When does Retrieval Induce Forgetting and When does it Induce Facilitation? Implications for Retrieval Inhibition, Testing Effect, and Text Processing

Jason C.K. Chan
*Iowa State University, ckchan@iastate.edu*

Follow this and additional works at: [http://lib.dr.iastate.edu/psychology_pubs](http://lib.dr.iastate.edu/psychology_pubs)

Part of the [Cognitive Psychology Commons](http://lib.dr.iastate.edu/cognitive_psychology), [Educational Assessment, Evaluation, and Research Commons](http://lib.dr.iastate.edu/assessment_evaluation), and the [Educational Psychology Commons](http://lib.dr.iastate.edu/education_psychology)

The complete bibliographic information for this item can be found at [http://lib.dr.iastate.edu/psychology_pubs/22](http://lib.dr.iastate.edu/psychology_pubs/22). For information on how to cite this item, please visit [http://lib.dr.iastate.edu/howtocite.html](http://lib.dr.iastate.edu/howtocite.html).
When does Retrieval Induce Forgetting and When does it Induce Facilitation? Implications for Retrieval Inhibition, Testing Effect, and Text Processing

Abstract
Retrieval practice can enhance long-term retention of the tested material (the testing effect), but it can also impair later recall of the nontested material – a phenomenon known as retrieval-induced forgetting (Anderson, M. C., Bjork, R. A., & Bjork, E. L. (1994). Remembering can cause forgetting: retrieval dynamics in long-term memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 20*(5), 1063–1087). Recent research, however, has shown that retrieval practice can sometimes improve later recall of the nontested material – a phenomenon termed retrieval-induced facilitation (Chan, J. C. K., McDermott, K. B., & Roediger, H. L. (2006). Retrieval-induced facilitation: initially nontested material can benefit from prior testing of related material. *Journal of Experimental Psychology: General, 135*, 553–571). What drives these different effects? Two experiments were designed to examine the conditions under which retrieval induces forgetting and facilitation. Two variables, the level of integration invoked during encoding and the length of delay between retrieval practice and final test, were revealed as critical factors in determining whether testing facilitated or hindered later retrieval of the nontested information. A text processing framework is advanced to account for these findings.

Keywords
Memory, testing effect, education, retrieval practice, retrieval-induced forgetting, retrieval induced facilitation, delay, integration, prose, situation model, discourse processing, retention interval, inhibition

Disciplines
Cognitive Psychology | Educational Assessment, Evaluation, and Research | Educational Psychology

Comments
NOTICE: this is the author's version of a work that was accepted for publication in Journal of Memory and Language. Changes resulting from the publishing process, such as peer review, editing, corrections, structural formatting, and other quality control mechanisms may not be reflected in this document. Changes may have been made to this work since it was submitted for publication. A definitive version was subsequently published in Journal of Memory and Language, [61, 2, (August 2009)] doi:10.1016/j.jml.2009.04.004.
When does Retrieval Induce Forgetting and When does it Induce Facilitation? Implications for Retrieval Inhibition, Testing Effect, and Text Processing

Jason C.K. Chan

Iowa State University of Science and Technology

Correspondence:

Jason C.K. Chan
Department of Psychology
W112 Lagomarcino Hall
Ames, IA 50011-3180
USA
ckchan@iastate.edu

NOTICE: this is the author’s version of a work that was accepted for publication in Journal of Memory and Language. Changes resulting from the publishing process, such as peer review, editing, corrections, structural formatting, and other quality control mechanisms may not be reflected in this document. Changes may have been made to this work since it was submitted for publication. A definitive version was subsequently published in Journal of Memory and Language, [61, 2, (August 2009)] doi:10.1016/j.jml.2009.04.004.
Abstract

Retrieval practice can enhance long-term retention of the tested material (the testing effect), but it can also impair later recall of the nontested material – a phenomenon known as retrieval-induced forgetting (Anderson, Bjork, & Bjork, 1994). Recent research, however, has shown that retrieval practice can sometimes improve later recall of the nontested material – a phenomenon termed retrieval-induced facilitation (Chan, McDermott, & Roediger, 2006). What drives these different effects? Two experiments were designed to examine the conditions under which retrieval induces forgetting and facilitation. Two variables, the level of integration invoked during encoding and the length of delay between retrieval practice and final test, were revealed as critical factors in determining whether testing facilitated or hindered later retrieval of the nontested information. A text processing framework is advanced to account for these findings.
When does Retrieval Induce Forgetting and When does it Induce Facilitation?

Although the idea of “frequent testing” may elicit negative reactions from students and educators alike, psychologists have long suspected that testing may have a positive influence on learning and its implications for education (Abbott, 1909; Ballard, 1913; Bjork, 1975; Brown, 1923; Gates, 1917; Naveh-Benjamin, 1990; Spitzer, 1939). In current scientific terms, this beneficial effect of retrieval is known as the **testing effect**. That is, taking an intervening test between learning and a delayed test boosts recall performance on that delayed test relative to a condition in which no initial test is taken (for reviews, see Crooks, 1988; Roediger & Karpicke, 2006).

During the earlier years of research on the testing effect, theorists suggested that the memorial benefits of testing are confined to materials that have been directly tested on the initial test (Duchastel, 1981; LaPorte & Voss, 1975; Nungester & Duchastel, 1982; Runquist, 1983, 1986). Recent studies, however, have revealed that testing, or retrieval practice, also influences later memory of the nontested materials. These findings are important from an educational perspective because critics of the testing effect have argued that test-enhanced learning has limited pedagogical generality because rigid learning of discrete, factual knowledge differs significantly from learning in the real world (Daniel & Poole, 2009). However, such a criticism misses an important aspect of retrieval. That is, retrieval serves more than to simply reinforce memory of a tested fact. For example, the effectiveness of testing as a flexible learning tool has been demonstrated in multiple approaches, including the beneficial effects of initial testing on subsequent new learning (Chan, Thomas, & Bulevich, 2009; Izawa, 1970; Robbins & Irvin, 1976; Szpunar, McDermott, & Roediger, 2008; Tulving & Watkins, 1974), which is, in effect, a type of transfer in learning (Phye & Sanders, 1992). Indeed, the flexibility of retrieval is
particularly apparent when its effects are demonstrated on the nontested materials. Therefore, a deeper understanding of the effects of retrieval on later memory of the nonretrieved items has important implications for educational practice.

Two literatures have independently investigated the later effects of testing on the nontested materials. Interestingly, they have arrived at different conclusions. For example, the conclusion from the retrieval-induced forgetting literature is that retrieval practice can impair later recall of the nontested materials (for reviews, Anderson, 2003; Bjork, Bjork, & MacLeod, 2006). In contrast, research from the adjunct questions tradition produced the opposite conclusion. Adjunct questions are questions embedded in the body of the text that students study. They can appear before (prequestions) or after the text (postquestions). Overall, research in this literature has found that answering adjunct questions facilitates later recall of the materials that are related to the adjunct questions (for reviews, see Crooks, 1988; Hamaker, 1986). The purpose of the current study is to elucidate variables that modulate the likelihood of obtaining facilitative vs. inhibitory effects of testing on the nontested materials. In the following sections, I first review the relevant literatures on retrieval-induced forgetting and retrieval-induced facilitation; I then present the logic behind the current experiments.

Evidence for Retrieval-Induced Forgetting

In an influential paper, Anderson, Bjork, and Bjork (1994) investigated the following question: After studying a list of categorized words (e.g., Fruit: orange, apple, banana, etc.), will performing retrieval practice on a subset of the exemplar words (e.g., orange) affect later recall of the nontested exemplar words (e.g., banana)? Anderson and colleagues’ experiment included four phases: a study phase, a retrieval practice phase, a distractor phase, and a final test phase. During the study phase, subjects studied category-exemplar pairs (e.g., fruit – orange, fruit –
banana, drinks – scotch, drinks – rum, etc.). During the retrieval practice phase, subjects performed a cued recall test on half of the exemplars from half of the categories (e.g., they might be tested on fruit – or ____ but not fruit – ba____, and not any items in the drinks category).

After a distractor phase (which typically lasts 5-30 min), subjects’ memory of the studied items is assessed on a final test. The practiced items are denoted $Rp^+$, the nonpracticed items from the practiced category are denoted $Rp^-$, and the items from the nonpracticed category are denoted $Nrp$. The general finding from this literature is that recall probability of the $Rp^-$ items is lower than that the $Nrp$ items, which suggests that retrieval practice of the $Rp^+$ items has impaired subsequent recall of their related ($Rp^-$) items. Anderson and his colleagues termed this finding retrieval-induced forgetting.

Briefly, the theoretical framework for retrieval-induced forgetting states that during the retrieval practice phase, suppression of the $Rp^-$ items serves to enhance retrieval of the $Rp^+$ items. $Rp^-$ items are suppressed because they are retrieval competitors against the $Rp^+$ items. This suppression/inhibition is later manifested as a reduction in the recall probability of the $Rp^-$ items during the delayed, final test. To ensure that retrieval inhibition occurs on the item level (i.e., the representation of banana itself) rather than on the association level (i.e., the linkage between fruit and banana), Anderson and colleagues demonstrated that retrieval-induced forgetting occurred even when an extra-list (or independent) cue was used to probe the $Rp^-$ item (e.g., yellow – ba____). Notably, though, the magnitude of retrieval-induced forgetting is typically smaller (and sometimes absent) with independent probes than that with studied (or intra-list) cues (Camp, Pecher, & Schmidt, 2007; Camp, Pecher, Schmidt, & Zeelenberg, 2009; Perfect, et al., 2004; Williams & Zacks, 2001), which suggests that retrieval inhibition may occur on both the item and association level.
The retrieval-induced forgetting paradigm has generated a wealth of research. Indeed, retrieval-induced forgetting has been shown in a wide variety of tasks (for reviews, see Anderson, 2003; Anderson & Neely, 1996; Bjork, et al., 2006; and for a recent report on the neural correlates of retrieval-induced forgetting, see Wimber, Rutschmann, Greenlee, & Bauml, 2008). Although agreement on the theoretical underpinnings of retrieval-induced forgetting has yet to be reached (e.g., Anderson & Spellman, 1995; Dodd, Castel, & Roberts, 2006; Norman, Newman, & Detre, 2007; Racsmany & Conway, 2006; Spitzer & Bauml, 2009; Williams & Zacks, 2001), the empirical conclusion from this literature is clear: retrieval practice can impair subsequent recall of the nontested-related material.

**Evidence for Retrieval-Induced Facilitation**

Although the literature on retrieval-induced forgetting may lead one to caution the memorial benefits of testing, a few recent studies (Callender & McDaniel, 2007; Carpenter, Pashler, & Vul, 2007; Chan, et al., 2006), in addition to studies in the adjunct questions literature (Hamaker, 1986), have suggested that retrieval practice can sometimes enhance later recall of the nontested materials. For example, Carpenter, Pashler, and Vul (2007) found that, in a paired associates learning task, retrieval practice of the target words enhanced subsequent recall of the cue words. That is, after learning word pairs such as “angle – corner” and performing retrieval practice (with corrective feedback) on “angle - ?”, delayed (18 – 48 hours later) recall of the cue word (angle) was enhanced relative to restudying the entire pair. Since exposure to the cue word was equated between the retrieval practice and restudy conditions, the enhanced recall of the cue can only be attributed to retrieval practice of the target. These researchers thus concluded that the testing benefit “spilled over to facilitate recall of information that was present on the test but
was not retrieved.” (p. 826, see also Kahana, 2002; Sommer, Schoell, & Buchel, 2008, for recent reviews of the vast literature on associative symmetry.)

More pertinent to the current purpose are results reported by Chan, McDermott, and Roediger (2006). In one experiment, subjects studied an article about the toucan bird and then either performed a cued recall test on that article (testing condition) or were dismissed (control condition). After a 24 hr delay, subjects completed the final test, which included questions that appeared during retrieval practice (Rp+) and questions that were related to the ones that appeared during retrieval practice (Rp-). For example, if the question “Where do toucans sleep at night?” (Answer: Treeholes) was presented during retrieval practice, then its related item “What other species are related to toucans?” (Answer: Woodpeckers) also appeared during the final test. If recalling a portion of the study materials (e.g., toucans sleep in treeholes) facilitates later recall of its related material, then subjects who completed the retrieval practice phase should outperform control subjects on the Rp- question set (e.g., the woodpeckers question). This was exactly what happened; subjects who completed retrieval practice outperformed control subjects on the Rp- items. Importantly, this retrieval-induced facilitation was not seen when subjects simply restudied (instead of attempting to retrieve) the Rp+ items, which suggests that retrieval serves a critical function in producing this benefit.

**Reconciling the Apparent Contradiction between Retrieval-Induced Facilitation and Retrieval-Induced Forgetting**

At first glance, the findings of retrieval-induced forgetting (henceforth RIFO) and retrieval-induced facilitation (henceforth RIFA) appear incompatible. How can such discrepant results occur from the same manipulation (i.e., retrieval practice of a subset of the learned materials)? To begin to answer this question, rather than declaring one pattern is real and the
other is false, a more productive endeavor would be to ask the following question: why would retrieval practice sometimes lead to facilitation and other times forgetting? Indeed, a better understanding of the conditions under which testing can help, or harm, later retrieval of initially nontested materials can serve two important purposes. First and from a theoretical perspective, it can extend the current understanding of the processes that are involved in retrieval. Second and from an applied perspective, it can be instrumental in helping psychologists to provide clear and effective advice for educators to improve student learning. A careful examination of the characteristics of the retrieval practice experiments suggests that two factors might be potentially critical in determining whether RIFO or RIFA is observed: First, the level of integration invoked at encoding; second, the length of delay introduced between retrieval practice and the final test. In the following sections, I briefly review the relevance of integration and delay on RIFO; I then illustrate their potential importance in obtaining RIFA.

Since Anderson et al.’s (1994) seminal paper, dozens of studies have confirmed the general finding of RIFO (for a review, see Anderson, 2003). Indeed, RIFO can be quite robust under many conditions. However, like most mental phenomena, boundary conditions exist for RIFO (Butler, Williams, Zacks, & Maki, 2001; Camp, Pecher, & Schmidt, 2005; Dodd, et al., 2006; Macrae & Roseveare, 2002; Perfect, Moulin, Conway, & Perry, 2002; Racsmany & Conway, 2006). Of particular interest here is that integrative encoding (Anderson & Bell, 2001; Anderson, Green, & McCulloch, 2000; Anderson & McCulloch, 1999; Bauml & Hartinger, 2002; Smith & Hunt, 2000) and delay (Bjork, et al., 2006; MacLeod & Macrae, 2001) are two such boundaries. For example, studies have shown that RIFO is eliminated when integration of the study materials is induced by a set of encoding instructions (e.g., Anderson & McCulloch, 1999) or by a high level of semantic similarity between the Rp+ and Rp- items (e.g., Bauml &
Hartinger, 2002); whereas other studies have shown that RIFO is eliminated when a long delay (24 hr) separates retrieval practice and the final test (e.g., MacLeod & Macrae, 2001; but see Garcia-Bajos, Migueles, & Anderson, 2009, for a recent demonstration of long-term RIFO).

Integration has been proposed to eliminate RIFO by reducing retrieval competition among response candidates and by mediated retrieval between the target and its associates (Anderson & McCulloch, 1999; Bjork, et al., 2006). The idea that integration reduces retrieval competition is not new. For example, integration has long been suggested to reduce response competition among retrieval candidates in the fan effect literature (Myers, O'Brien, Balota, & Toyofuku, 1984; Radvansky & Zacks, 1991; Smith, Adams, & Schorr, 1978) and in the classical verbal learning literature (Anderson & Neely, 1996; Osgood, 1946; Postman, 1971). According to Anderson and McCulloch (1999), retrieval competition between candidate items is the triggering mechanism for suppression. Therefore, if retrieval competition can be minimized via integration, then suppression of the target-related items becomes unnecessary and RIFO should not occur.

Mediated retrieval refers to the idea that subjects can use the target item as a retrieval cue to recall its related item (Postman, 1971). Based on this logic, integrative encoding may eliminate RIFO by strengthening the association between the target and its retrieval competitors. When participants attempt to recall an Rp- item during the final test, they can access this item through its association with the tested (Rp+) item. To illustrate, if recalling Orange during the initial test suppresses the representation of Banana, one can still recall Banana (when cued by Fruit) by first recalling Orange. One can then use the integrated association between Orange and Banana to recall Banana. The idea is termed mediated retrieval because when subjects encounter
Fruit – B____ during the final test, they recall Banana not based on the Fruit – Banana association, but based on the mediated association between Fruit – Orange – Banana.

Delay has been proposed to eliminate RIFO because retrieval inhibition is, theoretically, a transient, flexible, and adaptive mechanism (Bjork, et al., 2006). Inhibition of response competitors is supposed to make the target temporarily more retrievable. Parallels might be drawn between the gradual dissipation of retrieval inhibition over time with the phenomenon of spontaneous recovery from retroactive interference (Brown, 1991; Drosopoulos, Schulze, Fischer, & Born, 2007; Postman, Stark, & Fraser, 1968; Underwood, 1948; Wheeler, 1995), although the exact mechanism that underlies the effects of delay on retrieval inhibition is not well understood. It is important to note that studies have reported conflicting findings on whether retrieval inhibition is long lasting. Therefore, a more detailed review of the extant literature on the relation between delay and retrieval inhibition is presented in the General Discussion.

As the previous review shows, integration and delay are critical to the elimination of RIFO. Interestingly, these two variables might have also played a role in the occurrence of RIFA (Chan, et al., 2006). In the Chan et al. experiments, prose was used as the study material (which is more naturally integrative than word pairs – materials frequently employed in RIFO experiments) and a long (24 hr) retention interval separated the retrieval practice phase and the final test phase. If integration and delay can independently minimize retrieval inhibition, is it possible that they will produce facilitation when combined? This is an important question on multiple levels. First, from an applied perspective, understanding when retrieval practice can lead to enhancement and impairment of initially nontested materials would allow psychologists to make better recommendations to educators regarding pedagogical practices. Second, on the
empirical and theoretical levels, it is important to clearly delineate and explain conflicting findings in the literature. Although many studies have investigated the boundary conditions for RIFO, rarely are they designed to examine when retrieval inhibition would be reversed, because they tend to involve the variation of a single variable in isolation (e.g., integration was manipulated on its own, different memory tests were employed to investigate the generality of RIFO, etc.). Moreover, no study has yet to investigate the potential combined effects of multiple boundary variables on retrieval inhibition; therefore, it is still unclear whether these variables can have additive effects when combined. In sum, the current experiments were designed to address an important puzzle in the retrieval-induced forgetting and testing effect literatures.

In two experiments, I tested the roles that integration and delay play in the likelihood of RIFO and RIFA in a factorial design. Experiment 1 used the same prose materials as in Chan et al. (2006). Experiment 2 used simple propositional sentences that are similar to materials frequently found in the RIFO literature.

**Experiment 1**

**Design**

The experiment has a 2 (integration) X 2 (delay) X 3 (question type) design. Integration and delay were manipulated between subjects whereas question type was manipulated within subjects. Participants studied two articles (one on the Shaolin Temple and one on the Big Bang theory) and performed retrieval practice on one. After a period of distracting activities (i.e., delay), subject performed the final test on both articles. In keeping with the terms used in the RIFO literature, questions that appeared during the retrieval practice phase are termed Rp+, nonpresented questions from the retrieval practiced article are termed Rp-, and questions from the article that did not receive retrieval practice are termed Nrp.
During the encoding phase, integration was manipulated by presenting sentences in the passages in either a random or a coherent order. In the high-integration condition, sentences in the articles were presented individually in their natural, coherent order. In the low-integration condition, sentences were presented in a random order, which is expected to disrupt integration (cf. Balser, 1972; Carroll, Campbell-Ratcliffe, & Murnane, 2007; Engelkamp & Zimmer, 2002; Marschark, 1985; McDermott, 1996; Toglia, Neuschatz, & Goodwin, 1999). Note that sentences within the same paragraph were presented in a random order, but the paragraphs themselves were presented in their natural order. For example, subjects first studied sentences in the first paragraph presented in a random order, followed by sentences in the second paragraph presented in a random order, and so on. This procedure was chosen over a full randomization procedure to enhance the comprehensibility of the random presentation article. Specifically, the purpose here was to ensure that the manipulation of integration would lead to minimal difference in baseline (i.e., Nrp) performance. Indeed, any difference in Nrp performance between the high and low-integration groups can jeopardize the interpretability of the data. To preview, no such differences were found in the current data set.

To ensure that participants in the high-integration condition were indeed more likely to integrate the bits of information presented in the article, they were given explicit instructions to do so (although the integration instructions were believed to be redundant due to the coherent nature of the prose materials, see Einstein, McDaniel, Owen, & Cote, 1990). In addition to integration, retention interval was also manipulated in the current experiment. In the short delay condition, subjects took the final test 20 min after retrieval practice (a delay similar to the ones used in the RIFO paradigm). In the long delay condition, subjects took the final test 24 hr after retrieval practice (the same delay used in Chan et al., 2006).
Method

Participants

Ninety-six undergraduate students (24 in each condition) at Washington University in St. Louis received either course research credits or $20 for their participation.

Materials

The materials used in this experiment were similar to those in Experiments 2 and 3 of Chan et al. (2006). Specifically, subjects studied an article about the history of the Shaolin Buddhist temple and an article about the Big Bang theory. Both articles contained 13 paragraphs and were approximately 1900 words long. The length of the paragraphs varied between 4 and 10 sentences for the Shaolin article and between 3 and 12 sentences for the Big Bang article. Because the randomization procedure necessitated that sentences be presented one-at-a-time, minor changes were made to the original material to ensure that all sentences were sensible when presented individually. Most of these changes involved converting anaphoric inferences to names (e.g., “He” was replaced by “Einstein;” “it” was replaced by “the universe”).

Procedure

Subjects were tested individually or in groups of up to five people. They were tested either in different rooms or in the same room with dividers separating the computer terminals. Prior to the encoding phase, subjects were informed that they would learn something about the Big Bang theory and the Shaolin temple. All subjects were given intentional learning instructions and were told that sentences for each topic would be presented individually on the computer screen. They were further told that they might take an immediate test for one or both of the articles. Reading speed was self-paced, such that subjects could advance to the next sentence by pressing the Enter key. However, regardless of reading speed, all subjects were
given a total of 16 min to study each article. Therefore, a faster reader might have read all the sentences twice whereas a slower reader might have read most sentences only once. The total encoding time, however, was held constant across all subjects.

Subjects in the high-integration condition were told that they would read two articles and that they should try to integrate the materials within each article during encoding. In contrast, subjects in the low-integration condition were told that they would read facts, instead of an article, about each topic and that the facts would be presented in no particular order. Appendix A presents the full set of instructions for the high-integration condition and Appendix B presents the full set of instructions for the low-integration condition.

The retrieval practice phase occurred immediately following the encoding of one of the articles. It included two successive, identical tests for the same article. Multiple retrieval practice attempts are frequently employed in the design of RIFO experiments, although there is evidence that increasing the number of retrieval practice trials does not necessarily translate to a larger RIFO effect (Macrae & MacLeod, 1999; Shivde & Anderson, 2001). During each retrieval practice phase, subjects answered 12 questions, one-at-a-time, for 25 s each. Counterbalancing ensured that the two articles served in the retrieval-practice and control conditions equally often. After subjects had completed the study phase and the retrieval practice phase, those in the 24 hr condition were dismissed, whereas those in the 20 min condition completed the computerized Operation Span task (OSPAN, Conway, et al., 2005) and mental arithmetic problems for a total of 20 min. Subjects in the 24 hr condition completed the OSPAN task and arithmetic problems (also for a total of 20 min) immediately before the final test. As a result, the only difference between subjects in the two delay conditions was the length of the retention interval. During the final test, participants had 30 s to answer each question.
Results and Discussion

Statistical outcomes were reported with an alpha level of .05 unless otherwise noted. Partial eta squared ($\eta^2_p$) indicates effect size for analysis of variance (ANOVA) and Cohen’s $d$ indicates effect size for t-tests. Results from the retrieval practice phase are considered briefly for the sake of completeness, but the primary focus is on the results of the final test (and especially on the amount of RIFO and RIFA observed). The data are presented in the following order: (1) the retrieval practice results, (2) the final test results bearing on RIFO/RIFA, (3) the final test results concerning the testing effect. In addition, Appendix C displays all the questions used in this experiment, along with each question’s recall probability when the question was in the Rp+, Rp-, and Nrp condition.

Results from the Retrieval Practice Phase.

Subjects performed better on the second initial test ($M = .70$) than on the first initial test ($M = .66$), $F(1, 94) = 44.39$, $\eta^2_p = .32$ – an effect known as hypermnesia (Erdelyi & Becker, 1974; Payne, 1987). They also performed marginally better in the high-integration condition ($M = .71$) than in the low-integration condition ($M = .65$), $F(1, 94) = 3.13$, $\eta^2_p = .03$, $p = .08$.

However, the interaction between test number and integration condition was not significant, $F(1, 94) = 2.43$, $p > .10$.

Results from the Final Test.

Retrieval-Induced Facilitation and Retrieval-Induced Forgetting. Figure 1 displays the most important data of this experiment, which bear on the question of when retrieval practice leads to RIFO and RIFA. A comparison between the white bars and their adjacent gray bars in Figure 1 reveals that integration and delay contributed additively to RIFA. Specifically, when holding the other variable constant, integration and delay both increased the likelihood of RIFA as opposed
to RIFO. A 2 (question type: Nrp, Rp-) X 2 (integration) X 2 (delay) mixed ANOVA showed that delay had a main effect on recall performance, $F(1, 92) = 14.09, pes = .13$, such that recall probability declined over time. Not surprisingly, subjects performed more poorly at the 24 hr retention interval ($M = .53$) than at the 20 min retention interval ($M = .64$). Neither question type nor integration showed a main effect, both $F$s < 1.01. Integration and question type, however, produced a significant interaction, $F(1, 92) = 4.35, pes = .05$, so did delay and question type, $F(1, 92) = 5.01, pes = .05$. No three-way interaction was found, $F < 1$.

An examination of the interaction between integration and question type and the interaction between delay and question type revealed that integration and delay had similar and independent effects on RIFA. For example, the top panel in Figure 2 reveals that integration made RIFA more likely by enhancing the recall probability of the Rp- items but not the Nrp items. The bottom panel of Figure 2 shows that delay also enhanced the likelihood of RIFA, though in a different manner. Specifically, the Rp- items showed less forgetting (7% forgetting) than the Nrp items (16% forgetting) over the same retention interval. This is the first time retrieval practice has been shown to attenuate the forgetting function of Rp- materials.

Planned comparisons were conducted to establish that significant RIFA and RIFO were acquired in this data set. Specifically, subjects in the high-integration, long delay condition displayed significant RIFA, $t(23) = 2.15, d = .53$ ($M = .49$ for Nrp items and $M = .58$ for Rp-items, see the fourth pair of bars in Figure 1). In contrast, subjects in the low-integration, short delay condition demonstrated significant RIFO, $t(23) = 2.22, d = .57$ ($M = .66$ for Nrp items and $M = .57$ for Rp-items, see the first pair of bars in Figure 1). Remarkably, the magnitudes of RIFO and RIFA were the same in the present experiment (i.e., a 9% effect for both), and it is the first time RIFO has been found with these prose materials (cf. Chan, et al., 2006). Moreover,
neither RIFA nor RIFO was found in the low-integration, long delay condition (see the second pair of bars in Figure 1) or in the high-integration, short delay condition (see the third pair of bars in Figure 1), both ts < 1. Therefore, it appears that integration and delay are both important factors in creating RIFA. That is, the absence of both integration and delay led to RIFO, having one but not the other eliminated RIFO, and the presence of both produced RIFA.

**Testing Effect.** The effects of delay and integration on the testing effect can be seen in Figure 3 (the difference between the gray bar and its adjacent white bar indicates the size of the testing effect). Specifically, the testing effect was enhanced by both integration and delay. A 2 (within subjects, question type: Rp+, Nrp) X 2 (delay) X 2 (integration) ANOVA revealed a main effect of question type, $F(1, 92) = 37.73$, $pes = .29$, such that, overall, a testing effect was observed ($M = .70$ for Rp+ and $M = .58$ for Nrp; compare the white bars and their adjacent gray bars in Figure 3). Delay also produced a main effect, $F(1, 92) = 8.89$, $pes = .09$, such that recall probability dropped over time ($M = .68$ for 20 min and $M = .60$ for 24 hr). The main effect of integration, however, was not significant, $F = 1.00$.

More important than the main effects is the interaction between delay and question type, $F(1, 92) = 13.96$, $pes = .13$, such that the size of the testing effect increased with delay (.05 at 20 min and .20 at 24 hr). An examination of the bottom panel in Figure 4 shows that delay increased the testing effect by impairing the recall probability of the Nrp items, but it had very little impact on the recall probability of the Rp+ items. Indeed, the Rp+ items showed virtually no forgetting over the 24 hr delay, once again demonstrating the powerful influence of retrieval on subsequent forgetting (Carpenter, Pashler, Wixted, & Vul, 2008; Karpicke & Roediger, 2008). Similar to the results for the Rp- items, here integration also interacted (marginally) with question type, $F(1, 92) = 3.73$, $pes = .04$, $p = .06$, such that the testing effect was particularly
powerful in the high-integration condition (.08 for low-integration and .17 for high-integration). The top panel of Figure 4 shows that integration increased the testing effect by enhancing recall of the Rp+ items but not recall of the Nrp items.

Overall, several key findings emerged from Experiment 1. These results helped define the effects of integration and delay on the likelihood of obtaining RIFA and RIFO. First, RIFO occurred when integration was disrupted and when a short delay separated retrieval practice from the final test. Second, RIFA was found with integrative encoding and a long delay. Third and consistent with previous research (e.g., Anderson, et al., 2000; Anderson & McCulloch, 1999; Bauml & Hartinger, 2002; Carroll, et al., 2007; MacLeod & Macrae, 2001; Smith & Hunt, 2000), neither RIFO nor RIFA occurred when only one of the conditions was present. Fourth, retrieval practice slowed forgetting of the Rp- material, relative to the Nrp material (see bottom half of Figure 2). Fifth, integration augmented the magnitude of the later testing effect (see top half of Figure 4). In Experiment 2, I further examined the effects of integration and delay on RIFA and RIFO with more traditional RIFO-type materials.

**Experiment 2**

Experiment 1 provided evidence in favor of the hypothesis that integration and delay are important factors for the occurrence of RIFA. Nonetheless, the materials and procedures employed in Experiment 1 departed significantly from those used in a typical RIFO experiment. Therefore, in Experiment 2, I sought to extend Experiment 1’s findings with a set of simple propositional materials that are similar to the category-exemplar word pairs used in many RIFO studies. These propositional materials always involved an object and a location. Specifically, subjects studied sentences such as “The fork is in the nursery,” “the painting is in the nursery,” “the basket is in the attic,” etc. Retrieval always involved the recall of a target object given its
location. Paralleling Experiment 1, integration was manipulated by presenting sentences in a serial or a random order. For example, in the high-integration condition, participants would study “the fork is in the nursery,” “the painting is in the nursery,” “the mirror is in the nursery,” and “the radio is in the nursery” in a consecutive order, one-at-a-time. However, in the low-integration condition, all sentences were presented in a random order, so that it would be very unlikely for sentences that shared the same location to appear on back-to-back encoding trials. This presentation order, like that in Experiment 1, was expected to reduce the likelihood of integration during the encoding phase. In addition to integration, delay was also manipulated independently. Similar to Experiment 1, delay was varied by using a 20 min and a 24 hr retention interval.

Method

Participants

Ninety-six undergraduate students (24 in each condition) at Iowa State University participated for course research credits.

Materials

24 object and location names were chosen based on the following criteria: First, it must be reasonable for any of the objects to appear in any of the locations. Second, object names that shared the first two letters must not appear in the same location (because the first two letters of the object name were used during the recall tests). Third, all location and object names must be clearly distinguishable by meaning (e.g., because “university” was chosen as a location name, “college” was not). Participants studied 24 objects appearing in six unique locations. The 24 location names were rotated across participants. Counterbalancing ensured that each object
appear in each location equally often. Appendix D displays all the object and location names used in this experiment.

**Procedure**

Experiment 2 contained four phases. In Phase 1, participants studied the sentences one-at-a-time. Similar to Experiment 1, participants in the high-integration condition were told explicitly to integrate the objects presented in the same location, whereas participants in the low-integration condition were not. Appendix E displays the encoding instructions used in Experiment 2. A total of 24 object-location sentences were presented during the encoding phase. Presentation order was random; however, in the high-integration condition, sentences with the same location always appeared in consecutive trials to facilitate integration. Each sentence was presented for 8 s with a 1 s inter-stimulus interval.

In Phase 2 of the experiment, participants completed retrieval practice for half of the objects (i.e., two objects) in four of the locations (out of a total of six studied locations). This procedure thus created three item types. Specifically, eight objects were Rp+ items, eight objects were Rp- items (two nonpracticed objects from the four practiced locations), and eight objects were Nrp items (four objects from the two nonpracticed locations). During each retrieval practice trial, the computer presented a sentence with the object name cued by its first two letters (e.g., The fo______ was in the nursery). Participants had 10 s to answer each question and typed in their answer. Similar to Experiment 1 and many other RIFO experiments, the retrieval practice phase was repeated twice, each with a different randomization order.

In Phase 3 of the experiment, participants performed a distractor task by playing the video game “Tetris” – a popular falling-rock puzzle game. All participants played Tetris for 20 min. Afterwards, participants in the 20 min condition proceeded to Phase 4, the final test phase,
whereas participants in the 24 hr condition were dismissed and completed the final test upon their return on the next day.

In Phase 4 of the experiment, all participants completed the final test, which used the same presentation protocol as the retrieval practice phase with the following exceptions. First, the final test consisted of 24 (instead of 8) questions and second, participants had 15 s (instead of 10 s) to answer each question. Like Experiment 1, the final test questions appeared only once.

**Results and Discussion**

All analyses were performed with an alpha level of .05. Again, I first present results from the retrieval practice phase, follow by results concerning RIFA/RIFO, and then results concerning the testing effect.

**Results from the Retrieval Practice Phase.**

The results from the retrieval practice phase in Experiment 2 paralleled those in Experiment 1. Specifically, hypermnesia was found, such that recall probability increased across test trials (from .71 to .75), $F(1, 94) = 9.75, pes = .09$. In addition, integration led to better initial recall (.77 for high-integration and .69 for low-integration), $F(1, 94) = 3.89, pes = .04$. However, integration and test trial did not interact, $F < 1$.

**Results from the Final Test.**

**Retrieval-Induced Facilitation and Retrieval-Induced Forgetting.** Figure 5 displays the data of most interest to this experiment. A 2 (item type: Rp- vs. Nrp) X 2 (high- vs. low-integration) X 2 (20 min vs. 24 hr) ANOVA revealed an interaction between item type and integration, $F(1, 92) = 6.77, pes = 0.07$, and a main effect of delay, $F(1, 92) = 18.77, pes = .17$. No other interactions or main effects were significant. Most important for the current purpose was whether we replicated the major results in Experiment 1 – we did. Specifically, a 7% RIFO effect was found in the
low-integration, 20 min delay condition, \( t(23) = 2.18, d = .29 \). However, a 11% RIFA effect was found in the high-integration, long delay condition, \( t(23) = 2.83, d = .52 \). Moreover, similar to Experiment 1, integration and delay, when implemented separately, eliminated RIFO but did not produce RIFA, both \( ts < 1 \).

The interaction between integration (High, Low) and item type (Rp-, Nrp) is illustrated in the top half of Figure 6. As can be seen, this pattern closely mimics the one presented in the top half of Figure 2, such that integration enhanced subsequent recall of the Rp- items, but it had no impact on subsequent recall of the Nrp items, \( F(1, 94) = 6.76, pes = .07 \). However, the interaction between delay (20 min, 24 hr) and item type (Rp-, Nrp) was not significant (see bottom half of Figure 6), \( F(1, 94) = 2.00, p = .16, observed power = .29 \). Notably, although this interaction failed to reach conventional significance level, the Rp- items exhibited numerically less forgetting (15%) than did the Nrp items (21% forgetting) over the same time period (24 hr), which is consistent with the pattern observed in Experiment 1.

**Testing Effect.** Figure 7 shows the effects of integration and delay on subsequent recall of the Rp+ items and Nrp items. Importantly, though perhaps unsurprisingly, the main effect of item type was significant, such that there was a robust (22%) testing effect overall, \( F(1, 92) = 102.14, pes = .53 \). The main effect of delay was also significant, which indicates that reliable forgetting occurred over the 24 hr retention interval, \( F(1, 92) = 25.08, pes = .21 \). No main effect for integration was observed, \( F(1, 92) = 2.46, p = .12 \). Moreover, replicating Experiment 1, there was a significant interaction between integration and item type (see the top panel of Figure 8), \( F(1, 92) = 9.93, pes = .02 \), such that integrative encoding led to a greater testing effect. No other interactions were significant, all \( Fs < 1.62 \). It is unclear why the testing effect did not increase significantly with delay in this experiment, although the difference was in the expected direction.
(such that the testing effect was numerically larger 24 hr after retrieval practice \([M = .24]\) than 20 min after retrieval practice \([M = .21]\)). Overall, the results of Experiment 2 largely replicated and extended those from Experiment 1.

**General Discussion**

In two experiments, I have demonstrated the importance of integration and delay on the effects of retrieval practice on later memory. Experiment 1 showed that when integrative encoding was combined with a 24 hr delay, initial testing enhanced later recall of the nontested material. However, when integration was made difficult and a short delay separated the initial test and final test, retrieval practice hampered later recall of the nontested material. These findings are extended to the learning of spatial relations between an object and a location. In this General Discussion, I first review the relevant literature on the effects of integration on RIFO/RIFA. I then propose a theoretical explanation for how integration and delay can protect against RIFO and can induce RIFA in prose. This account incorporates ideas from the literatures on retrieval inhibition, spreading activation, and text processing, and it provides a useful framework for thinking about the processes underlying RIFA and RIFO in the context of prose.

**The Importance of Integration in Producing RIFA or RIFO**

Several methods have been devised to induce integration in the retrieval practice paradigm. These methods include giving subjects a set of integration instructions (Anderson & McCulloch, 1999; Smith & Hunt, 2000), using materials that are naturally integrative (Bauml & Hartinger, 2002; Carroll, et al., 2007; Conroy & Salmon, 2005; Garcia-Bajos, et al., 2009; Little, Bjork, & Bjork, 2007; Migueles & Garcia-Bajos, 2007), and asking subjects to make similarity judgments on the study materials (Anderson, et al., 2000). Although only a small subset of these studies reveal a RIFA effect (Anderson, et al., 2000; Garcia-Bajos, et al., 2009), the others have
all shown that integrative encoding can reduce or eliminate RIFO. Hence, integration has been regarded as a general boundary condition for retrieval-induced forgetting (Anderson, 2003).

Anderson and colleagues (Anderson, et al., 2000; Anderson & McCulloch, 1999; Bjork, et al., 2006) provided what is perhaps the most widely accepted explanation for the effects of integration on RIFO. According to this explanation, integration cancels the debilitating effects of retrieval inhibition by reducing retrieval competition among study items.¹ This idea is more clearly delineated under Anderson and Spellman’s (1995) pattern suppression model (see also Anderson, 2003), which specifies that integration reduces the likelihood of retrieval inhibition by enhancing the encoding of the shared features among retrieval competitors. For example, if Orange and Lemon are retrieval competitors under the Fruit category, subjects might integrate these items by thinking about their shared features such as their citrus quality and similar texture, in addition to the fact that they are both fruits (the retrieval cue). When subjects recall Orange during the retrieval practice phase, the shared features between Lemon and Orange are facilitated, thus allowing Lemon (the Rp-item) to be retrieved at levels comparable to control (Nrp) items during the final test (see Figure 9 for a graphical depiction of this logic).

The pattern suppression model can be easily extended to explain the effects of integration on RIFA. Specifically, RIFA is more likely to occur at longer retention intervals because the facilitated features of the Rp-material might no longer need to counteract the previously inhibited features (due to the purportedly fleeting nature of retrieval inhibition). Despite the intuitive appeal of the pattern suppression framework, a problem may limit its application to the current results. In particular, although it is relatively simple to explain integration from a shared features perspective when one is dealing with objects like Orange and Lemon, it is more difficult to imagine what constitutes shared features when the materials are complex prose. For example,
what are the common features between the concepts “Shaolin means young forest,” and “the Shaolin temple was built on a piece of land that had been ravaged by fire, because the builders planted many new trees”? Clearly, the connection between these facts is based more on a conceptual level than on a feature level. Therefore, an alternative, or modified, version of the pattern suppression model is required to adequately address the current findings.

A Text Processing Account

When dealing with prose materials (rather than word lists), instead of thinking about related ideas in prose as features within objects, it might be more appropriate to think of them as idea units within mental models. According to the discourse processing literature, three levels of representation are created when people comprehend prose (Johnson-Laird, 1983; van Dijk & Kintsch, 1983). These representations differ on how closely they adhere to the original text. The lowest level of representation – the surface level – contains the exact wording of the sentences and is highly transient. The middle level of representation is known as the text base, which contains a paraphrased version of the text while preserving its original meaning. The highest level of representation is the situation model, which represents ideas of the text in an abstract format (e.g., the people, the actions, the setting, the events, and the inferences). Information in the situation model interacts with pre-existing knowledge (i.e., episodic and semantic memory) of the reader, so that situation models can be updated online as one continues reading/listening to the text (Fletcher, 1994; Graesser, Millis, & Zwaan, 1997).

In relation to the current study, this idea suggests that when people encounter new information as they read sentences in an article, they may update an existing situation model or they may create a new one to accommodate the newly acquired information. One factor that determines whether the former or the latter occurs is the ease with which new information can be
consolidated into existing situation models (Gernsbacher, 1997). Therefore, when information is presented in a coherent, integrative manner, fewer situation models should be formed because it is easier to update existing situation models with the new information. However, when information is presented as separate, randomly-ordered facts, many more situation models might be required (Radvansky, 1999; Radvansky & Zacks, 1991). Importantly, recent evidence in the fan effect literature shows that multiple pieces of information stored in the same situation model do not interfere with each other (Radvansky & Copeland, 2006; Radvansky & Zacks, 1991; see also, Anderson & Bell, 2001).³ The fan effect refers to an increase in response time when subjects have learned multiple pieces of information associated with a given target (e.g., the fireman is in the park, the fireman is in the house), relative to having learned just one piece of information about a given target (e.g., the mug is in the office).

Related to the current study, Radvansky and Zacks (1991) had participants study object-location sentences similar to the ones in Experiment 2. Three object-location learning conditions were employed. In Condition 1, one object appeared in one location; in Condition 2, *multiple objects appeared in one location*; in Condition 3, *one object appeared in multiple locations*. Interestingly, the fan effect occurred in Condition 3 (one object – multiple locations) but not in Condition 2 (multiple objects – one location). Radvansky and Zacks argued that no interference was observed in Condition 2 because participants were able to integrate multiple objects into a single situation model when these objects appear in the same location. However, it was difficult (and unnatural) for subjects to form a single situation model that encompasses the same object appearing in different locations; as a result, different mental models must be formed for each object-location pairing, leading to the fan interference effect.
Applying this logic to the present findings suggests that integration reduced the likelihood of RIFO because it led subjects in the high-integration condition to incorporate more related ideas into the same situations models, thus reducing the overall level of interference across situation models. In contrast, subjects in the low-integration condition may have to repeatedly create new situation models to incorporate new facts. Therefore, when subjects in the low-integration condition retrieved a concept during retrieval practice, related information stored in the many other situation models might interfere and therefore need to be inhibited (Anderson & Bell, 2001; Radvansky, 1999). This inhibition was later manifested as RIFO. Importantly and similar to the pattern suppression model, this idea assumes that retrieval inhibition occurs in both the high- and low-integration conditions, but the level of inhibition varies between subjects in these conditions. That is, because subjects in the high-integration condition are expected to generate fewer situation models than their low-integration counterparts, they should exhibit less retrieval-induced forgetting.

To adequately account for the finding of RIFA, one assumption of the current explanation needs to be made explicit. Specifically, when multiple pieces of information are stored in the same mental model, not only are the nontested items not inhibited by retrieval practice of their associates, they are strengthen by such retrieval practice. This enhancement can occur based on automatic spreading activation or based on a strategic retrieval process (Burgess & Shallice, 1996; Chan, et al., 2006). When the practiced (Rp+) and nonpracticed (Rp-) items are integrated into the same mental model, retrieval of the Rp+ item can activate the Rp- item, and this activation is not counteracted by retrieval inhibition. However, retrieval practice of a given item can still trigger inhibition of items in other mental models (i.e., the Nrp items), but such inhibition should be much less robust than would be for directly-related items (like the Rp-
items), which are more likely to serve as retrieval competitors. This idea is consistent with the finding that the strength of competition modulates the level of retrieval inhibition (see Experiment 3 of Anderson, et al., 1994; Anderson & Spellman, 1995). As a result, this text processing account argues that RIFA and RIFO always occur together (see Racsmany & Conway, 2006, for a similar idea about a memory representation being activated and inhibited simultaneously), but the factor that determines whether RIFA or RIFO is observed is the ratio between activation and inhibition, if the number of activated Rp-traces outweighs those of inhibited traces, than RIFA would be observed (and vice versa).

One important implication of this logic is that RIFO and RIFA are not incompatible phenomena; rather, they represent different sides of a continuum based on the observable difference between the overall activation and inhibition. This idea is interesting because it generates some clearly testable hypotheses. For example, one may attempt to manipulate the activation/inhibition ratio by varying the number of situation models formed – with the amount of retrieval inhibition positively associated with the number of mental models formed. Alternatively, readers with different structure building ability (Callender & McDaniel, 2007; McDaniel, Hines, & Guynn, 2002) may vary on how well they can update already-formed situation models to incorporate new ideas, so it is possible that low structure builders are more likely to experience retrieval inhibition in complex prose than are high structure builders.

Although applying the idea of mental models to the pattern suppression model (Anderson & Spellman, 1995) presents a promising framework for thinking about RIFO/RIFA in prose, one important question remains. Specifically, how does delay affect the likelihood of retrieval-induced facilitation and inhibition? Based on the extant literature (MacLeod & Macrae, 2001; Saunders & MacLeod, 2002; but see Tandoh & Naka, 2007) and the current finding, it appears
that delay has a powerful influence on retrieval inhibition, such that retrieval inhibition weakens as the retention interval between retrieval practice and the final test increases. Several studies, however, have demonstrated that RIFO can have long-term (i.e. 24 hr or longer) consequences. However, many of these experiments shared one common methodological characteristic that differed from the typical retrieval practice paradigm: subjects in these experiments performed retrieval practice spaced over three (Conroy & Salmon, 2005, 2006) or more (Ford, Keating, & Patel, 2004) days. In a direct comparison between massed and spaced testing, no long-term RIFO was observed when massed retrieval practice was employed, which is the common procedure in RIFO experiments; however, when retrieval practice sessions were spaced over days, long-term RIFO was found (Conroy & Salmon, 2005). Two other studies also showed long-term RIFO effects (Migueles & Garcia-Bajos, 2007; Storm, Bjork, Bjork, & Nestojko, 2006). However, because delay was manipulated within subjects in these experiments, the RIFO effect was confounded with the testing effect. To date, the most promising evidence for long-term retrieval inhibition was reported by Garcia-Bajos et al. (2009), who showed that RIFO can be demonstrated over a 1-week retention interval.

Based on these conflicting findings, it is clear that it would be premature to conclude that delay is a general boundary condition for RIFO, though it is equally clear that evidence exist to support the idea that retrieval inhibition can subside with time. More generally, our current knowledge about the time course of retrieval inhibition is rather limited, and more empirical evidence needs to be accumulated to better understand the effects of delay on retrieval inhibition. More important for current purposes, though, is whether RIFA possesses a different time course than RIFO? The present findings suggest that the enhancement in recall produced by retrieval practice is perhaps longer lasting than the degradation in recall produced by retrieval inhibition.
According to Chan et al. (2006), RIFA occurs because subjects actively search for related (e.g., the Rp-) concepts when they attempt to answer a question (similar to the conscious search component proposed by Burgess and Shallice [1996] in autobiographical memory retrieval); therefore, the enhancement in recall for the Rp- concepts (i.e., RIFA) should behave in a similar fashion to the enhancement in recall for the Rp+ concepts (i.e., the testing effect). Critically, one of the best-known properties of the testing effect is that it is very long-lasting (Roediger & Karpicke, 2006); therefore, assuming that retrieval inhibition fades with time in the current study, RIFA and RIFO might, tentatively speaking, have different time courses. Again, more research needs to be conducted to further shed light on this issue.

**Concluding Remarks**

Traditionally, retrieval is treated as an evaluative tool that reveals what people remember and what they have forgotten. However, retrieval does not only show what people know, it also changes what people know. Retrieval practice is one of the most powerful ways to enhance long-term retention (Karpicke & Roediger, 2008), and the influence of retrieval practice is not confined to only the directly tested items. In the current study, I have shown that two factors – integration and delay – govern when retrieval induces forgetting and when it induces facilitation. Armed with this knowledge, one might advice students not to perform retrieval practice immediately before an exam to avoid the negative influence of retrieval inhibition, especially if the to-be-tested materials are similar to the category-exemplar materials used in the typical RIFO paradigm (e.g., one may want to avoid practicing retrieval of a subset of the items in the periodic table immediately before a chemistry exam). Clearly, a better understanding of when RIFO and RIFA occur would allow educators to better apply testing as a learning tool to education.
References


Author Notes

Part of the work reported here was supported by a grant from James S. McDonnell Foundation (220020041) to Kathleen McDermott and Roddy Roediger. Experiment 1 was conducted as part of the dissertation for Jason Chan at Washington University. Experiment 2 was conducted at Iowa State University and was not supported by the McDonnell grant. I thank Kathleen McDermott, Roddy Roediger, Mark McDaniel, Jan Duchek, Alan Lambert, Jim Wertsch, and Pascal Boyer for helpful comments on the dissertation version of this manuscript. The data from Experiment 1 was presented at the 2007 meeting of the Psychonomic Society at Long Beach, CA. I also thank Moses Langley, Stephen Anderson, Katie Paulsen, and Ben Davis for their help with participant testing in Experiment 2.
Footnotes

1 The idea that integration minimizes RIFO by reducing retrieval competition may explain why a recent study failed to observed RIFA (Camp, de Bruin, le Febre, & Matil, 2009). In this study, the authors were able to eliminate RIFO by presenting prose materials in an integrative, coherent manner and by using non-competitive questions during the retrieval practice phase; however, RIFA was not observed. It is possible that RIFA was not found in this study because their two manipulations were both designed to reduce retrieval competition, thus producing a sub-additive (or redundant) effect. In contrast, the two manipulations employed here reduced retrieval inhibition by different underlying (and presumably additive) mechanisms, thereby making RIFA more likely.

2 For current purposes, the terms mental model and situation model are used interchangeably.

3 For an alternative account, see J.R. Anderson and Reder (1999).

4 Two other recent studies have reported reliable long-term RIFO. However, details on the methodology of one of these studies are scarce because it was published in Japanese (Tandoh & Naka, 2007). The other study found that long-term (12 hr) retrieval inhibition was observed only when a period of sleep separated retrieval practice and the final test or when participants actively rehearsed the Rp+ items during the retention interval (Racsmany, Conway, & Demeter, 2009).
Figure Captions

Figure 1. Probability of correct recall on the final test as a function of encoding condition (low vs. high-integration) and delay (20 min vs. 24 hr) in Experiment 1. The white (left) bars represent performance on the Nrp (Control) questions and the gray (right) bars represent performance on the Rp- (Nontested-Related) questions. The comparison between the two leftmost bars shows retrieval-induced forgetting (RIFO), whereas the comparison between the two rightmost bars shows retrieval-induced facilitation (RIFA). Error bars are within subjects .95 CI.

Figure 2. Disparate effects of integration and delay on RIFA in Experiment 1. The top panel shows that integration augmented RIFA by enhancing recall of the Rp- items but not the Nrp items. The bottom panel shows that testing attenuated forgetting of the Rp- items relative to the Nrp items. Error bars are within subjects .95 CI.

Figure 3. Probability of correct recall on the final test as a function of encoding condition (low vs. high-integration) and delay (20 min vs. 24 hr) in Experiment 1. The white bars represent performance on the Nrp questions, and the dark bars represent performance on the Rp+ questions. A comparison of the white and dark bars shows the testing effect. Error bars are within subjects .95 CI.

Figure 4. Disparate effects of integration and delay on the testing effect in Experiment 1. The top panel shows that integration augmented the testing effect by enhancing recall of the Rp+ items but not the Nrp items. The bottom panel shows that delay augmented the testing effect by impairing recall of the Nrp items but not the Rp+ items. Error bars are within subjects .95 CI.

Figure 5. Probability of correct recall on the final test as a function of encoding condition (low vs. high-integration) and delay (20 min vs. 24 hr) in Experiment 2. The white bars represent
performance on the Nrp (Control) questions and the gray bars represent performance on the Rp- (Nontested-Related) questions. The comparison between the two leftmost bars shows retrieval-induced forgetting (RIFO), whereas the comparison between the two rightmost bars shows retrieval-induced facilitation (RIFA). Error bars are within subjects .95 CI.

Figure 6. Disparate effects of integration and delay on RIFA in Experiment 2. The top panel shows that integration augmented RIFA by enhancing recall of the Rp- items but not the Nrp items. The bottom panel shows that testing attenuated forgetting of the Rp- items relative to the Nrp items. Error bars are within subjects .95 CI.

Figure 7. Probability of correct recall on the final test as a function of encoding condition (low vs. high-integration) and delay (20 min vs. 24 hr) in Experiment 2. The white bars represent performance on the Nrp questions, and the dark bars represent performance on the Rp+ questions. A comparison of the white and dark bars shows the testing effect. Error bars are within subjects .95 CI.

Figure 8. Disparate effects of integration and delay on the testing effect in Experiment 2. The top panel shows that integration augmented the testing effect by enhancing recall of the Rp+ items but not the Nrp items. The bottom panel shows that, curiously and unlike Experiment 1, delay had similar effects on the Rp+ and Nrp items. Error bars are within subjects .95 CI.

Figure 9. The effects of retrieval practice on materials that are integrated (top panel) and materials that are not integrated (bottom panel). Black dots stand for features that are enhanced by retrieval practice, whereas the crossed out dots stand for features that are inhibited. When items have been integrated during encoding, retrieval practice enhances a larger portion of the features in the retrieval competitor than when the items have not been integrated, thus
counteracting the impairments of retrieval inhibition. The illustration is adopted from Anderson, Green, and McCulloch (2000).
Figure 1
Figure 2
Figure 3
Figure 4

Integration Probability of Correct Recall

Nrp Rp+

Probability of Correct Recall

Low High

20 min 24 hr Delay

Figure 4
Figure 5
Figure 6
Figure 7

The figure shows the probability of correct recall over time for different integration levels. The x-axis represents time points of 20 min and 24 hr, and the y-axis represents the probability of correct recall. Two conditions are compared: Low Integration and High Integration. The error bars indicate the variability in the data.
Figure 8
Integrative Encoding

Non-Integrative Encoding

Figure 9
Appendix A

Instructions for participants in the high-integration condition in Experiment 1.

You will now read an article about the (Shaolin temple/Big Bang theory). Please pay close attention to the article and try your best to memorize it because your memory for this article will be tested later in the experiment. The computer will present sentences in the article to you one-at-a-time. When you finish reading the sentence on screen, simply press the “Enter” key to see the next sentence. You can read this article at your own pace. Note that you will not be able to go back to a previously seen sentence. You will have up to 16 minutes to read this article.

However, since this article is not very long, I anticipate that you will be able to finish reading the entire article in less than 16 minutes. When this happens, the computer will present the article from the beginning again – this will be your only chance to re-read sentences that you have read before. Please use the entire 16 minutes and take full advantage of the re-presentation of the article.

Another important thing for you to do is to try your best to integrate information presented in the article. Because the sentences are presented individually, it is important that you keep information that you have just learned in mind while reading the new sentences. Try to integrate the information and to relate them to each other as well as you can. This should help you in memorizing the article.
Appendix B

Instructions for participants in the low-integration condition in Experiment 1.

You will now learn some facts about the Big Bang theory/Shaolin temple. Please pay close attention to the facts and try your best to memorize them because your memory for these facts will be tested later. The computer will present these facts one-at-a-time. When you finish reading the fact on screen, simply press the "Enter" key to see the next fact. You can read these facts at your own pace. Note that you will not be able to go back to a previously seen fact. You will have up to 16 minutes to read these facts. Note that you will not be able to go back to a previously seen fact. However, since there is a limited set of facts, I anticipate that you will be able to finish reading the whole set of facts in less than 16 minutes. When this happens, the computer will present the facts from the beginning again - this will be your only chance to re-read facts that you have read before. Please use the entire 16 minutes and take full advantage of the re-presentation of the facts. This should help you in memorizing the facts.
Appendix C

Questions (and their answers) used in Experiment 1. Proportion of accurate recall is displayed as a function of item type (i.e., Rp+, Rp-, Nrp) separated by integration (i.e., High and Low) and delay (20 min, 24 hr). Questions from the Big Bang article is presented first, followed by questions from the Shaolin temple article. Related questions appear in the following table in consecutive order (e.g., Q1 is related to Q2, Q3 is related to Q4, etc.).

<table>
<thead>
<tr>
<th>Question</th>
<th>Answer</th>
<th>Rp+</th>
<th>Rp-</th>
<th>Rp+</th>
<th>Rp-</th>
<th>Rp+</th>
<th>Rp-</th>
<th>Rp+</th>
<th>Rp-</th>
</tr>
</thead>
<tbody>
<tr>
<td>According to the flat and open models, the universe will</td>
<td>expand infinitely</td>
<td>0.83</td>
<td>0.83</td>
<td>1.00</td>
<td></td>
<td>0.58</td>
<td>0.60</td>
<td>1.00</td>
<td></td>
</tr>
<tr>
<td>According to the oscillating closed universe model, the universe will</td>
<td>big crunch</td>
<td>0.58</td>
<td>1.00</td>
<td>0.67</td>
<td>0.67</td>
<td>0.43</td>
<td>0.50</td>
<td>0.71</td>
<td>0.60</td>
</tr>
<tr>
<td>alternate between a big bang and a</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>In 1964, astronomers Arno Penzias and Robert Wilson inadvertently</td>
<td>extraterrestrial origin</td>
<td>0.83</td>
<td>0.83</td>
<td>0.83</td>
<td>0.67</td>
<td>0.57</td>
<td>0.33</td>
<td>0.75</td>
<td>0.60</td>
</tr>
<tr>
<td>discovered a noise that they thought originated from an</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Later on, it was obvious that what astronomers Arno Penzias and</td>
<td>cosmic background radiation</td>
<td>0.50</td>
<td>0.33</td>
<td>0.67</td>
<td>0.33</td>
<td>0.67</td>
<td>0.43</td>
<td>0.50</td>
<td>0.86</td>
</tr>
<tr>
<td>Robert Wilson heard was</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>After the Big Bang, gravity condensed clumps of matter together and</td>
<td>galaxies</td>
<td>0.92</td>
<td>1.00</td>
<td>1.00</td>
<td>0.75</td>
<td>0.83</td>
<td>0.57</td>
<td>0.58</td>
<td>0.71</td>
</tr>
<tr>
<td>these clumps eventually formed</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>What theory predicted the exaggerated outward expansion of the</td>
<td>inflation theory</td>
<td>0.42</td>
<td>0.17</td>
<td>0.00</td>
<td>0.17</td>
<td>0.29</td>
<td>0.00</td>
<td>0.33</td>
<td>0.60</td>
</tr>
<tr>
<td>universe (where particles were flying outward faster than the speed of</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>light)?</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Edwin Hubble discovered that a galaxy's velocity is proportional to its</td>
<td>distance from Earth</td>
<td>0.83</td>
<td>0.67</td>
<td>0.83</td>
<td>0.83</td>
<td>0.67</td>
<td>0.57</td>
<td>0.75</td>
<td>0.86</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>It took every galaxy</td>
<td>the same amount of time</td>
<td>0.50</td>
<td>0.67</td>
<td>0.00</td>
<td>0.42</td>
<td>0.43</td>
<td>0.17</td>
<td>0.67</td>
<td>0.40</td>
</tr>
<tr>
<td>to move from a common starting position to its current position.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
At the beginning of the universe, there was an unequal amount of two types of particles, when these two types of particles collide and destroy each other, ______________ is created.

<table>
<thead>
<tr>
<th>pure energy</th>
<th>1.00</th>
<th>1.00</th>
<th>0.83</th>
<th>0.58</th>
<th>0.57</th>
<th>0.83</th>
<th>0.58</th>
<th>0.60</th>
<th>0.57</th>
<th>0.25</th>
<th>0.20</th>
<th>0.67</th>
</tr>
</thead>
</table>

Immediately after the Big Bang, the universe was extremely hot as a result of two types of particles rushing apart in all directions, what are they?

<table>
<thead>
<tr>
<th>matter and antimatter</th>
<th>0.83</th>
<th>1.00</th>
<th>1.00</th>
<th>0.42</th>
<th>0.50</th>
<th>0.71</th>
<th>0.50</th>
<th>0.57</th>
<th>0.60</th>
<th>0.50</th>
<th>1.00</th>
<th>0.80</th>
</tr>
</thead>
</table>

Due to Doppler shifting, the wavelength emitted by something travelling away from us is shifted to a ______________ frequency (length-wise).

<table>
<thead>
<tr>
<th>lower or longer</th>
<th>1.00</th>
<th>0.83</th>
<th>0.67</th>
<th>0.83</th>
<th>0.50</th>
<th>0.29</th>
<th>0.83</th>
<th>1.00</th>
<th>0.80</th>
<th>0.77</th>
<th>1.00</th>
<th>0.80</th>
</tr>
</thead>
</table>

When visible wavelengths are emitted by objects moving away from us, their wavelengths are said to be ______________. (Name the phenomenon)

<table>
<thead>
<tr>
<th>redshifted</th>
<th>0.83</th>
<th>0.67</th>
<th>0.33</th>
<th>0.58</th>
<th>0.57</th>
<th>0.67</th>
<th>0.67</th>
<th>0.80</th>
<th>0.57</th>
<th>0.54</th>
<th>0.20</th>
<th>0.83</th>
</tr>
</thead>
</table>

Einstein resisted the idea of the beginning of the universe by introducing a constant called ______________ into his equations.

<table>
<thead>
<tr>
<th>fudge factor</th>
<th>0.92</th>
<th>0.83</th>
<th>0.67</th>
<th>0.67</th>
<th>0.86</th>
<th>0.67</th>
<th>0.75</th>
<th>1.00</th>
<th>0.86</th>
<th>0.92</th>
<th>0.80</th>
<th>1.00</th>
</tr>
</thead>
</table>

Einstein ultimately gave grudging acceptance to "the necessity for a beginning" and eventually to "the presence of a ______________".

<table>
<thead>
<tr>
<th>superior reasoning power</th>
<th>0.33</th>
<th>0.17</th>
<th>0.17</th>
<th>0.25</th>
<th>0.33</th>
<th>0.14</th>
<th>0.08</th>
<th>0.29</th>
<th>0.20</th>
<th>0.08</th>
<th>0.17</th>
<th>0.00</th>
</tr>
</thead>
</table>

Scientists believe that there was one helium nucleus for every ______________ protons within the first three minutes of the universe.

<table>
<thead>
<tr>
<th>ten</th>
<th>0.92</th>
<th>1.00</th>
<th>0.67</th>
<th>0.58</th>
<th>0.86</th>
<th>0.33</th>
<th>0.83</th>
<th>1.00</th>
<th>0.86</th>
<th>0.67</th>
<th>1.00</th>
<th>0.50</th>
</tr>
</thead>
</table>

When protons and neutrons react, they form heavy hydrogen. Another way to call these heavy hydrogen is ______________.

<table>
<thead>
<tr>
<th>deuterium</th>
<th>0.50</th>
<th>0.33</th>
<th>0.33</th>
<th>0.33</th>
<th>0.67</th>
<th>0.57</th>
<th>0.25</th>
<th>0.57</th>
<th>0.80</th>
<th>0.58</th>
<th>0.67</th>
<th>0.60</th>
</tr>
</thead>
</table>

NASA created a satellite that orbits around the Earth. What was one question that the creation of this satellite attempted to answer?

<table>
<thead>
<tr>
<th>COBE (Cosmic Background Explorer)</th>
<th>0.33</th>
<th>0.67</th>
<th>0.67</th>
<th>0.42</th>
<th>0.33</th>
<th>0.57</th>
<th>0.67</th>
<th>0.43</th>
<th>1.00</th>
<th>0.17</th>
<th>1.00</th>
<th>0.00</th>
</tr>
</thead>
</table>

NASA created the ______________ to detect radiation emanating from the universe. (You may use abbreviations)

<table>
<thead>
<tr>
<th>q</th>
<th>0.25</th>
<th>0.17</th>
<th>0.17</th>
<th>0.08</th>
<th>0.29</th>
<th>0.17</th>
<th>0.42</th>
<th>0.40</th>
<th>0.43</th>
<th>0.08</th>
<th>0.40</th>
<th>0.17</th>
</tr>
</thead>
</table>
What is the term that refers to how fast the velocities of the galaxies increase with their distance from the Earth?  

<table>
<thead>
<tr>
<th>Hubble constant</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.42 0.33 0.33 0.08 0.17 0.43 0.33 0.29 0.40 0.08 0.50 0.00</td>
</tr>
</tbody>
</table>

If the discovery by the Hubble telescope is confirmed then scientists will need to modify their theory because such a heavy and complex atom should not have existed.  

<table>
<thead>
<tr>
<th>heavy and complex</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.58 0.83 0.50 0.42 0.86 0.33 0.75 0.80 0.86 0.67 0.80 0.67</td>
</tr>
</tbody>
</table>

The Hubble telescope found the heavy element boron in extremely ancient stars.  

<table>
<thead>
<tr>
<th>boron</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.67 0.83 1.00 0.67 0.50 0.57 0.67 0.57 0.40 0.75 0.83 0.80</td>
</tr>
</tbody>
</table>

Arthur Eddington said that "We must allow evolution an infinite amount of time to get started."  

<table>
<thead>
<tr>
<th>evolution</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.50 0.83 0.50 0.42 0.33 0.29 0.67 1.00 0.60 0.67 0.67 0.20</td>
</tr>
</tbody>
</table>

Arthur Eddington, who opposed to the idea that there is a beginning of time, became famous for experimentally confirming Einstein's general theory of relativity.  

<table>
<thead>
<tr>
<th>Einstein's general theory of relativity</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.92 0.83 0.83 0.33 0.86 0.83 0.75 0.60 0.86 0.77 1.00 0.67</td>
</tr>
</tbody>
</table>
Shaolin Temple

<table>
<thead>
<tr>
<th>Question</th>
<th>Answer</th>
<th>Low, 20 min</th>
<th>Low, 24 hr</th>
<th>High, 20 min</th>
<th>High, 24 hr</th>
</tr>
</thead>
<tbody>
<tr>
<td>One story suggests that the Shaolin temple was built on a piece of land that had recently been ravaged by fire, because the builders planted new trees.</td>
<td>0.58 0.83 0.67</td>
<td>0.46 0.67 0.67</td>
<td>0.92 1.00 1.00</td>
<td>0.82 1.00 1.00</td>
<td></td>
</tr>
<tr>
<td>What does the word Shaolin mean?</td>
<td>Young forest</td>
<td>0.75 0.67 0.17</td>
<td>0.85 0.83 0.17</td>
<td>0.75 1.00 0.67</td>
<td>0.55 0.80 0.63</td>
</tr>
<tr>
<td>Damo's approach to Buddhism was a radical departure from the traditional Buddhist sutras at that time, which emphasized scholasticism.</td>
<td>0.42 0.33 0.50</td>
<td>0.15 0.33 0.50</td>
<td>0.42 0.33 0.33</td>
<td>0.09 0.25 0.60</td>
<td></td>
</tr>
<tr>
<td>The teachings of Damo are based on self-cultivation through meditation.</td>
<td>0.58 0.67 0.67</td>
<td>0.54 1.00 0.67</td>
<td>0.67 0.83 1.00</td>
<td>0.36 0.80 0.50</td>
<td></td>
</tr>
<tr>
<td>The moving exercises that Damo taught the monks were based on the movements of the 18 main animals (Hint: give a number).</td>
<td>0.33 0.67 0.50</td>
<td>0.62 0.17 0.50</td>
<td>0.50 0.67 0.50</td>
<td>0.27 0.75 0.40</td>
<td></td>
</tr>
<tr>
<td>When Damo joined the monks, he observed that they were weak because most of their routines paralleled those of the Irish Monks.</td>
<td>0.92 0.67 0.83</td>
<td>0.62 1.00 0.50</td>
<td>0.83 0.83 0.83</td>
<td>0.73 1.00 0.63</td>
<td></td>
</tr>
<tr>
<td>The Qing government banned the Shaolin Temple and the practice of martial arts because of fear of rebellion.</td>
<td>0.92 1.00 0.83</td>
<td>0.92 0.83 0.83</td>
<td>1.00 0.67 0.83</td>
<td>0.82 1.00 1.00</td>
<td></td>
</tr>
<tr>
<td>The way for Shaolin to preserve martial arts was to train secretly and to spread it to the layman.</td>
<td>0.58 0.67 0.67</td>
<td>0.31 0.33 0.33</td>
<td>0.58 1.00 0.50</td>
<td>0.55 0.75 0.60</td>
<td></td>
</tr>
<tr>
<td>Damo has two major contributions to Shaolin. One of them is on the philosophical side, which is the introduction of Zen Buddhism.</td>
<td>0.58 0.83 0.00</td>
<td>0.23 0.83 0.50</td>
<td>0.50 1.00 0.33</td>
<td>0.36 0.80 0.43</td>
<td></td>
</tr>
<tr>
<td>Damo has two major contributions to Shaolin. One of them is on the physical side, which is the introduction of Martial Arts.</td>
<td>0.92 0.83 0.83</td>
<td>0.92 1.00 1.00</td>
<td>0.83 1.00 0.50</td>
<td>0.82 1.00 1.00</td>
<td></td>
</tr>
</tbody>
</table>
A few years after Damo's death, he was seen in the mountains of Central Asia carrying ______________________________.

A staff from which hung a single sandal 0.92 1.00 0.50 0.85 1.00 0.67 0.75 1.00 0.67 0.55 1.00 0.75

How long did Damo meditate in the Shaolin mountain?

9 years 1.00 1.00 0.83 0.77 0.83 0.50 1.00 1.00 1.00 0.73 1.00 0.88

What is the name of the stone that Damo faced when he meditated?

wall-facing rock 0.75 0.67 0.50 0.54 0.67 0.17 0.58 0.83 0.83 0.45 0.63 0.80

By the late 1800s, China was divided into national zones, which was similar to _______________ but on a bigger scale.

WWII Berlin 0.58 0.50 0.33 0.62 1.00 0.33 0.92 0.83 0.50 0.27 0.60 0.57

The temporary defeat for the martial artists against military powers of the occupation government led to a modern reformation that included adoption of _______________.

military weapons and tactics 0.50 0.33 0.17 0.31 0.17 0.67 0.50 0.33 0.33 0.45 0.86 0.40

The end to suffering can be achieved by the daily practice of the _______________.

Eight-Fold Path 0.67 0.83 0.33 0.69 0.83 0.50 0.92 1.00 0.83 0.36 0.60 0.50

The teachings of the Buddha are based on the tenets of the _______________.

Four Noble Truths 0.67 0.33 0.67 0.23 0.83 0.83 0.75 0.83 0.50 0.18 0.63 0.60

Practicing the Shaolin martial arts as a _______________ represents a contradiction to Buddhist philosophy.

sport 0.67 0.33 0.50 0.54 0.67 0.50 0.58 0.83 0.33 0.18 0.63 0.40

The practice of Shaolin martial arts should lead to control of the _______________.

go 0.25 0.33 0.33 0.08 0.83 0.33 0.50 0.83 0.17 0.27 0.60 0.00

The Emperor believed that having other people perform good actions in his name is the path to _______________, but Damo disagreed.

Nirvana 0.58 0.50 0.50 0.46 0.67 0.67 0.67 0.67 0.83 0.55 1.00 0.75

When Damo first travelled to China to see the Emperor, the Emperor had started a project that required local Buddhist monks to _______________.

translate Buddhist texts 0.83 1.00 0.83 0.85 1.00 0.83 1.00 1.00 0.67 0.64 1.00 1.00
After Li Shimin was enthroned as the Emperor of the Tang Dynasty, he rewarded Shaolin with ________________ acres of land.

<table>
<thead>
<tr>
<th></th>
<th>0.58</th>
<th>0.83</th>
<th>0.67</th>
<th>0.31</th>
<th>0.83</th>
<th>0.50</th>
<th>0.75</th>
<th>0.83</th>
<th>0.36</th>
<th>0.80</th>
<th>0.75</th>
</tr>
</thead>
</table>

The Tang Dynasty Emperor Li Shimin gave the monks a royal dispensation permitting them to ________________.

|  | 0.58 | 1.00 | 1.00 | 0.54 | 0.83 | 1.00 | 0.83 | 0.50 | 0.83 | 0.73 | 0.86 | 0.60 |
### Appendix D

Object and location names used in Experiment 2.

<table>
<thead>
<tr>
<th>Objects</th>
<th>Locations</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fork</td>
<td>Bottle</td>
</tr>
<tr>
<td>Painting</td>
<td>Pencil</td>
</tr>
<tr>
<td>Mirror</td>
<td>Soda</td>
</tr>
<tr>
<td>Radio</td>
<td>Lock</td>
</tr>
<tr>
<td>Rope</td>
<td>Needle</td>
</tr>
<tr>
<td>Seat</td>
<td>Ribbon</td>
</tr>
<tr>
<td>Dog</td>
<td>Phone</td>
</tr>
<tr>
<td>Table</td>
<td>Match</td>
</tr>
<tr>
<td>Basket</td>
<td>Horn</td>
</tr>
<tr>
<td>Coat</td>
<td>Bench</td>
</tr>
<tr>
<td>Pipe</td>
<td>Book</td>
</tr>
<tr>
<td>Chain</td>
<td>Cake</td>
</tr>
<tr>
<td></td>
<td>Nursery</td>
</tr>
<tr>
<td></td>
<td>University</td>
</tr>
<tr>
<td></td>
<td>Attic</td>
</tr>
<tr>
<td></td>
<td>Store</td>
</tr>
<tr>
<td></td>
<td>Trailer</td>
</tr>
<tr>
<td></td>
<td>Hall</td>
</tr>
<tr>
<td></td>
<td>Palace</td>
</tr>
<tr>
<td></td>
<td>Factory</td>
</tr>
<tr>
<td></td>
<td>Market</td>
</tr>
<tr>
<td></td>
<td>Yard</td>
</tr>
<tr>
<td></td>
<td>Office</td>
</tr>
<tr>
<td></td>
<td>Factory</td>
</tr>
<tr>
<td></td>
<td>Barn</td>
</tr>
<tr>
<td></td>
<td>Apartment</td>
</tr>
<tr>
<td></td>
<td>Warehouse</td>
</tr>
<tr>
<td></td>
<td>Lawn</td>
</tr>
</tbody>
</table>
Appendix E

Instructions for subjects in the high-integration condition in Experiment 2.

In this experiment, you will be asked to memorize a number of objects appearing in a number of locations. For example, you may learn that “the ice cream is in the fridge,” that “the orange is in the fridge,” and that “the rose is in the garden.” I would like you to try your best to remember where each object appears. Some objects will appear in the same location while others will appear in different locations. When the objects appear in the same location, try to form a mental image with all these objects in that location. Each object-location sentence will appear for 8 seconds. Try your best to commit them to memory. After the learning phase, you will be tested on some of these sentences.

Instructions for subjects in the low-integration condition in Experiment 2.

In this experiment, you will be asked to memorize a number of objects appearing in a number of locations. For example, you may learn that “the ice cream is in the fridge,” that “the orange is in the fridge,” and that “the rose is in the garden.” I would like you to try your best to remember where each object appears. Some objects will appear in the same location while others will appear in different locations. Each object-location sentence will appear for 8 seconds. Try your best to commit them to memory. After the learning phase, you will be tested on some of these sentences.