanced learning: Taking memory tests improves long-term retention. *Psychological Science*, 17(3), 249–255. <http://dx.doi.org/10.1111/j.1467-9280.2006.01693.x>.
- Smith, S. M., Glenberg, A. M., & Bjork, R. A. (1978). Environmental context and human memory. *Memory & Cognition*, 6(4), 342–353. <http://dx.doi.org/10.3758/BF03197465>.
- Surprenant, A. M., & Neath, I. (2009). *Principles of memory*. New York: Psychology Press.
- Wang, P. (2000). *Modern Japanese-Chinese and Chinese-Japanese dictionary*. Beijing: Foreign Language Teaching and Research Press.