

Iowa State University

From the Selected Works of Jason C.K. Chan

November, 2006

Retrieval-induced facilitation: Initially nontested material can benefit from prior testing of related materia

Jason C.K. Chan, *Washington University in St Louis*

Kathleen B. McDermott, *Washington University in St Louis*

Henry L. Roediger III, *Washington University in St Louis*



SELECTEDWORKS™

Available at: http://works.bepress.com/jason_chan/4/

Running head: RETRIEVAL-INDUCED FACILITATION

Retrieval-induced Facilitation: Initially Nontested Material Can Benefit From Prior Testing of
Related Material

Jason C.K. Chan

Kathleen B. McDermott

Henry L. Roediger, III

Washington University in St. Louis

In press in Journal of Experimental Psychology: General

Correspondence to:

Jason Chan
Department of Psychology
Washington University – Box 1125
St. Louis MO 63130-4899
USA

Phone: 314-935-8892
Email: ckchan@wustl.edu

This is a manuscript of an article from *Journal of Experimental Psychology: General* 135 (2006): 553, doi:10.1037/0096-3445.135.4.553. Posted with permission. This article may not exactly replicate the final version published in the APA journal. It is not the copy of record.

Abstract

Classroom exams can only assess students' knowledge of a subset of the material taught in a course. What are the implications of this approach for long-term retention? Three experiments ($n = 210$) examined how taking an initial test affects later memory for prose materials not initially tested. Experiment 1 showed that testing enhanced recall 24 hours later for the initially nontested material. Importantly, this facilitation was not seen for participants given additional study opportunities without initial testing. Experiment 2 extended this facilitative effect to a within-subjects design. Experiment 3 demonstrated that this facilitation can be modulated by conscious strategies. These results have implications for educational practice and the theoretical developments of the testing effect, associative memory, and retrieval inhibition.

(Word Count: 119)

Keywords: memory, testing effect, semantic priming, retrieval-induced forgetting, education, adjunct questions

Retrieval-induced Facilitation: Initially Nontested Material Can Benefit From Prior Testing of Related Material

Using memory tests as an evaluation tool is a standard practice not only among memory researchers, but also among educators (Bangert-Drowns, Kulik, & Kulik, 1991; Roediger & Karpicke, 2006a; Rohm, Sparzo, & Carson, 1986). While it is true that memory tests can be conceived of as a measurement tool, this conception loses sight of an important role of testing – it is a powerful memory enhancer (Bjork, 1975). For example, when comparing the final, delayed recall performance between two groups of participants, one with and one without an initial recall test following the study episode, the former group consistently outperforms the latter (Abbott, 1909; Gates, 1917; Spitzer, 1939).

Under most circumstances, taking an initial test strengthens recall of the studied material on a subsequent, delayed test. The focus of the current manuscript is on materials *not* directly tested during the initial test (*nontested* materials). The *nontested* materials, as defined in the current manuscript, are limited to materials that are *related* to the tested materials. For example, if students learn that photosynthesis happens during the day and that plants breathe in CO₂ during the day, would recalling the first fact during an initial test affect later memory for the second, initially-nontested fact (relative to a condition in which neither fact was initially tested)?

Findings from the current research have important implications for educators and memory researchers alike. If taking initial tests enhances later memory for the nontested material, then educators might consider increasing the frequency of testing to enhance long-term retention for both the tested and related, nontested materials (Bangert-Drowns et al., 1991; Keys, 1934). On the other hand, if taking initial tests were somehow detrimental to memory for the

nontested material, educators may weigh the benefits of testing against such detriments accordingly. Results from four literatures provide empirical and theoretical bases for predicting three potential patterns of results: that testing might benefit later recall of related materials (a positive effect), that it might not affect later recall of related materials (no effect), or that it might harm later recall of these materials (a negative effect). We briefly review the logic behind these competing predictions.

Positive Effect: Associative Memory and Adjunct Questions

Associative memory theories such as spreading activation (Collins & Loftus, 1975; Collins & Quillian, 1972), Adaptive Control of Thought-Rational (ACT-R, Anderson, 1996), and Search of Associative Memory (SAM, Raaijmakers, 2003; Raaijmakers & Shiffrin, 1981) all share one assumption: activation of one concept in memory should produce facilitative effects for related concepts. For example, spreading activation suggests that activation of one semantic node would spread to other semantic nodes through its associative network (e.g., thinking *bed* would activate *sleep*). Although spreading activation was originally designed as a theory for semantic memory, it has recently been applied quite extensively to episodic memory research (McDermott & Watson, 2001; Roediger, Balota, & Watson, 2001; Roediger & McDermott, 1995).

More relevant to the current question, however, are ACT-R and SAM. Both theories suggest that retrieving a subset of the studied material should facilitate retrieval of the remaining material, provided that some associative links exist between them. In particular, both ACT-R and SAM specify that the probability of activation (or retrieval) of a target item is proportional to the combined strength in which the retrieval context and the retrieval cue elicit memories for that

target item. Applying the logic of SAM and ACT-R to the current context, the probability of retrieving a previously nontested target would be proportional to the associative strength between the tested item and the previously nontested target item. Therefore, if the associative strength between the tested and the nontested item is increased through prior testing (e.g., retrieval of a subset of the studied items activates related concepts in its associative network), the probability of retrieving the previously nontested target should also increase.

The literature on adjunct questions (Frase, 1968; Hamaker, 1986; Rickards, 1979) would also lead one to predict a positive outcome for the nontested material in the current experiments. Adjunct questions are embedded directly in the text that participants study, and participants answer these questions while they are reading the to-be-remembered text. A final test may include questions that are identical to the adjunct questions (henceforth tested), new questions that are related to the adjunct questions (henceforth nontested-related), or new questions that are unrelated to the adjunct questions (henceforth control). Adjunct questions can either precede a paragraph (prequestions) or follow a paragraph (postquestions, a condition that more closely resembles the current research). Hamaker conducted a comprehensive review of the literature and concluded that “postquestions facilitate the learning of material covered either directly [tested] or indirectly [nontested-related] by them; they have no general positive or negative effect on the learning of text material unrelated to them” (p. 236).

No Effect: The Testing Effect and Verbal Learning Literatures

Several researchers in the testing effect and verbal learning literatures have investigated the effect of testing on nontested items using within-subjects designs. An assumption inherent in a within-subjects design is that being tested on a subset of the studied items will not affect

memory for the nontested items (because they serve as control items). A few studies have indeed provided empirical evidence in favor of this assumption (Nungester & Duchastel, 1982; Runquist, 1983; Slamecka & Katsaiti, 1988). However, one critical property of these studies is that no obvious relation exists between the tested and the nontested materials. Importantly, this null effect (that is, no effect of testing on later performance of unrelated, nontested items) fits well within the associative memory frameworks and the adjunct questions literature that we just outlined. Specifically, associative memory theories would *not* predict any facilitative effect for items that are unrelated to the tested items, or that such facilitation, if any, would be negligible. Moreover, as mentioned earlier, the general finding from the adjunct questions literature indicates that adjunct questions do not affect performance on unrelated questions. For cases in which related materials were used, however, conclusive evidence failed to emerge (Duchastel, 1981; LaPorte & Voss, 1975; Mandler & Rabinowitz, 1981). A full consideration of these studies is covered in the General Discussion.

Negative Effect: Retrieval-Induced Forgetting

Counter to the predictions generated from associative memory theories and results on adjunct questions, findings from the retrieval-induced forgetting literature might lead one to predict that performing retrieval practice on a subset of the studied material would inhibit later retrieval of the remaining material (Anderson, 2003; Anderson, Bjork, & Bjork, 1994). A typical retrieval-induced forgetting experiment consists of three phases: (1) a study phase, (2) a retrieval practice phase in which participants practice retrieving only a subset of the studied material, and (3) a final test that includes both materials that did and did not receive retrieval practice. Two important characteristics of this paradigm require clarification. First, materials used in the

retrieval-induced forgetting paradigm are typically category-exemplar word lists (e.g., Fruit: orange, banana, etc.), although other kinds of materials have also been used (see Anderson, 2003, for a review). Second, exemplars that receive retrieval practice are denoted Rp+ (tested), exemplars from these categories that do not receive retrieval practice are denoted Rp- (nontested-related), and exemplars that belong to (entire) categories that do not receive retrieval practice are denoted Nrp (control). Consistent with findings from the testing effect literature, Rp+ items are better recalled than the other two types of items. The more important finding, however, is that final recall performance for Rp- items is worse than for items from the Nrp categories, which signals that retrieval practice reduces accessibility of the Rp- items. The theoretical construct that Anderson, Bjork, and colleagues have advanced to account for the Rp- inhibition can be briefly delineated as follows: When participants perform retrieval practice on a subset of the studied items (e.g., perform retrieval practice for the exemplar *orange* with the probe *Fruit – Or ____*), they must inhibit related competitors (*banana*, an Rp- item) that may intrude (because of mechanisms similar to spreading activation). This inhibition thus manifests as a reduction of subsequent recall for the Rp- items.

At first glance, the logic of retrieval-induced forgetting might lead one to predict negative effects for the nontested-related material in the current experiments. However, a more thorough consideration suggests that predictions emanating from this literature may not be as straightforward as it appears. For example, recent evidence has shown that retrieval-induced forgetting is eliminated (and in one case, reversed, Anderson, Green, & McCulloch, 2000) when participants integrate the study material (Anderson & McCulloch, 1999; Bauml & Hartinger, 2002; Smith & Hunt, 2000) or when a 24-hour delay is inserted between the retrieval practice

and the final test (MacLeod & Macrae, 2001; Saunders & MacLeod, 2002). The first finding suggests that integrative encoding may allow associative activation to overcome inhibition. The second finding suggests that retrieval-induced forgetting, though replicable and robust, is relatively fleeting. Importantly, these findings may lead one to predict *no retrieval-induced forgetting* for the current experiments because we used well-integrated prose materials (Hunt & McDaniel, 1993; Jefferies, Lambon Ralph, & Baddeley, 2004) and inserted a 24-hour delay between the initial tests and the final test. The reasoning behind these specific manipulations is delineated below.

Theoretical and Applied Implications of the Current Experiments

The current experiments are designed to employ educationally-relevant experimental designs and materials in the laboratory; as a result, prose materials were used as our study material and a 24-hour delay was implemented between the initial test and the final test. The prose materials were designed to mimic college-level textbook content, and the 24-hour delay allowed us to examine relatively long-term effects of initial testing. The 24-hour delay between study and final test is also representative of students' typical studying habits (i.e., most students study the night before an exam, Indig, 2005; Leeming, 2002; Michael, 1991). Moreover, materials conducive to spreading activation were created (details to be presented in the Method). If spreading activation does occur among related concepts in testing, then we should be able to observe a benefit for the initially nontested material if it is related to the tested material. In Experiment 1, participants first studied an article about the toucan bird. Afterwards, participants in the *testing* condition performed two successive, identical tests on a subset of the article, participants in the *extra study* condition studied a subset of the article two additional times, and

participants in the *control* condition were dismissed. All participants returned for the final recall test on the entire article 24 hours later. Importantly, the comparison between the testing and the extra study condition allowed us to examine whether the facilitative effect on the nontested-related material, if any, is specific to testing. In Experiment 2, we expanded the material set and sought to conceptually replicate Experiment 1's findings in a within-subjects design. In Experiment 3, we attempt to identify potential mechanisms that are responsible for retrieval-induced facilitation. The logic behind Experiment 3 is discussed after presentation of results from the first two experiments.

Experiment 1

Method

Participants. Eighty-four undergraduate students at Washington University participated for partial fulfillment of a course research requirement or were paid \$5 for each half hour of participation. There were 28 participants in each of the three practice conditions: testing, extra study, and control. Participants were tested individually or in groups of two to three.

Materials. An article about the toucan bird was created for the current experiment. The information contained in this article was taken from several online sources (see Appendix A for details). This article (approximately 2700 words long) contained information about the biological characteristics and living habits of toucans.

Test materials were 40 short answer questions, of which four were filler questions. The 36 target questions contained two *related* sets of 18 questions (denoted as set A and set B). For example, one question in set A was, "Where do toucans sleep at night?" and the related question in set B was, "What other bird species is the toucan related to?" The answer for the former

question was *tree holes* and the answer for the latter question was *woodpeckers*. As was evident in the article, these two answers were related to each other because toucans cannot make tree holes with their soft bills; instead, they sleep in tree holes made by woodpeckers. Related questions were constructed based on both conceptual relations and on where the information appeared in the article. Specifically, most related questions asked details about information that appeared in the same paragraph of the article (see Appendix B for examples). An important criterion in creating these questions was that none of the information presented (including the answer) in one set would answer the questions in the other set. For example, woodpeckers never appeared in set A questions and tree holes never appeared in set B questions. Question sets were counterbalanced across participants (see Appendix B for a sample of the materials used in this experiment). All materials used in the current study are available as supplementary materials online and upon request.

One may question what constitutes relatedness. For our purposes, we broadly defined related information as any *pair* of facts in the article that (a) are conceptually similar or (b) appeared in close proximity physically and thus also episodically (e.g., in two consecutive sentences). Quantitative estimates of relatedness can be obtained using Latent Semantic Analysis (Foltz, Kintsch, & Landauer, 1998; Landauer, Foltz, & Laham, 1998), which uses a multi-dimensional semantic space to analyze relatedness of verbal materials based on their co-occurrence in everyday language.

To provide some metric of relatedness (apart from our intuition), we obtained relatedness ratings via the sentence comparison feature on the Latent Semantic Analysis website at <http://lsa.colorado.edu>. Specifically, we obtained the relatedness of each pair of questions either

with normal pairing (e.g., by computing the relatedness between the “toucans sleep” question and the “species related to toucan” question) or randomized pairing (e.g., by computing the relatedness between the “toucans sleep” question and a randomly paired question such as “what is the most colorful toucan species?”). If the question pairs were related based on specific information in addition to the fact that they were related to the topic of toucans, then the averaged relatedness of the normal question pairs should exceed the random question pairs. Indeed, this was the case. The averaged relatedness rating (a correlation) was .38 for normal pairing and .07 for randomized pairing, $t(34) = 4.19$, $d = 1.44$.

Procedure. The experiment was conducted in two sessions scheduled 24 hours apart. During the Day 1 session, all participants were given 25 minutes to read the article under intentional learning instructions. They were told that if they finished reading the article early, they should re-read the article until time expired. Afterwards, participants in the *testing* condition answered 22 questions on the computer one at a time for 25 seconds each (18 target questions from set A/B and 4 filler questions). Participants typed in their answers and were told to be as accurate as possible and not to guess. The same test was administered twice in immediate succession (with a different random order of questions for each test). No corrective feedback was given. Participants in the *extra study* condition read 22 statements twice for 25s each (with a fresh random order for each block of 22 statements). For example, one statement was, “Toucans sleep in tree holes at night.” The extra study condition is similar to providing students with all the questions and their answers for an upcoming test (essentially a “cheating” condition). This extra study condition served to isolate the effect of retrieval from simply strengthening a subset of the studied items via repeated exposures. Participants in the control

group were dismissed after reading the article.

On the second day, participants took a final test with 40 questions (18 set A, 18 set B, 4 fillers). The initially nontested questions were always presented in the first half of the experiment to avoid output interference (Roediger, 1974) that may mask a facilitation effect. Participants were given 30 seconds to answer each question, and they were again told not to guess.

Results and Discussion

The critical results are displayed in Figure 1. As can be seen from the right side of the figure, which displays results for questions not initially practiced on Day 1, participants who had been tested on related questions on Day 1 (the Testing group, white bar) outperformed participants who had received extra study on related questions on Day 1 (the Extra Study group, the middle bar) or participants in the control condition (the Control group, the right bar). Surprisingly, this 9% facilitative is retrieval-specific because participants in the extra study condition showed no such benefit. To our knowledge, this is the first demonstration of a *retrieval-specific* facilitative effect on recall of previously nontested (yet related) materials. Figure 1 also shows a robust testing effect for items directly tested on Day 1 even though no feedback was given (compare the leftmost to the rightmost bar of the entire figure). We now describe results from inferential statistical analyses; the full set of means underlying these analyses can be seen in Table 1. All analyses were done with an alpha level of .05 (two-tailed unless stated otherwise). Partial eta squared (η^2) indicates effect size for omnibus tests and Cohen's d indicates effect size for t-tests.

Recall Probabilities on Day 1. As can be seen in Table 1, participants performed slightly

better on the second test (.72) than they did on the first test (.69), $t(27) = 2.22$, $d = .23$, a hypermnesia effect (Erdelyi & Becker, 1974). A closer examination revealed that 14 of the 28 participants showed hypermnesia, whereas only 5 showed overall inter-test forgetting.

Recall Probabilities on Day 2. Separate analyses were conducted for questions that had been presented on Day 1 (either via testing or extra study) and those not presented on Day 1. We first discuss results for the *presented* questions. An independent samples t-test indicates that participants in the extra study condition (.79) outperformed participants in the testing condition (.71), $t(54) = .60$, $d = .57$ (see left side of Figure 1). This result was somewhat surprising given that previous studies have shown the opposite pattern (Hogan & Kintsch, 1971; Roediger & Karpicke, 2006a), although two methodological characteristics of our experiment might have contributed to this discrepancy. First, participants in our extra study condition only performed additional study opportunities on a subset of the original material, whereas participants in other studies re-studied the entire content of the original material. Second, previous studies typically used a longer delay (7 days in Roediger and Karpicke and 48 hours in Hogan and Kintsch). Therefore, it is perfectly possible that the repeated study superiority would not hold up under all circumstances (e.g., if we had used a longer retention interval).

A comparison between the leftmost bar and the rightmost bar in Figure 1 reveals a testing effect, $t(54) = 5.42$, $d = 1.35$. That is, participants in the testing condition (.71) outperformed participants in the control condition (.50) on Day 2. Some researchers have suggested that it is important to tease apart the effect of testing into two components: retrieval and repeated exposure of the studied material (Carrier & Pashler, 1992; Kuo & Hirshman, 1996). Specifically, when participants recall an item from memory during an initial test, the recalled

item can serve as an extra study episode; therefore, the effect of prior retrieval on later memory performance can only be shown through a comparison between initially recalled items and items that have been repeatedly studied sans retrieval processes. Although partialing the effect of re-encoding from retrieval is not the major interest of the present research, one can isolate the effect of retrieval by comparing the conditional probability of final recall given initial recall (test 2 of Day 1) for the testing group to the probability of final recall for the extra study group. The probability of final recall given initial recall ($M = .95$, $SD = .06$) revealed that hardly any forgetting occurred for the recalled items during the 24-hour retention interval. Clearly, this conditional probability was higher than the final recall probability of the extra study group (.79), $t(54) = 5.34$, $d = 1.49$.

Although this comparison shows that taking an initial test almost eliminated forgetting, it suffers from possible item selection issues. That is, items that participants could recall during the initial tests might be particularly easy (for whatever reason). Therefore, the s score, which provides an estimate of the influence of initial retrieval on later retrieval with item selection partialled out (Modigliani, 1976), was computed to compare the testing condition with the extra study condition. An s score that is greater than 0 means that initial retrieval has asserted a positive influence on subsequent retrieval compared to the extra study condition over and above any item selection artifacts. Detailed instructions on how to calculate the s score can be found in McDaniel and Masson (1985; see also, Modigliani, 1976). The s score for this comparison (i.e., testing vs. extra study) was .21, which indicates that initial retrieval asserted a larger beneficial effect on the later memorability of previously recalled information than did restudying, even when item selection has been taken into account.

More important for the current purposes were results from the nontested questions (i.e., questions not presented on Day 1). An analysis of variance (ANOVA) revealed a main effect of practice condition (testing, extra study, control) on recall probabilities, $F(2, 81) = 4.07$, $p_{es} = .09$. Fisher's protected t-tests confirmed that the testing group outperformed the extra study group, $t(54) = 2.63$, $d = .69$, and the control group, $t(54) = 2.32$, $d = .56$, whereas the extra study and the control groups had nearly identical performance, $t < 1$. As stated previously, the fact that facilitation of the nontested items is only seen in the testing condition indicates that this is a retrieval-based effect.

To further examine this retrieval-induced facilitation effect, we plotted cumulative recall curves as a function of response time (RT) during the final test for participants in the three practice conditions (cf. Roediger & Thorpe, 1978; Wixted & Rohrer, 1994). As can be seen in Figure 2, the advantage that the testing group had over the control and extra study groups seemed to grow with response time. Specifically, no recall advantage was seen for the testing group over the control group for responses that occurred within 5s of the stimulus (question) onset. However, the advantage for the testing group started to emerge for responses that occurred between 5-10s after the stimulus onset, and this advantage grew bigger as response time increased. To quantify this interaction, we separated response times into two categories, one from 0-15s and the other from 16-30s. A 2 (response time) X 3 (practice condition) mixed ANOVA showed that this interaction was significant, $F(2, 81) = 3.31$, $p_{es} = .08$. The fact that retrieval-induced facilitation increased with response time may be interpreted as implying a consciously driven search mechanism. For example, participants might employ a memory search strategy in which they used related information to help retrieve target information, and the longer

they had to retrieve such related information, the more likely it would be for them to retrieve the target. In contrast, if automatic spreading activation were the sole cause of the facilitative effect, one would not expect facilitation to increase with response time or that it would take more than 5 seconds for facilitation to surface. This evidence, coupled with the lack of facilitation in the extra study condition, casts serious doubts on the validity of an account that relies solely on automatic activation. We return to a more detail analysis of this effect after presentation of results in Experiment 2.

Because of the relative novelty of retrieval-induced facilitation, we sought to replicate it using a within-subjects design in Experiment 2. We also expanded our material set in an attempt to increase the generality of our findings. The extra study condition was dropped in Experiment 2 but used again in Experiment 3.

Experiment 2

Method

Participants. Seventy-two undergraduate students from Washington University participated.

Materials. In addition to the toucan article used in Experiment 1, three new articles were created for this experiment: (1) The Big Bang Theory, (2) The History of Hong Kong, and (3) The Shaolin Temple (see Appendix for sources). These topics were chosen because they were relatively unfamiliar to most undergraduate psychology majors. Each of the four articles, including the toucan article (which was shortened for this experiment), was approximately 1900 words long. Each participant read only two of the four articles, and the four articles were rotated across participants in a Latin square design so that each article served equally often as the

experimental (with initial tests) and the *control* (without initial tests) article. We created 24 questions for each of the four articles. Similar to Experiment 1, the 24 questions were divided into two related sets of 12 questions each. Unlike in Experiment 1, however, no filler questions were included. Assignment of question sets to the tested and the nontested conditions (for the experimental article) was counterbalanced across participants.

Relatedness ratings were again obtained for questions in each article using Latent Semantic Analysis. The new toucan questions now had a relatedness rating of .36, the Big Bang questions .39, the Shaolin questions .26, and the Hong Kong questions .34. These relatedness ratings were always higher than when pairings were randomized (.16, .18, .10, and .08, respectively), all $ps < .05$ except for the toucan comparison, in which $t(22) = 1.85, p = .08$.

Procedure. Participants studied the two articles with intentional learning instructions and were told that they may or may not be tested immediately following each article. Study order of the control and the experimental article was random. Participants were given 16 minutes to read each article and 25 seconds to answer each question¹.

The final recall test was scheduled 24 hours after the first session. It contained 48 questions that can be separated into three groups: (1) 12 questions tested twice on Day 1 from the experimental article (*tested*), (2) 12 nontested questions from the experimental article (*nontested-related*), and (3) 24 questions from the *control* article. Participants had 30 seconds to answer each question. Whether questions for the control or the experimental article would be presented first was determined randomly, but the nontested questions for the experimental article were always presented first to avoid output interference².

Results and Discussion

An examination of the two bars on the right of Figure 3 reveals a retrieval-induced facilitation effect for the nontested-related questions. The magnitude of the effect was similar to that obtained in Experiment 1 even though there were fewer items and fewer participants in this experiment (due to the within-subjects design).

Recall Probabilities on Day 1. Similar to Experiment 1, a small effect of hypermnesia was observed when recall performance was compared between test 1 (.65) and test 2 (.67) on Day 1, although the significance level of this comparison was marginal by conventional standards, $t(71) = 1.76$, $d = .11$, $p = .08$. Out of the 72 participants, 21 showed hypermnesia and 13 showed inter-test forgetting.

Recall Probabilities on Day 2. A repeated-measures ANOVA indicated a main effect of question type (tested, nontested-related, control), $F(2, 70) = 68.29$, $p = .66$. Not surprisingly, participants performed better on the tested questions (.66) than both the nontested-related questions (.49), $t(71) = 7.71$, $d = .89$, and the control questions (.42), $t(71) = 11.40$, $d = 1.41$. Moreover, replicating the finding from Experiment 1, if participants had answered a question correctly on the second test of Day 1, they almost always answered the same question correctly on Day 2 (.92, $SD = .13$, s score compared to control = .56). Most important for the current purposes was that participants performed better on the nontested-related questions than on the control questions, $t(71) = 3.21$, $d = .39$. This finding represents a retrieval-induced facilitation effect in a within-subjects design, thus conceptually replicating the finding from Experiment 1.

We plotted cumulative recall curves for the control and nontested-related questions as a function of response time (Figure 4). Similar to the pattern in Experiment 1, we only observed retrieval-induced facilitation for slower responses, and the facilitation seemed to increase with

response time. In fact, there was no sign of facilitation for the nontested-related questions over the control questions within the first 15 seconds of the question onset, which was evidenced by the significant interaction between response time (1-15s, 16-30s) and item type (nontested-related, control), $F(1, 70) = 8.80$, $p_{es} = .11$. Again, this finding might be taken as support for the involvement of a conscious search mechanism that tends to reveal its benefits later in a recall period.

An additional response time analysis was possible for this experiment due to its within-subjects design. That is, we could ask whether the magnitude of retrieval-induced facilitation varies with how long participants took to answer questions on Day 1. If retrieval-induced facilitation is caused by an active search process during the initial tests, then the longer it took participants to answer questions during the initial test, the more likely it would be for participants to activate related information. This notion is similar to ideas in the feeling-of-knowing (FOK) and tip-of-the-tongue (TOT) literatures. That is, when participants attempt to retrieve some information in memory, they tend to activate partial information of the target item and its associates, and these associated thoughts can come to participants' conscious awareness (Koriat, 1993; Schwartz & Smith, 1997). Applying this logic to the current scenario, when participants face a question that they have trouble answering, they may search for related information that might then lead to the correct answer. Moreover, related information is more likely to be activated if participants have more time to search for the correct answers. The prediction that emerges from this analysis is that participants who took longer to answer the questions during the initial tests should show greater retrieval-induced facilitation than those who were relatively fast. To test this prediction, we conducted a median split analysis based on participants' mean

correct response time on Day 1. Specifically, we averaged the correct response times of the two initial tests on Day 1 for each participant and classified participants into a fast (mean correct RT = 8.5s) and a slow (mean correct RT = 12.4s) group. An examination of Figure 5 shows that retrieval-induced facilitation was particularly apparent (12%) among participants who took longer to answer questions correctly during the initial tests (although there was still a 4% facilitative effect for the fast participants), which was exemplified by an interaction between response time on Day 1 (fast, slow) and item type (nontested-related, control), $F(1, 69) = 3.48$, $p = .05$, $p < .10$. This finding is again supportive of a conscious retrieval/search mechanism behind retrieval-induced facilitation. A closer examination of the data suggests that retrieval-induced facilitation was particularly beneficial to the slower participants, who, on average, exhibited worse recall performance than the faster participants. This conclusion was reached because slower participants exhibited lower recall performance on both the control (8% worse) and tested (7% worse) questions. However, the benefit of retrieval-induced facilitation allowed the slower participants to perform as well as the faster participants on the nontested-related questions.

Taken together, the response time analyses from Experiments 1 and 2 suggest that a conscious search mechanism might be responsible for the facilitative effect of initial testing on related materials. However, these results, though consistent and interesting, are correlational, so they do not provide definitive evidence in support of a conscious search model. In Experiment 3, we sought to provide more solid support for a conscious search model by manipulating the way participants retrieve information during the initial tests.

Experiment 3

The purpose of Experiment 3 was to uncover potential underlying mechanisms that are responsible for retrieval-induced facilitation. To that end, it is beneficial to briefly summarize the major findings we have accumulated thus far. First, as predicted, taking an initial test on a subset of studied material facilitates later retrieval of related, but previously nontested, material. Importantly, this facilitation was observed only when retrieval was required during the initial tests – no facilitation occurred for participants who restudied the original material. Further, the effect of retrieval-induced facilitation increased with both response time during the initial tests and during the final test. Taken together, these results made us reconsider the theoretical framework behind retrieval-induced facilitation.

Originally, we hypothesized that facilitation of the nontested-related material would support an automatic spreading/associative activation account. However, the retrieval-specificity and slow nature of the effect required us to abandon a framework that relies *solely* on automatic processes. Indeed, we know of no spreading activation account that would have predicted the retrieval-specificity of this facilitation or its gradual nature in arising over time. To that end, a dual-process account of retrieval-induced facilitation is postulated: In addition to automatic activation, we hypothesize that facilitation of related facts occurred because, during the initial tests, participants spontaneously adopted a retrieval strategy that broadened their search efforts. For example, when participants attempted to answer “where do toucans sleep at night?” nontarget memories such as trees and woodpeckers might receive activation via automatic spreading activation or because participants were actively searching for facts related to the target memory. Participants might adopt such a search strategy if they believed that retrieval of related facts would help them in retrieving the correct response (via more retrieval cues). If this were

the case, then the longer it took participants to provide a correct answer, the more likely related information would come to mind, which would then be exhibited as retrieval-induced facilitation in our paradigm. One advantage to this dual-process account is that it can explain the retrieval-specificity of the facilitation. According to this account, participants in the extra study condition simply reviewed the studied facts; there was no demand for active, conscious retrieval of the target and its related information. As a result, long lasting facilitation of related facts is not expected in this condition.

We formally tested this conscious search hypothesis by experimentally manipulating the retrieval strategies participants were instructed to use during the initial tests in Experiment 3. Specifically, participants were told to adopt either a *broad* or a *narrow* retrieval strategy during the initial recall tests. Retrieval strategies used during the Day 2 final test, however, were held constant. If participants adopt a broad/general retrieval strategy during the initial tests and this is indeed the cause of retrieval-induced facilitation, then retrieval-induced facilitation should be minimized when they are told to use a narrow retrieval strategy. In addition, we attempted to replicate the retrieval-specificity effect by including an extra study condition. The prediction is that retrieval-induced facilitation would occur only in the broad retrieval condition.

Method

Participants. Fifty-four undergraduate students at Washington University participated in this experiment. Each initial practice condition (broad, narrow, extra study) contained 18 participants. Participants' age ranged from 19 to 25 ($M = 19.80$, $SD = 1.49$) and there were 38 females and 16 males.

Materials. Articles for the Shaolin temple and the Big Bang theory were used in this

experiment. These articles had produced the most consistent retrieval-induced facilitation in Experiment 2. Participants studied two articles and one of them served as the experimental article and the other the control article.

Procedure. The experimental procedures were similar to those in Experiment 2 except that (a) participants in the extra study group were told that they would get a chance to restudy (on the computer) some of the facts presented in the article, though they would not know which article they would restudy, (b) participants in the two testing groups (broad, narrow) were given specific instructions on how to approach the initial tests. Because the instructions were critical to our manipulation, they are represented here in their entirety. The broad and narrow retrieval instructions are identical for the first half but different for the second half, which are presented here in *Italics* typeface. Instructions for the broad condition appear in the first paragraph, which is followed by instructions for the narrow condition.

We want you to do this memory test in a specific manner. When you work on a normal memory test and you encounter a question that you have trouble answering, you may try to think of everything you remember that is related to the question at hand. For example, if a question asks you to name the state that is located just east of Missouri, you may think of all the states surrounding Missouri such as Iowa, Kansas, Illinois, Indiana, Arkansas, etc., and realize that Illinois is the correct answer, because Iowa is north of Missouri, Arkansas is south of Missouri, etc. Such broad and general retrieval strategies may help you in locating the correct fact and to help you in remembering related details. You may do this very explicitly (think of all the related facts in an organized manner) or

more automatically (all the related ideas just pop into your head and then you just pick the one that seems right). *I would like you to MAXIMIZE this type of broad retrieval strategy. That is, when you are answering the upcoming questions, try to think of all the facts that are related to the question at hand from the article that you just read. Such a strategy should help you better answer the questions. In addition, if you have time left after you answer the question, you should STILL try to remember facts that are related to the question at hand. Do this for as long as possible (you will have 25 seconds for each question). Although you are encouraged to think of as many related facts as possible, it is also important that you try to answer the questions with only the correct answer and not some random facts that you generated in this process (this is not a free association task). In other words, do not guess.*

The following are instructions specific to the narrow retrieval condition:

In this memory test, I would like you to MINIMIZE this type of broad retrieval strategy. That is, when you are answering the upcoming questions, try to come up with the correct answers without thinking about anything else. Just think about the question and remember the answer. You may find this relatively difficult to do, but please do try your best to do it. One way to do this is not to allow yourself to think about anything other than what directly answers the question (probably more similar to answering fill-in-the-blanks questions than to essay questions). Answer the questions CONCISELY; do not guess. Once you have answered the

questions, press the “Enter” key and the next question will appear. You will have 25 seconds for each question.

Note that participants in both the broad and extra study condition had exactly 25 seconds per question/restudy statement, but participants in the narrow condition had *up to* 25 seconds per question. Participants in the narrow condition were allowed to advance the trials by pressing the Enter key. This manipulation was included to stop any potential post-retrieval thoughts that may activate related concepts.³

On Day 2, all participants completed the final cued recall test on both articles. The test instructions informed participants in the two testing conditions that they should abandon the retrieval strategy they were told to use on Day 1 and instead complete the tests as they would normally in a real exam. All participants, including participants in the extra study condition, were told to use whatever strategies they could to maximize their performance but were discouraged from guessing. After participants in the broad and narrow conditions completed the recall tests for both articles, a post-test question assessed how well they were able to implement the broad/narrow retrieval strategy the day before. For obvious reasons, participants in the extra study condition did not encounter this post-test question.

Results and Discussion

An examination of Figure 6 shows that, as predicted, retrieval-induced facilitation for the nontested-related questions was seen only in the broad retrieval condition – no facilitation for the nontested-related questions was observed in the narrow or extra study condition.

Recall Probabilities on Day 1. One possible concern with the broad and narrow retrieval instructions was that they would affect participants' response criteria. For example, it was possible for participants in the broad condition to interpret the instructions as encouraging them to guess, whereas the opposite would happen to participants in the narrow condition. If this were the case, participants in the broad condition would answer more questions during the initial tests than participants in the narrow condition. Such concerns seem unwarranted as results in Table 3 show that participants in the narrow condition did not leave more questions unanswered (.13 for Test 1 and .09 for Test 2) than participants in the broad condition (.16 for Test 1 and .13 for Test 2) in both Test 1 and Test 2, both $t_s < 1$.

Correct recall probabilities also supported the idea that there was no difference in response criterion between the two groups. An examination of Table 3 reveals remarkable similarity in initial recall performance between participants in the broad and narrow conditions. There was a main effect of test number (test 1, test 2), $F(1, 34) = 7.02$, $p_{es} = .17$, but neither the main effect of retrieval instructions (broad, narrow) nor the interaction was significant, both $F_s < 1$. The main effect of test number indicated a small but reliable hypermnesia effect similar to those in Experiment 1 and 2. That is, on average, participants performed better in the second recall test (.66) than in the first (.62). Most important for the current purpose, there was no difference in recall performance between the broad and narrow groups by the end of the second test on Day 1 (.66 for broad and .65 for narrow). We now examine the effects of different initial practice conditions on Day 2 recall performance.

Recall Probabilities on Day 2. A 3 (practice condition: broad, narrow, extra study) X 3 (question type: tested, nontested-related, control) mixed ANOVA indicated a significant effect of

question type, $F(2, 50) = 58.57$, $pes = .70$, but neither the main effect of practice condition, $F < 1$, nor the interaction between question type and practice condition was significant, $F < 2$, $p > .10$. The main effect of question type basically indicated a testing/extra study effect. That is, overall, correct recall probability was higher for initially practiced questions (.66) than for both nontested-related (.44) and control questions (.42). Planned comparisons between the recall probabilities of tested/restudied questions and control questions showed that the benefits of testing or extra study were apparent in all three practice conditions, all $ts > 5$, $ds > 1.21$. In fact, the magnitude of this initial practice effect was remarkably similar across conditions (all between 23-25%), which is also evident from Figure 6 (comparing the leftmost with the rightmost bar in each Day 1 practice condition). Interestingly, the benefit of testing now equaled that of extra studying, which is different from the pattern in Experiment 1. The discrepancy could be due to some methodological differences between Experiment 1 and 3 such as different articles being used, varying amounts of study times, and the within-, as opposed to between-, subjects design of Experiment 3. Regardless of the reason for this discrepancy, the question of interest in this experiment was whether different retrieval instructions during the initial tests effected different magnitudes of retrieval-induced facilitation. To that end, planned comparisons were conducted to examine differences in recall probabilities between the nontested-related questions and the control questions for the three practice conditions.

Replicating the retrieval-specificity effect in Experiment 1, there was no sign of facilitation for the nontested-related questions (.38) over the control questions (.42) in the extra study condition, $t < 1$. There was also no hint of retrieval-induced facilitation in the narrow condition, $t < 1$. In fact, the recall probabilities for the nontested-related questions and the

control questions were identical (.44) in this condition. Importantly, there was a significant benefit associated with initial testing in the broad retrieval condition, $t(17) = 2.83$, $d = .54$, where initial testing led to enhanced recall for nontested-related questions (.51) relative to control questions (.41).

Cumulative recall curves like those reported in Experiment 1 and 2 are displayed in Figure 7. As can be seen, results in the broad retrieval condition (see top panel) mirrored those in the two previous experiments. That is, as response time increased so did the magnitude of retrieval-induced facilitation, although the interaction between item type (control, nontested-related) and response time (1-15s, 16-30s) was not statistically significant, $F(1, 17) = 2.8$, $pes = .14$, $p = .11$, observed power = .35. In contrast, an examination of the bottom two panels (for the narrow and extra study conditions) reveals no retrieval-induced facilitation regardless of response time.

A median split analysis similar to that in Experiment 2 was also conducted here for participants in the broad and the narrow conditions (see Figure 8). The top panel (broad condition) of Figure 8 shows that participants who answered (correctly) faster on Day 1 showed less retrieval-induced facilitation than participants who answered slower, again suggesting a conscious mechanism at work. These results replicated those in Experiment 2, and they both seemed to indicate that retrieval-induced facilitation might be particularly influential among participants who had poorer recall (as shown in the recall probabilities of the control and tested items) and were slower to answer questions, $F(1, 16) = 3.27$, $pes = .17$, $p < .10$. Furthermore, similar to Experiment 2, slower participants seemed to benefit more from retrieval-induced facilitation. The slower participants again performed more poorly than the faster participants on

both the control (14% worse) and the tested (7% worse) questions but not the nontested-related questions (only 1% worse).

Data from participants in the narrow condition (see bottom panel of Figure 9), however, revealed an interesting pattern that differed dramatically from the pattern in the broad condition - there was a crossover interaction between participants' response time on Day 1 and the magnitude of retrieval-induced facilitation, $F(1, 16) = 5.45, p = .025$. Specifically, the slower participants performed better on the nontested-related questions than the control questions but this effect was reversed among the faster participants. That is, the faster participants in the narrow condition actually showed a tendency towards retrieval-induced forgetting. This crossover interaction was unexpected, and we are not able to provide a definitive explanation for this effect. However, to speculate, it was possible that in an attempt to come up with the correct answer during the initial tests without thinking about related information, participants in the narrow condition might have deliberately attempted to inhibit related memories, and faster participants had better inhibitory control than slower participants. There are reasons to believe that faster participants might be better inhibitors than slower participants, when one takes into account that processing speed is often correlated with inhibitory control (Salthouse, 1996; Salthouse & Meinz, 1995). This explanation is admittedly post-hoc and speculative, but it served as a potential starting point in explaining this interesting dissociation. Another way to look at this data set is to compare faster to slower participants. Similar to the data reported thus far, the slower participants generally exhibited lower recall performance than the faster participants in both the control (6% worse) and tested (7% worse) questions. However, in the narrow condition, the slower participants actually outperformed the faster participants on the

nontested-related questions (7% better), which again demonstrated that retrieval-induced facilitation might be particularly beneficial to those who typically perform worse on recall.

A post-test question administered at the end of the experiment on Day 2 asked participants to rate how well they were able to perform the broad/narrow retrieval task on Day 1 on a five-point scale, with 1 representing “I did not try to use a broad/narrow retrieval strategy at all” and 5 representing “it wasn’t a problem at all. That’s what I normally do anyway”. Interestingly, results from this post-test question indicated that participants in the broad and narrow condition did not perceive either of the retrieval tasks as more difficult, at least in retrospect, ($M = 3.44$ for broad and $M = 3.39$ for narrow), $t < 1$.

General Discussion

In three experiments, we have demonstrated that initial testing can benefit later retention for related, yet nontested materials. Importantly, no such benefits occurred for participants who were given additional study opportunities in the absence of testing or when they were told to employ a narrow retrieval strategy during the initial tests. Further, the magnitude of retrieval-induced facilitation increases with response time and is greater among participants who are slower and have lower baseline recall probabilities. We now discuss the theoretical and practical implications of these findings.

Retrieval-Induced Facilitation and Inhibition Theories

On face value, the current results appear to be at odds with findings from the retrieval-induced forgetting literature, but a closer examination of that literature suggests that this is not necessarily the case. In particular, as we have mentioned in the Introduction, several boundary conditions have been found for retrieval-induced forgetting. For example, integrative processing

(Anderson, 2003) during encoding, long delays between retrieval practice and the criterial test (MacLeod & Macrae, 2001), nonrandom selection of items presented during the retrieval practice phase (Dodd, Castel, & Roberts, 2006) and perceptual implicit tests such as word fragment completion (Butler, Williams, Zacks, & Maki, 2001; Camp, Pecher, & Schmidt, 2005) typically eliminate retrieval-induced forgetting. The effects of integration and delay on retrieval-induced forgetting bear particular importance to the current study and the application of our findings to educational domains. Our experiments were designed as analogs of real-world educational situations, where coherent text material and 24-hour delays are typical. Because the detrimental effects of retrieval induced-forgetting did not occur under these situations, we can make a relatively unreserved recommendation for frequent testing in the classroom: testing aids later recall of both tested and nontested (but related) material.

Other than the nature of the study materials, another major methodological difference exists between the retrieval-induced forgetting paradigm and the current paradigm. Whereas the Rp+ (tested) and Rp- (nontested-related) items share the same retrieval cue (i.e., a common category) in the retrieval-induced forgetting paradigm, the corresponding items in the current experiments do not share the same retrieval cue (i.e., different questions probe different, yet related concepts). Using unique retrieval cues for the tested and nontested-related questions should reduce the possibility of retrieval competition between related concepts, which is a hallmark of retrieval-induced inhibition. In fact, the retrieval dynamics in a typical retrieval-induced forgetting procedure and the current procedure may be quite different. In retrieval-induced forgetting experiments, attempting to recall the target item for the probe Fruit – Or _____ may automatically trigger the thought of apple, which must be inhibited to avoid its intrusion.

This inhibition is usually assumed to be unconscious. In contrast, we have proposed that processes that lead to retrieval-induced facilitation are subject to conscious control. A more thorough discussion of the roles that controlled and automatic processes may play in the facilitation of nontested-related material will follow shortly (see also, Chan & McDermott, 2006).

One concern with the current experiments is whether facilitation only occurs when there is a substantial (namely, 24-hour) delay between the initial test and the final test because all three of our experiments used this delay. Since retrieval-induced forgetting is usually found in tests with short delays (typically 20 minutes) but not with long delays, one may wonder whether retrieval-induced forgetting would have occurred had we used a shorter delay. To address this issue, another experiment has been conducted in our lab in which we have observed retrieval-induced facilitation in a single test session. In this experiment, participants studied an article and then performed three successive tests that were each separated by a brief period (5 min) of filler activity. The first and third tests were always identical. The major manipulation was the nature of the second, intervening test, which was either a test identical to the first and third test; a test that asked questions related to, but were not covered in the first and third test (a nontested-related test); or a test that asked questions about a completely different, nonstudied topic. Performance on the third test in comparison to the first test (i.e., Test 3 – Test 1 performance) served as an index of the effect of intervening activity. Of interest is whether participants who took the related intervening test would show a bigger benefit on Test 3 (relative to Test 1) than participants who took the unrelated intervening test. If retrieval-induced facilitation occurs in a single test session, then participants who took the related intervening test should perform better

than participants who took the unrelated intervening test. Indeed, this was the case. Specifically, participants who took the related intervening test showed a 4% enhancement in their final test performance relative to the initial test, $t(61) = 2.55$, $d = .19$, whereas no such benefit was seen among participants who took the unrelated intervening test (0.0% improvement). Remarkably, participants who took the related intervening test outperformed (though not significantly) even those who took the same test three times (3% improvement from test 1 to test 3)! We present this experiment here in the General Discussion rather than as a separate experiment because (a) it used a slightly different method compared to our other experiments, and (b) it did not directly address the question of interest, that is, what are some potential underlying mechanisms for retrieval-induced facilitation? The outcome of this experiment was straightforward; therefore, we believe it is appropriate to present this data set in a shortened format.

Previous Findings Consistent with or in Conflict with our Results

As mentioned in the Introduction, several prior studies have examined issues similar to the current research, but conclusive evidence failed to emerge from these studies (Duchastel, 1981; LaPorte & Voss, 1975; Mandler & Rabinowitz, 1981). We now provide a brief review of these studies and raise possible reasons for their inconsistencies.

LaPorte and Voss (1975) used prose materials and failed to show any facilitation or inhibition for the nontested material one-week following the initial test. However, since no sample materials were provided, it is unclear whether the tested and nontested questions were related. For example, one of the articles that LaPorte and Voss used was a biography of Sir Isaac Newton. Suppose one question in the tested condition asked for the date of birth of Newton, and one question in the nontested condition asked for the name of the university that Newton

attended. From this example, one would not expect retrieving the date of birth for Newton to facilitate later recall for the name of the university he attended. Therefore, it may be crucial to establish a clear link between the tested and nontested materials in order to examine subsequent effects of initial testing on the nontested material. Moreover, the complexity of their study material might have contributed to their null results. They noted that 100 questions were created for a 1500-word article (in contrast, we created 36 questions for a 2700-word article in Experiment 1 and 24 questions for the 1900-word articles in Experiment 2), meaning that a question was created for every 15 words of the article. We suspect that in order to create 100 usable questions for such a short article, the article would have to be filled with individual facts. Such a narrative structure might have reduced the coherent, integrative nature of prose, which might have limited participants' ability to integrate the study material.

Duchastel (1981) reported finding no facilitation or inhibition for nontested material that was related to the tested material after a 2-week delay; however, upon closer examination of the data he reported in Table 2, it appears that facilitation of the nontested materials did occur. Specifically, Duchastel compared the recall performance between the testing group and the control group on three types of questions: tested, nontested-related, and control questions. However, since the control group never received an initial test, there should be no "tested" and "nontested-related" questions per se for these participants. It is unclear whether the questions that appeared in the initial test were counterbalanced across participants (i.e., whether the initially-tested questions for one participant would serve as initially-nontested questions for a different participant). If the questions were not counterbalanced, then the results might be subject to item selection problems. In contrast, if the questions were counterbalanced, then the

recall scores on all the questions for the control participants should be collapsed. When this is done, it can be seen that participants who took an initial test ($M = .47$, $SD = .16$) outperformed control participants on the nontested-related questions ($M = .31$, SD pooled over question type = .18). Moreover, even when no collapsing is done for the control group, there was still a trend for the testing group to outperform the control group (.23 vs. .18 for recall of topics and .47 vs .39 for short answers cued recall).

Mandler and Rabinowitz (1981) conducted an experiment in which participants studied category-exemplar lists, which was followed immediately by a recognition test and then a recall test a week later. The important finding is that for categories in which half of the exemplars were presented during the initial recognition test, delayed recall of the remaining exemplars from these categories ($M = .32$, $SD = .18$) was better than recall of exemplars from categories that were not represented during the initial recognition test ($M = .24$, $SD = .08$). The results from this experiment are particularly noteworthy because Mandler and Rabinowitz used verbal materials similar to those typically used in retrieval-induced forgetting paradigms. Two methodological differences, however, set this experiment apart from the typical retrieval-induced forgetting experiments. First, Mandler and Rabinowitz used recognition as their initial test. Similar to the present experiments in which unique cues are linked to each to-be-recalled item, recognition should eliminate (or at least reduce) retrieval competition. Second, Mandler and Rabinowitz waited a week before administering their final recall test. As we have noted earlier, retrieval-induced inhibition, even if it had occurred, might not have survived the 7-day retention interval.⁴

Facilitation as a Retrieval-specific Mechanism

One crucial finding from Experiments 1 and 3 is the absence of facilitation for the

nontested material when participants reviewed the studied material. This finding implicates retrieval during the initial tests as a crucial component in the later facilitation of nontested material. The importance of the initial tests is further emphasized by the finding that retrieval-induced facilitation increases with response time on Day 2 but only for participants in the testing condition. That is, if retrieval-induced facilitation can be attributed solely to participants using recalled items as extra retrieval cues during the final test, then facilitation should have occurred in the slower responses in both the narrow and extra study conditions. This outcome, however, did not materialize (see Figure 8). The finding that no retrieval-induced facilitation occurred in the narrow and extra study conditions suggests that when there are no initial tests, much of the learned information is no longer accessible to retrieval after the 24-hour retention interval, regardless of response time during the final test. However, when there is an initial test, these would-be-inaccessible memories are enhanced so that they are recallable later (Chan & McDermott, 2006).

The dissociation between retrieval and repeated studying on retrieval-induced facilitation points to the interesting idea that a distinction might be drawn between associative activation based on encoding processes and retrieval processes. The fact that facilitation of the nontested materials is observed after a 24-hour delay is especially impressive when one considers the short-lived nature of semantic priming, which is a type of associative activation driven by encoding processes (Masson, 1995; Zeelenberg & Pecher, 2002). Our results suggest that conscious retrieval of associated thoughts is particularly likely during retrieval as opposed to during encoding (re-studying). Specifically, it is possible that a broad retrieval strategy can augment the automatic associative activation of related memories, thus making it more likely for long lasting

facilitative effects to occur for these memories. The idea that automatic processes can be modulated by controlled processes is not new (see Balota, Black, & Cheney, 1992; Chan, McDermott, Watson, & Gallo, 2005; Cunningham et al., 2004), and future research will surely shed more light on how these processes interact to produce and eliminate retrieval-induced facilitation.

Applied Implications

There is a recent surge of interest in applying well-established cognitive theories to the practice of education, as evidenced by numerous educationally-relevant symposia at psychology conferences (McDaniel, 2005; Roediger, 2005) and significant government grants (e.g., from the Institute of Education Sciences). Memory researchers may be able to help educators devise practices that maximize learning and retention. For example, experiments in the testing effect literature have shown time and again that retrieval practice enhances retention and slows forgetting (Hogan & Kintsch, 1971; Roediger & Karpicke, 2006a; but see also Slamecka & Katsaiti, 1988, for a counter argument). Participants in the current experiments showed hardly any forgetting on initially-recalled material even after a 24-hour retention interval. The beneficial effects of testing on later memory is a well-known phenomenon (Bjork, 1999; Roediger & Karpicke, 2006b); however, the current research demonstrates that even related, nontested material can be enhanced through testing.

The amount of information that textbooks and lectures present to students throughout the course of a semester is overwhelming. It is impractical, if not impossible, for teachers to provide exhaustive tests for the courses they teach. Therefore, teachers typically sample important and overarching concepts to be covered in exams, which begs the question of how well students can

maintain knowledge that is not tested in the exams. Results from the current experiments imply that as long as students have retrieved a concept, other related concepts should also receive a boost. In fact, according to the adjunct questions literature, this facilitative effect is particularly powerful when the initial tests involve higher-order questions (questions that require reasoning and manipulation of knowledge) and when questions are administered as short answers (or cued recall) rather than multiple choice questions (Hamaker, 1986). Moreover, if facilitation of related memories are indeed based on conscious strategies, then it should be particularly likely for such facilitation to occur when students are asked more overarching, conceptual questions, as opposed to questions that focus on minute details, which are particularly prominent in multiple choice exams. Our findings might be especially encouraging to teachers who regularly use essay or short answer questions in their exams.

Retrieval of concepts related to the target information occurs frequently in real-life testing situations, especially in essay exams. Teachers and teaching assistants (including the first author of the present manuscript) frequently complain about students “writing too much” in their exam booklets. That is, much to the chagrin of the graders, students often answer a simple short answer question with as much information as they can retrieve (perhaps in hopes of receiving partial credit or not to omit any information potentially critical to the scoring criteria of the instructor), regardless of whether the retrieved information is indeed the correct answer. In fact, retrieval of such related information might be even more likely to occur in real-life situations than in our experiments, because in real-life exams students are almost never penalized for guessing. One may even go so far as stating that real-life testing is analogous to a forced recall procedure in which students are encouraged to guess. Ironically, findings from our study suggest

that this sort of all-inclusive retrieval strategy, though perhaps annoying to the grader, might be beneficial to retention in the long run.

Another potentially important finding from the current experiments is that retrieval-induced facilitation increases with response time. This result highlights the importance of providing students with ample time during exams. Indeed, had we given participants 10, instead of 30, seconds to answer each question, we might have incorrectly concluded that retrieval-induced facilitation does not occur. The present results also have implications for students' studying strategies. For example, textbooks usually include end-of-chapter questions for review. However, it is unclear whether students actually use such questions when they are preparing for an exam. Our finding suggests that doing practice exams is a good way to improve long-term retention for both the directly recalled items and their associates.

Conclusions

Under conditions that simulate educational situations, we have demonstrated that retrieval practice not only benefits memory for the directly tested material, but also (to a lesser extent) its related, nontested counterparts. We have also shown that this finding is caused by mechanisms that are specific to the retrieval process. From an applied perspective, these results have highlighted the importance of giving students ample response time and of testing as opposed to restudying. From a theoretical perspective, these results suggest that theories may need to be revised to take into account the different facilitative effects caused by encoding and retrieval, and that inhibition of memories related to previously retrieved targets is not a necessary consequence of retrieval practice.

References

- Abbott, E. E. (1909). On the analysis of the factors of recall in the learning process. *Psychological Monographs, 11*, 159-177.
- Anderson, J. R. (1996). ACT: A simple theory of complex cognition. *American Psychologist, 51*, 355-365.
- Anderson, M. C. (2003). Rethinking interference theory: Executive control and the mechanisms of forgetting. *Journal of Memory & Language, 49*, 415-445.
- 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.
- Anderson, M. C., & Green, C. (2001). Suppressing unwanted memories by executive control. *Nature, 410*, 366-369.
- Anderson, M. C., Green, C., & McCulloch, K. C. (2000). Similarity and inhibition in long-term memory: Evidence for a two-factor theory. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 26*, 1141-1159.
- Anderson, M. C., & McCulloch, K. C. (1999). Integration as a general boundary condition on retrieval-induced forgetting. *Journal of Experimental Psychology: Learning, Memory, & Cognition, 25*, 608-629.
- Balota, D. A., Black, S. R., & Cheney, M. (1992). Automatic and attentional priming in young and older adults: Re-evaluation of the two-process model. *Journal of Experimental Psychology: Human Perception and Performance, 18*, 485-502.

- Bangert-Drowns, R. L., Kulik, J. A., & Kulik, C. C. (1991). Effects of frequent classroom testing. *Journal of Educational Research, 85*, 89-99.
- Bauml, K.-H., & Hartinger, A. (2002). On the role of item similarity in retrieval-induced forgetting. *Memory, 10*, 215-224.
- Bjork, R. A. (1975). Retrieval as a memory modifier: An interpretation of negative recency and related phenomenon. In R. L. Solso (Ed.), *Information processing and cognition: The Loyola symposium* (pp. 123-144). New York: John Wiley.
- Bjork, R. A. (1999). Assessing our own competence: Heuristics and illusions. In D. Gopher & A. Koriat (Eds.), *Attention and performance XVII: Cognitive regulation of performance: Interaction of theory and application* (pp. 435-459). Cambridge, MA: MIT Press.
- Bulevich, J. B., Roediger, H. L., Balota, D. A., & Butler, A. C. (2006). Failures to find suppression of episodic memories in the think/no-think paradigm. *Memory & Cognition, in press*.
- Butler, K. M., Williams, C. C., Zacks, R. T., & Maki, R. H. (2001). A limit on retrieval-induced forgetting. *Journal of Experimental Psychology: Learning, Memory, & Cognition, 27*, 1314-1319.
- Camp, G., Pecher, D., & Schmidt, H. G. (2005). Retrieval-induced forgetting in implicit memory tests: The role of test awareness. *Psychonomic Bulletin & Review, 12*, 490-494.
- Carrier, M., & Pashler, H. (1992). The influence of retrieval on retention. *Memory & Cognition, 20*, 633-642.
- Caughey, J. E., Boroumand, Y., Bjork, R. A., & Bjork, E. L. (2005). *The testing effect and retrieval-induced forgetting: Will testing suppress untested educational materials?* Paper

presented at the The 17th Annual Meeting of the American Psychological Society, Los Angeles, CA.

- Chan, J. C. K., & McDermott, K. B. (2006). Initial testing enhances later recollection in recognition memory: A dual-process perspective. *Manuscript in preparation*.
- Chan, J. C. K., McDermott, K. B., Watson, J. M., & Gallo, D. A. (2005). The importance of material-processing interactions in inducing false memories. *Memory & Cognition, 33*, 389-395.
- Collins, A. M., & Loftus, E. F. (1975). A spreading-activation theory of semantic processing. *Psychological Review, 82*, 407-428.
- Collins, A. M., & Quillian, M. R. (1972). Experiments on semantic memory and language comprehension. In L. W. Gregg (Ed.), *Cognition in learning and memory*. Oxford, England: John Wiley & Sons.
- Cunningham, W. A., Johnson, M. K., Raye, C. L., Gatenby, J. C., Gore, J. C., & Banaji, M. R. (2004). Separable neural components in the processing of black and white faces. *Psychological Science, 15*, 806-813.
- Dodd, M. D., Castel, A. D., & Roberts, K. E. (2006). A strategy disruption component to retrieval-induced forgetting. *Memory & Cognition, 34*, 102-111.
- Duchastel, P. C. (1981). Retention of prose following testing with different types of tests. *Contemporary Educational Psychology, 6*, 217-226.
- Erdelyi, M. H., & Becker, J. (1974). Hypermnesia for pictures: Incremental memory for pictures but not words in multiple recall trials. *Cognitive Psychology, 6*, 159-171.

- Foltz, P. W., Kintsch, W., & Landauer, T. K. (1998). The measurement of textual coherence with latent semantic analysis. *Discourse Processes, 25*, 285-307.
- Frase, L. T. (1968). Questions as aids to reading: Some research and theory. *American Educational Research Journal, 5*, 319-332.
- Gates, A. I. (1917). Recitation as a factor in memorizing. *Archives of Psychology, 6*, 104.
- Hamaker, C. (1986). The effects of adjunct questions on prose learning. *Review of Educational Research, 56*, 212-242.
- Hogan, R. M., & Kintsch, W. (1971). Differential effects of study and test trials on long-term recognition and recall. *Journal of Verbal Learning & Verbal Behavior, 10*, 562-567.
- Hunt, R. R., & McDaniel, M. A. (1993). The enigma of organization and distinctiveness. *Journal of Memory & Language, 32*, 421-445.
- Indig, S. (2005). Cramming Canadians: Two-thirds of university students start studying for exams no more than a week in advance. Retrieved August 9, 2005, from <http://www.kumon.com/pressroom/pressreleases/article-canada.asp?articlenum=17&language=Canada>
- Jefferies, E., Lambon Ralph, M. A., & Baddeley, A. D. (2004). Automatic and controlled processing in sentences recall: The role of long-term and working memory. *Journal of Memory & Language, 51*, 623-643.
- Keys, N. (1934). The influence on learning and retention of weekly as opposed to monthly tests. *Journal of Educational Psychology, 25*, 427-436.
- Koriat, A. (1993). How do we know that we know? The accessibility model of the feeling of knowing. *Psychological Review, 100*, 609-639.

- Kuo, T.-M., & Hirshman, E. (1996). Investigations of the testing effect. *American Journal of Psychology, 109*, 451-464.
- Landauer, T. K., Foltz, P. W., & Laham, D. (1998). An introduction to latent semantic analysis. *Discourse Processes, 25*, 259-284.
- LaPorte, R. E., & Voss, J. F. (1975). Retention of prose materials as a function of postacquisition testing. *Journal of Educational Psychology, 67*, 259-266.
- Leeming, F. C. (2002). The exam-a-day procedure improves performance in psychology classes. *Teaching of Psychology, 29*, 210-212.
- MacLeod, M. D., & Macrae, C. N. (2001). Gone but not forgotten: The transient nature of retrieval-induced forgetting. *Psychological Science, 12*(2), 148-152.
- Mandler, G., & Rabinowitz, J. C. (1981). Appearance and reality: Does a recognition test really improve subsequent recall and recognition? *Journal of Experimental Psychology: Human Learning & Memory, 7*, 79-90.
- Masson, M. E. J. (1995). A distributed memory model of semantic priming. *Journal of Experimental Psychology: Learning, Memory, & Cognition, 21*, 3-23.
- Mazzoni, G. (2006). *Intentional suppression of true and false memories*. Paper presented at the The 46th Annual Meeting of the Psychonomic Society, Toronto, ON, Canada.
- McDaniel, M. A. (2005). *Applying Cognition to Education*. Paper presented at the Symposium Conducted at the 46th Annual Meeting of the Psychonomic Society, Toronto, ON, Canada.
- McDaniel, M. A., & Masson, M. E. J. (1985). Altering memory representations through retrieval. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 11*, 371-385.

- McDermott, K. B., & Watson, J. M. (2001). The rise and fall of false recall: The impact of presentation duration. *Journal of Memory and Language*, *45*(1), 160-176.
- Michael, J. (1991). A behavioral perspective on college teaching. *Behavior Analyst*, *14*, 229-239.
- Modigliani, V. (1976). Effects on a later recall by delaying initial recall. *Journal of Experimental Psychology: Human Learning and Memory*, *2*, 609-622.
- Nungester, R. J., & Duchastel, P. C. (1982). Testing versus review: Effects on retention. *Journal of Educational Psychology*, *74*, 18-22.
- Raaijmakers, J. G. (2003). Spacing and repetition effects in human memory: Application of the SAM model. *Cognitive Science*, *27*, 431-452.
- Raaijmakers, J. G., & Shiffrin, R. M. (1981). Search of associative memory. *Psychological Review*, *88*, 93-134.
- Rickards, J. P. (1979). Adjunct postquestions in text: A critical review of methods and processes. *Review of Educational Research*, *49*, 181-196.
- Roediger, H. L. (1974). Inhibiting effects of recall. *Memory & Cognition*, *2*(2), 261-269.
- Roediger, H. L. (2005). *Bringing cognitive science into the classroom*. Paper presented at the Symposium Conducted at the 17th Annual Meeting of the American Psychological Society, Los Angeles, CA.
- Roediger, H. L., Balota, D. A., & Watson, J. M. (2001). Spreading activation and arousal of false memories. In H. L. Roediger, J. S. Nairne, I. Neath & A. M. Surprenant (Eds.), *The nature of remembering: Essays in honor of Robert G. Crowder* (pp. 95-115). Washington, DC: American Psychological Association.

- Roediger, H. L., & Karpicke, J. D. (2006a). Test-enhanced learning: Taking memory tests improves long-term retention. *Psychological Science, 17*, 249-255.
- Roediger, H. L., & Karpicke, J. D. (2006b). The power of testing memory: Basic research and implications for educational practice. *Perspectives on Psychological Science, in press*.
- Roediger, H. L., & McDermott, K. B. (1995). Creating false memories: Remembering words not presented in lists. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 21*(4), 803-814.
- Roediger, H. L., & Thorpe, L. A. (1978). The role of recall time in producing hypermnesia. *Memory & Cognition, 6*, 296-305.
- Rohm, R. A., Sparzo, F. J., & Carson, M. B. (1986). College student performance under repeated testing and cumulative testing conditions: Report on five studies. *Journal of Educational Research, 80*(2), 99-104.
- Runquist, W. N. (1983). Some effects of remembering on forgetting. *Memory & Cognition, 11*, 641-650.
- Salthouse, T. A. (1996). The processing-speed theory of adult age differences in cognition. *Psychological Review, 103*, 403-428.
- Salthouse, T. A., & Meinzig, E. J. (1995). Aging, inhibition, working memory, and speed. *Journal of Gerontology: Psychological Sciences & Social Sciences, 50B*, 297-306.
- Saunders, J., & MacLeod, M. D. (2002). New evidence on the suggestibility of memory: The role of retrieval-induced forgetting in misinformation effects. *Journal of Experimental Psychology: Applied, 8*, 127-142.

- Schwartz, B. L., & Smith, S. M. (1997). The retrieval of related information influences tip-of-the-tongue states. *Journal of Memory and Language*, *36*, 68-86.
- Slamecka, N. J., & Katsaiti, L. T. (1988). Normal forgetting of verbal lists as a function of prior testing. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *14*, 716-727.
- Smith, R. E., & Hunt, R. R. (2000). The influence of distinctive processing on retrieval-induced forgetting. *Memory & Cognition*, *28*, 503-508.
- Spitzer, H. F. (1939). Studies in retention. *Journal of Educational Psychology*, *30*, 641-656.
- Tulving, E., & Watkins, M. J. (1974). On negative transfer: Effects of testing one list on the recall of another. *Journal of Verbal Learning & Verbal Behavior*, *13*, 181-193.
- Wixted, J. T., & Rohrer, D. (1994). Analyzing the dynamics of free recall: An integrative review of the empirical literature. *Psychonomic Bulletin & Review*, *1*, 89-106.
- Zeelenberg, R., & Pecher, D. (2002). False memories and lexical decision: Even twelve primes do not cause long-term semantic priming. *Acta Psychologica*, *109*, 269-284.

Author Notes

The experiments in this report were supported by a grant from the Institute of Education Sciences (R305H030339). Some of the findings from this research report have been presented as a poster at the 2005 APS convention in Los Angeles, and others were presented as a poster at the 2005 meeting of the Psychonomic Society. We thank Pooja Agarwal, Karl Szpunar, Sean Kang, Jeff Karpicke, Andrew Butler, and Jes Logan for helpful comments and discussions.

Correspondence of this manuscript should be addressed to J.C.K. Chan (email: ckchan@wustl.edu).

Footnotes

¹Because presentation order of the study articles was random, one possible concern was that participants who read the control article before the experimental article would be impaired in learning the experimental article due to negative transfer, also known as the priority effect (Tulving & Watkins, 1974). For example, when participants learn A-D following the learning of A-B, the learning of A-D is impaired if no test is given following the learning of A-B (and prior to the learning of A-D). Such concern is probably not applicable to the current experiment because the priority effect has only been shown in the classic paired associates A-B, A-D (but not A-B, C-D) paradigm. The priority effect has never been shown with prose materials or when the two sets of to-be-remembered materials are heterogeneous like the articles used here.

²Caughey and colleagues (2005) showed that cross-article output interference may mask the effect of retrieval-induced facilitation. Specifically, if the control article is tested prior to the experimental article, output interference from the control article (control items) may impair recall for the experimental article (the nontested-related and tested items), thus creating effects that ostensibly represent retrieval-induced forgetting. However, no cross-article output interference was evident in our results. The data for participants who were tested on the experimental article first were virtually identical to the data for participants who were tested on the control article first. Moreover, for participants in the control condition (in Experiment 1 and 2) and participants in the extra-study condition in Experiment 1, we conducted a split-half analysis to examine whether recall probabilities for questions presented in the first half of the experiment (on Day 2) differ from those presented in the second half. Once again, virtually no

output interference was observed in any of the three conditions (all differences within 1.5 percentage points).

³The broad/narrow manipulation may remind one of the think/no-think procedure (Anderson & Green, 2001) because participants were told not to think about facts related to the target memory. Although there is some similarity between these procedures, there are also some importance differences. For example, in the think/no think procedure, participants are told to not think about the *target* response when given the cue. However, in our experiment, participants were told explicitly to retrieve the target to their best ability. Moreover, no inhibition was found in our narrow retrieval condition, which is ostensibly similar to the no-think condition (see also, Bulevich, Roediger, Balota, & Butler, 2006; Mazzoni, 2006).

⁴One may wonder whether Mandler and Rabinowitz's (1981) free recall procedure presents a condition in which retrieval-induced forgetting is masked by enhanced category access for the testing condition over the control condition (Anderson, 2003). This concern is likely unwarranted because Mandler and Rabinowitz's results were actually a combination of free and cued recall data. Specifically, participants were provided with the category names as retrieval cues if they were unable to recall any items in a given category, and the reported results have taken into account items that were recalled *after* subjects were cued. That is, items were scored as recalled if they were recalled in either the free or cued recall phases.

Table 1

Recall Probability as a Function of Practice Condition and Question Type in Experiment 1 (SD in parentheses)

	<u>Day 1</u>		<u>Day 2</u>	
	<u>Test 1</u>	<u>Test 2</u>	<u>Presented on Day 1</u>	<u>Not Presented on Day 1</u>
<u>Testing</u>				
Correct	.69 (.13)	.72 (.13)	.71 (.14)	.59 (.15)
Incorrect	.26 (.12)	.25 (.13)	.23 (.13)	.30 (.10)
Unanswered	.05 (.06)	.04 (.05)	.06 (.07)	.11 (.11)
<u>Extra study</u>				
Correct			.79 (.14)	.49 (.14)
Incorrect			.15 (.10)	.39 (.12)
Unanswered			.06 (.08)	.12 (.14)
<u>Control</u>				
Correct				.50 (.17)
Incorrect				.34 (.17)
Unanswered				.17 (.16)

Table 2

Recall Probability as a Function of Question Type in Experiment 2 (SD in parentheses)

	<u>Day 1</u>		<u>Day 2</u>		
	<u>Test 1</u>	<u>Test 2</u>	<u>Tested</u>	<u>Nontested-Related</u>	<u>Control</u>
<u>Testing</u>					
Correct	.65 (.18)	.67 (.18)	.66 (.18)	.49 (.20)	.42 (.16)
Incorrect	.31 (.17)	.30 (.17)	.29 (.17)	.43 (.20)	.49 (.19)
Unanswered	.04 (.09)	.03 (.07)	.05 (.11)	.08 (.15)	.09 (.16)

Table 3

Recall Probability as a Function of Initial Practice Condition and Question Type in Experiment 3 (SD in parentheses)

	<u>Day 1</u>		<u>Day 2</u>		
	<u>Test 1</u>	<u>Test 2</u>	<u>Tested /Restudied</u>	<u>Nontested-Related</u>	<u>Control</u>
<u>Broad</u>					
Correct	.63 (.23)	.66 (.22)	.66 (.23)	.51 (.19)	.41 (.18)
Incorrect	.21 (.15)	.21 (.13)	.18 (.13)	.25 (.14)	.32 (.15)
Unanswered	.16 (.16)	.13 (.13)	.16 (.18)	.23 (.14)	.27 (.19)
<u>Narrow</u>					
Correct	.61 (.21)	.65 (.20)	.67 (.19)	.44 (.16)	.44 (.13)
Incorrect	.26 (.16)	.26 (.16)	.24 (.14)	.38 (.15)	.39 (.16)
Unanswered	.13 (.13)	.09 (.14)	.09 (.11)	.19 (.14)	.17 (.12)
<u>Extra Study</u>					
Correct			.65 (.19)	.38 (.24)	.42 (.18)
Incorrect			.24 (.13)	.39 (.24)	.37 (.10)
Unanswered			.11 (.11)	.19 (.14)	.22 (.13)

Figure Captions

Figure 1. Mean probabilities of accurate recall on Day 2 as a function of question type (presented on Day 1 or not) and practice condition (testing, extra study, control) in Experiment 1. Participants in the testing condition never received corrective feedback. In retrieval-induced forgetting terms, the first bar from the left is Rp+, the third bar is Rp-, and the fifth bar is Nrp. Error bars display between-subjects .95 CI.

Figure 2. Day 2 cumulative probabilities of accurate recall for questions not presented on Day 1 as a function of response time and practice condition (testing, extra study, control).

Figure 3. Mean probabilities of accurate recall on Day 2 as a function of question type (tested, nontested-related, and control) in Experiment 2. In retrieval-induced forgetting terms, the first bar from the left is Rp+, the second bar is Rp-, and the third bar is Nrp. Error bars display within-subjects .95 CI.

Figure 4. Day 2 cumulative probabilities of accurate recall for initially nontested questions as a function of response time and question type (nontested-related, control).

Figure 5. Mean probabilities of accurate recall on Day 2 as a function of question type (tested, nontested-related, control) and fast vs. slow participants based on their mean correct response speed on Day 1. Error bars display within subjects .95 CI.

Figure 6. Mean probabilities of accurate recall on Day 2 as a function of question type (tested, nontested-related, and control) and initial practice condition (broad, narrow, extra study) in Experiment 3. Error bars display .95 CI.

Figure 7. Separate plots illustrating Day 2 cumulative probabilities of accurate recall for initially nontested questions as a function of response time, question type, and initial practice condition.

Figure 8. Mean probabilities of accurate recall on Day 2 as a function of question type (nontested-related, control) and fast vs. slow participants based on their mean correct response speed on Day 1 in the broad retrieval condition. Top panel shows results for participants in the broad condition and bottom panel shows results for participants in the narrow condition. Error bars display within subjects *.95 CI*.

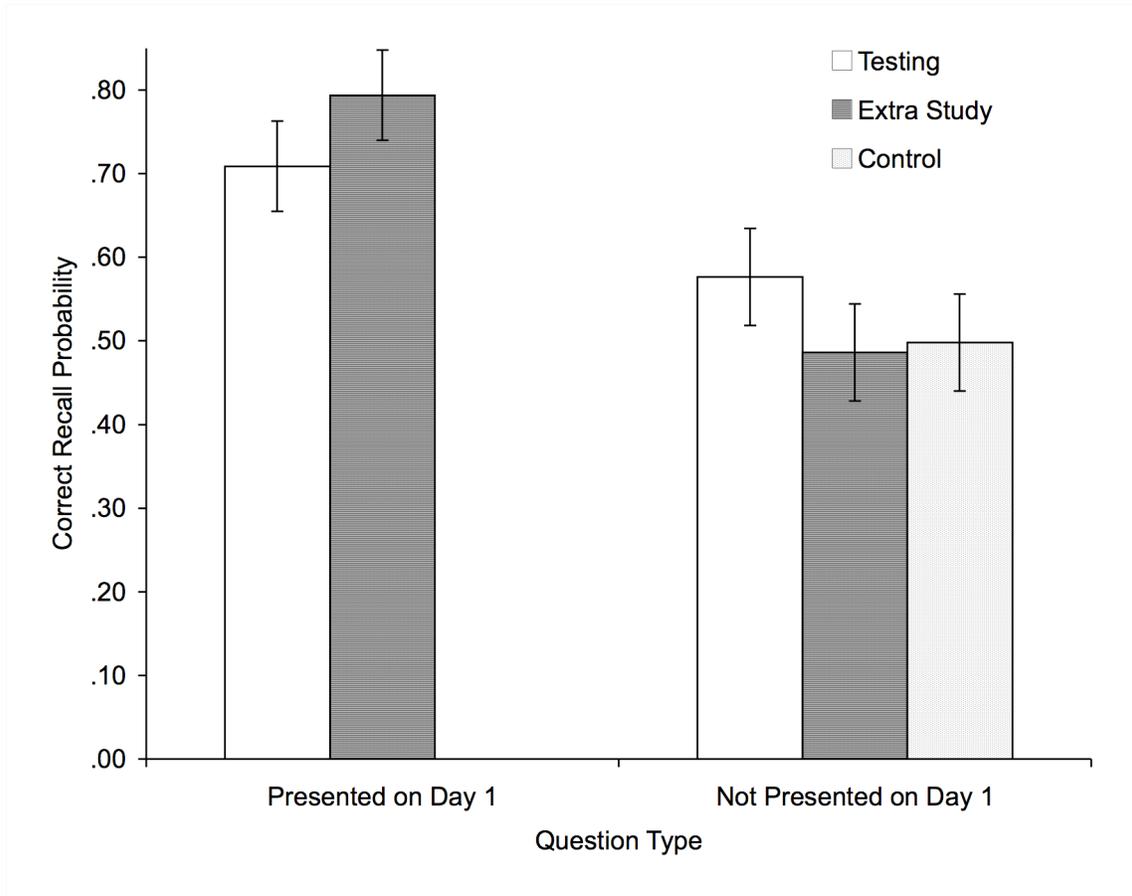
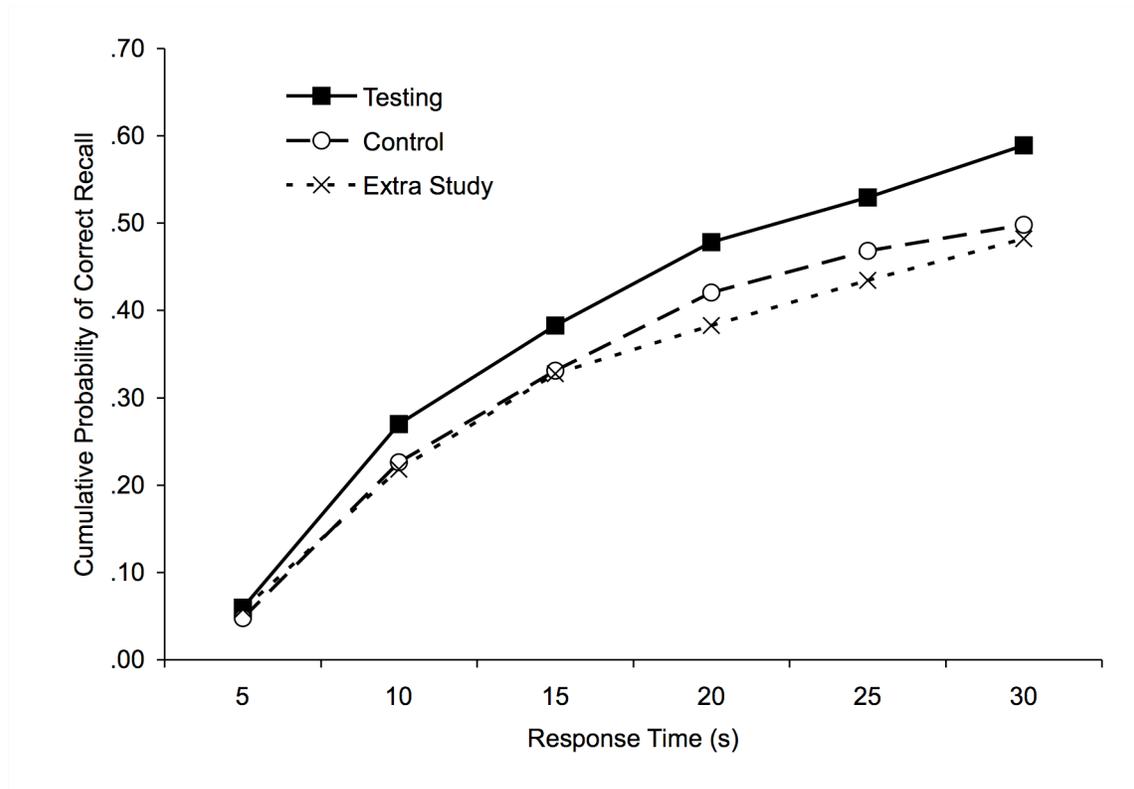


Figure 1



Note that response time and probability of correct recall does not start at time = 0 because “time 0” was the onset of the question presentation, and reading the questions alone should take a few seconds after stimulus onset.

Figure 2

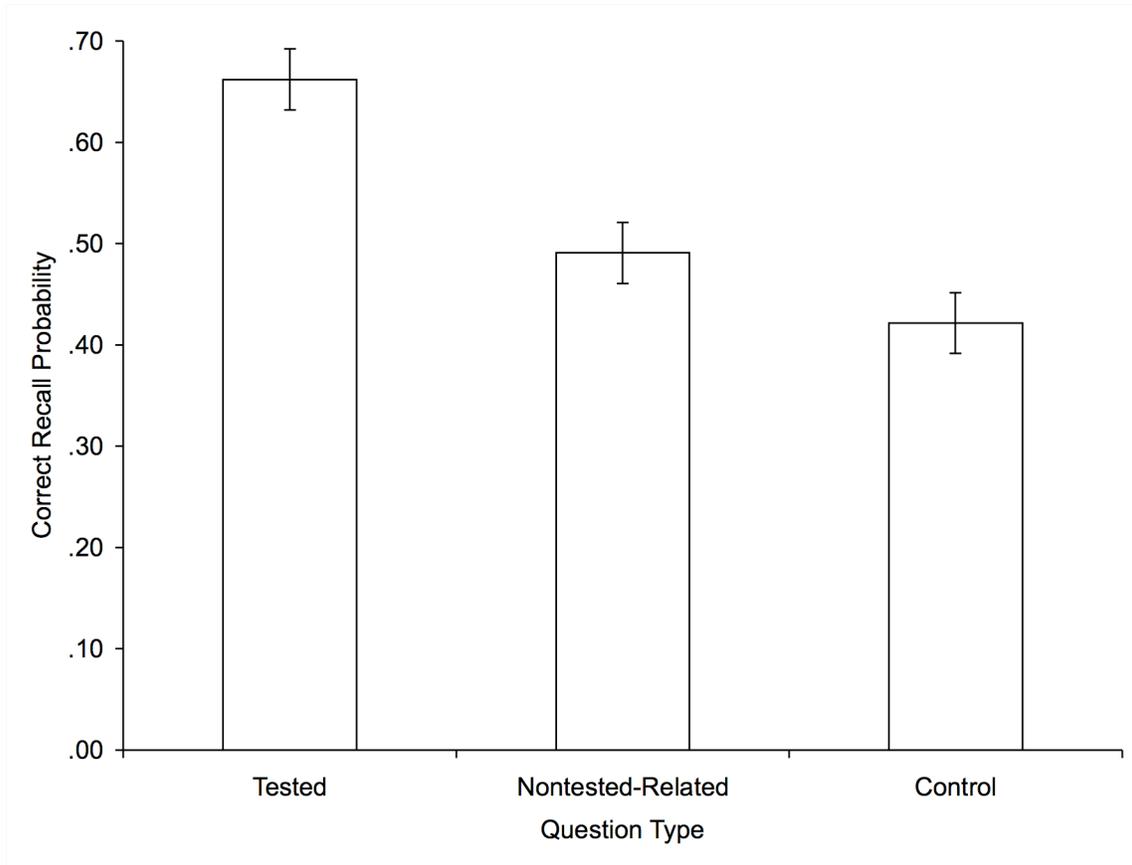


Figure 3

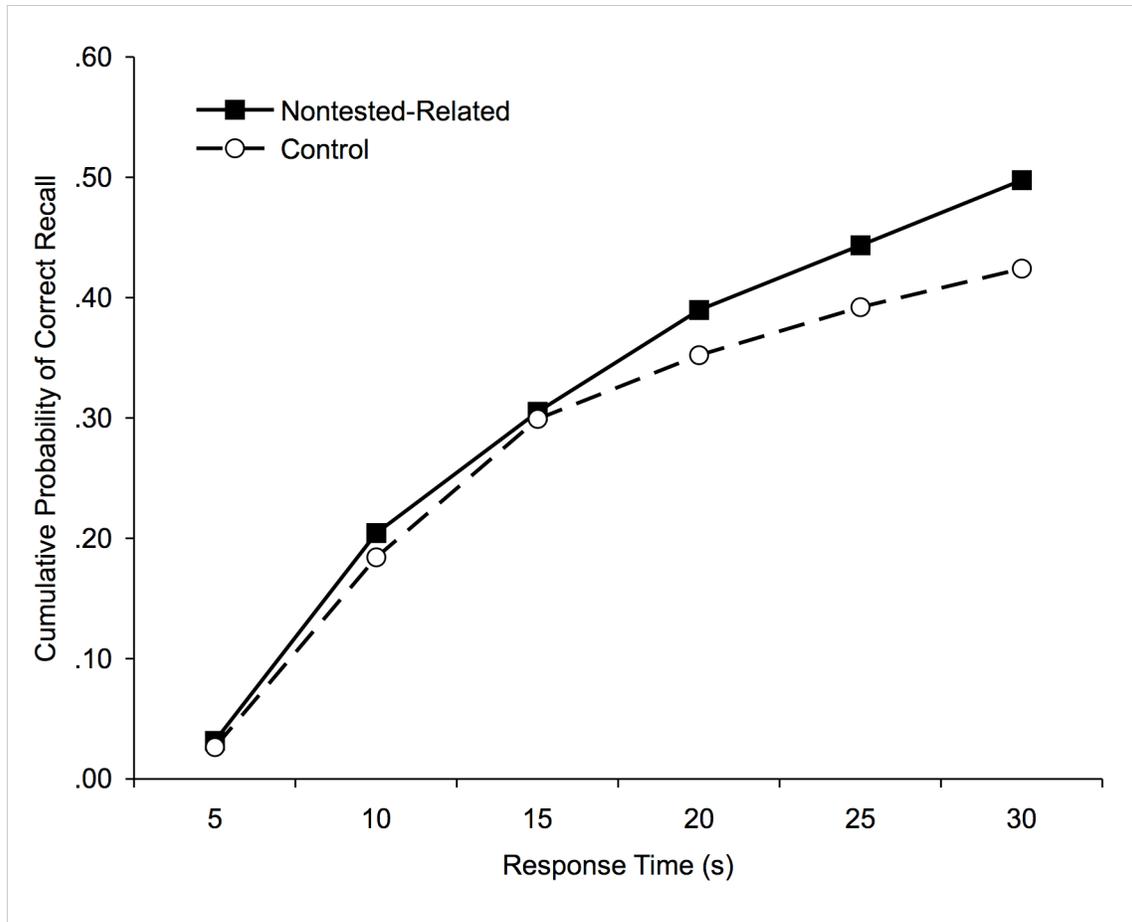


Figure 4

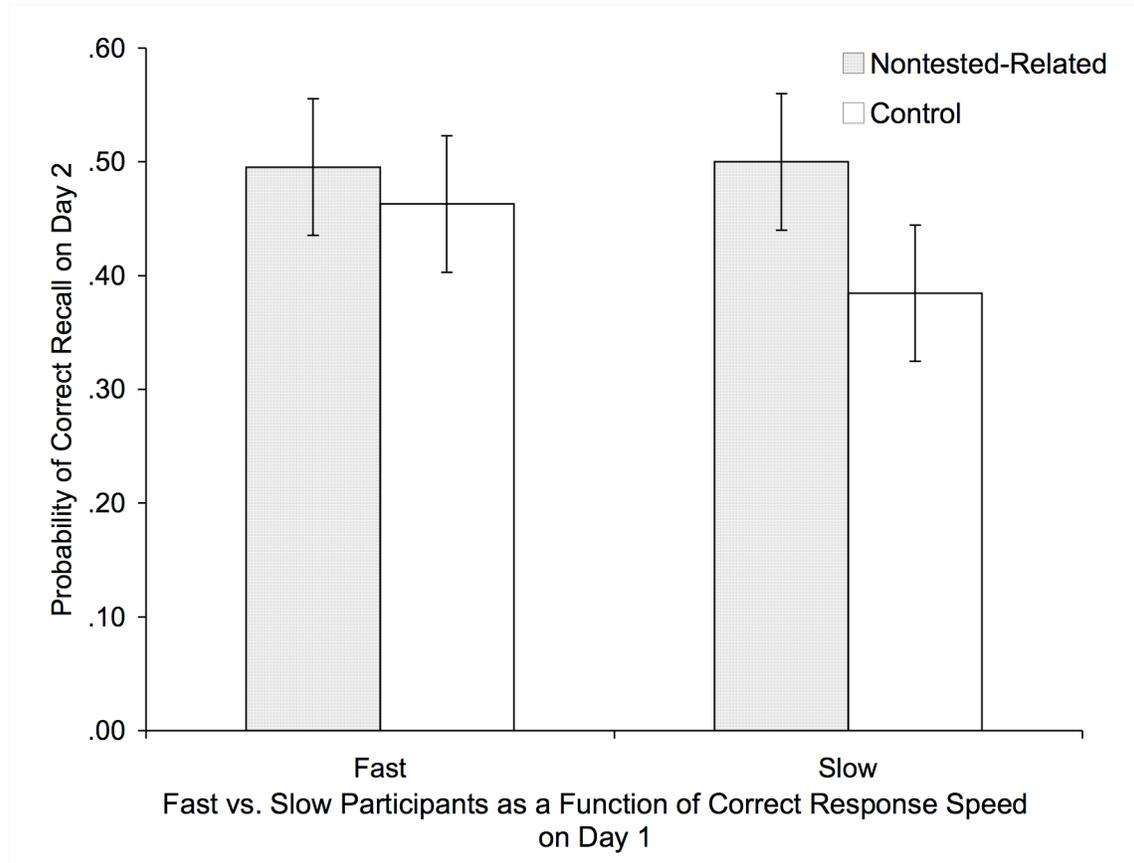


Figure 5

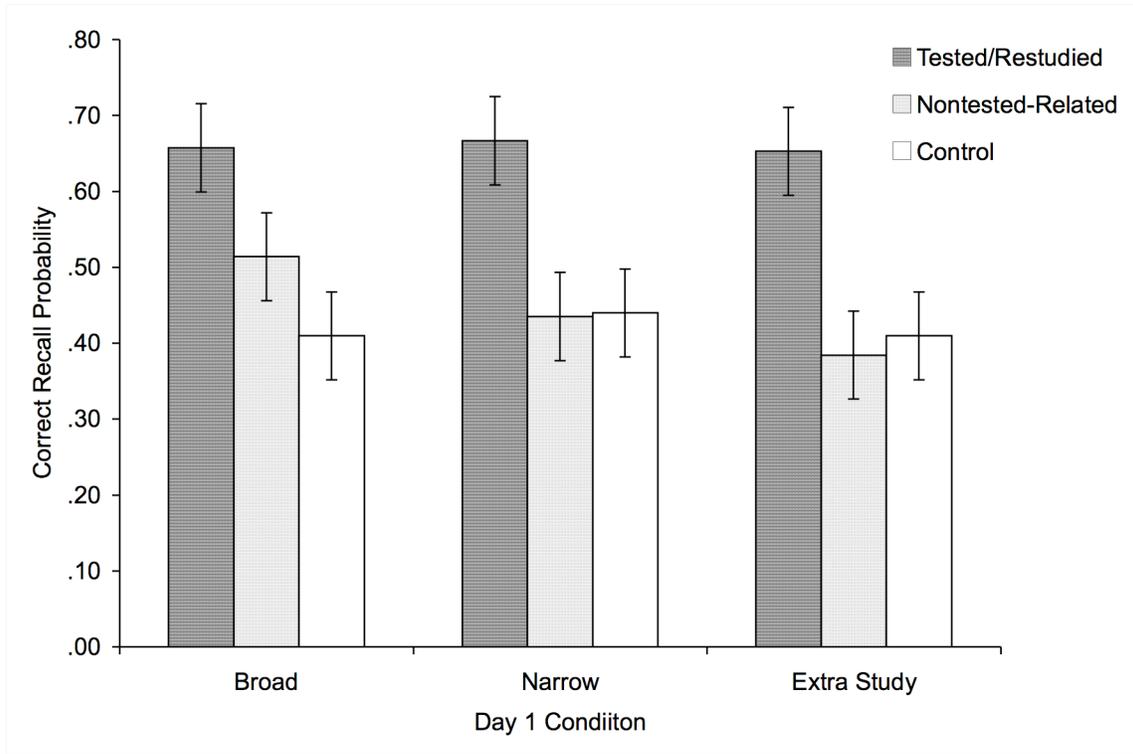


Figure 6

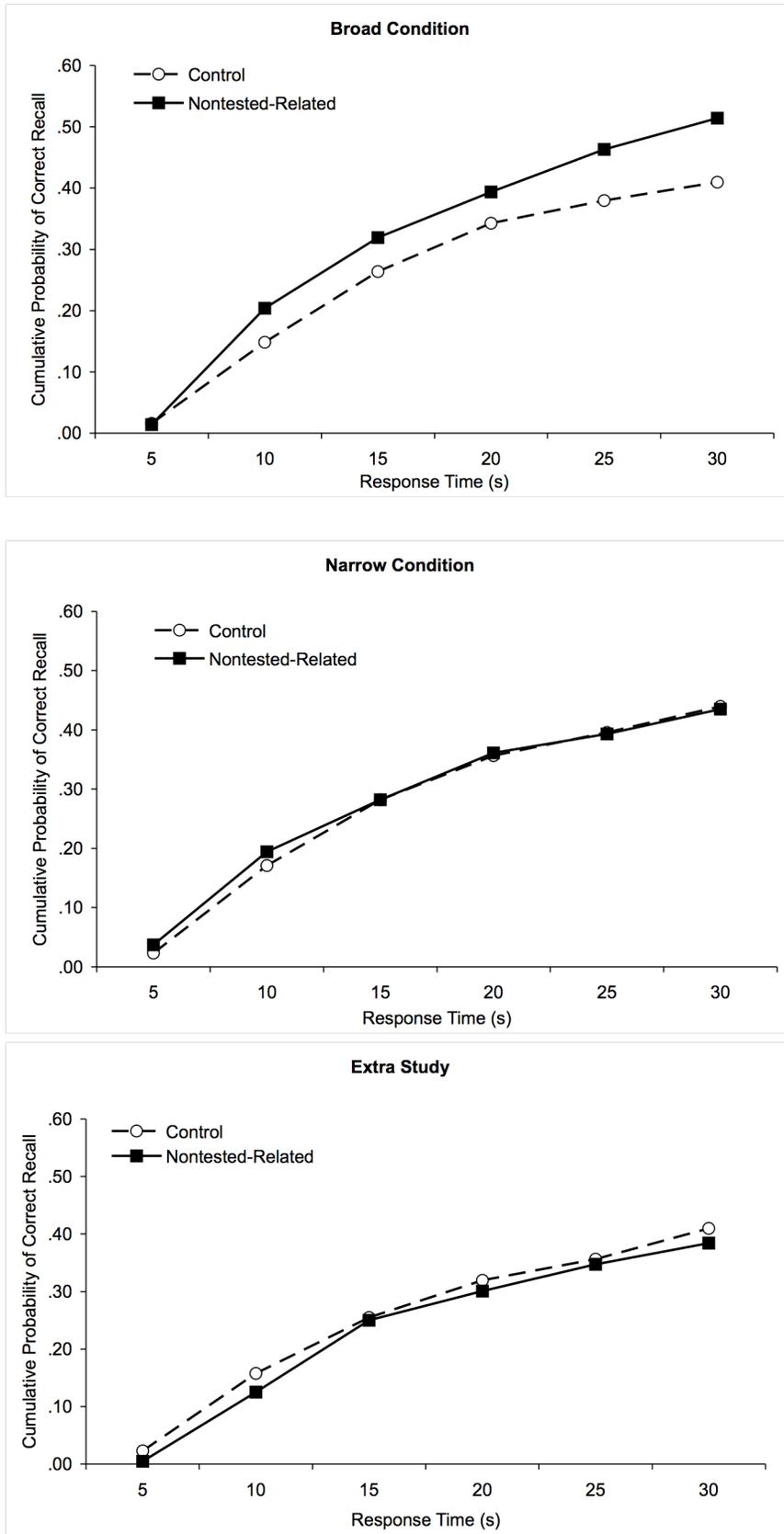


Figure 7

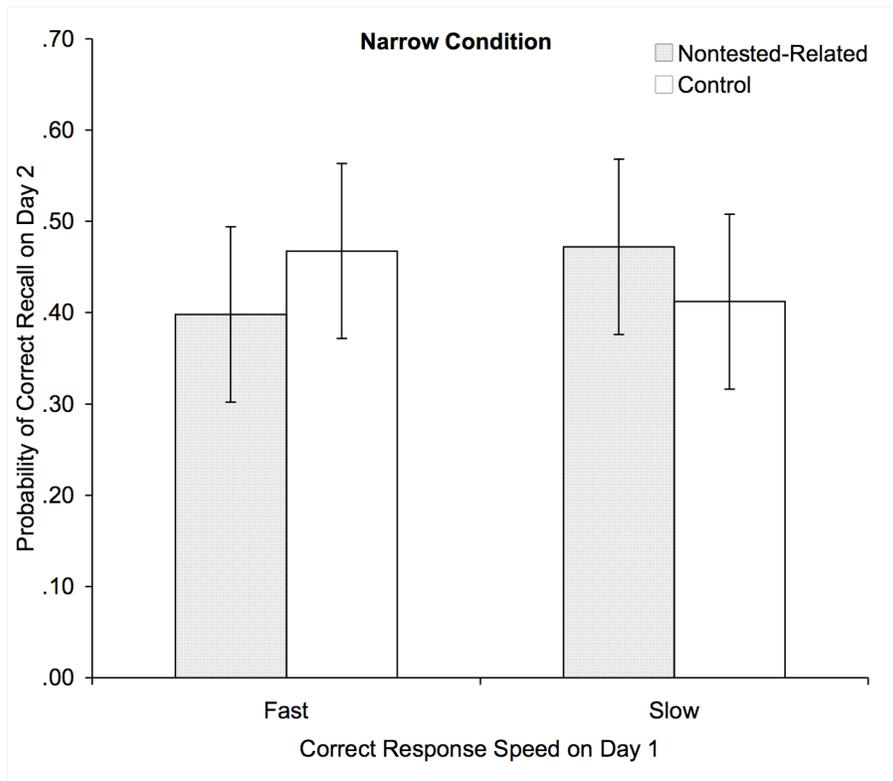
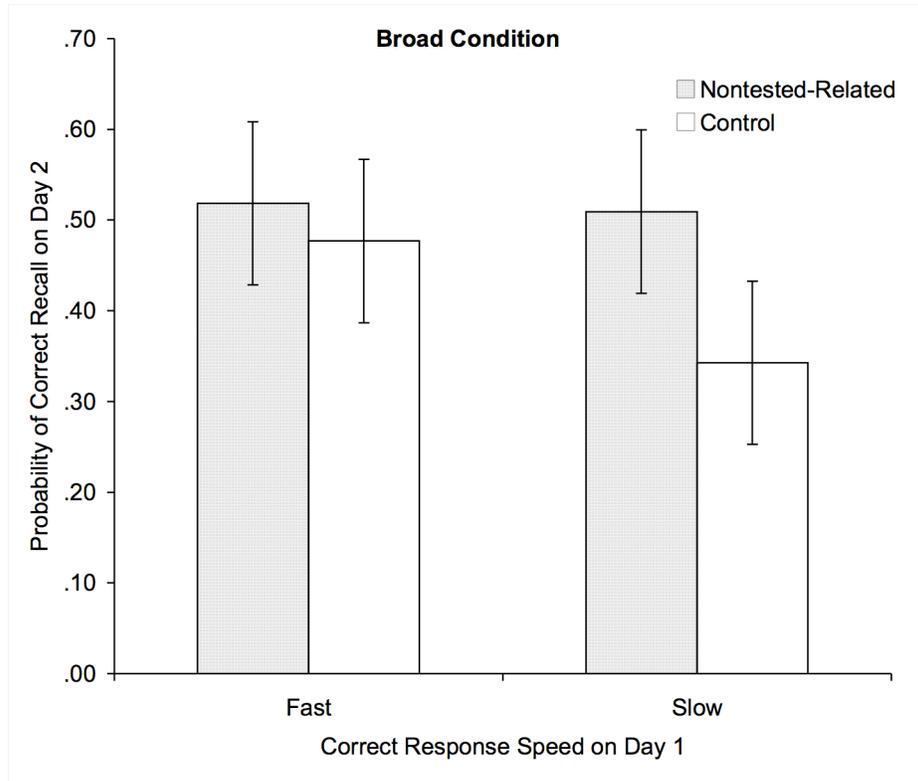


Figure 8

Appendix A

Online sources for the articles

Toucan

<http://www.summersbirds.com/toucans/toucans.htm>

<http://www.petpublishing.com/birdtimes/breeds/toucan.shtml>

<http://birding.about.com/od/birdstoucans/a/toucans.htm>

Big Bang Theory

<http://www.umich.edu/~gs265/bigbang.htm>

http://www.damtp.cam.ac.uk/user/gr/public/bb_home.html

http://cosmology.berkeley.edu/Education/IUP/Big_Bang_Primer.html

<http://www.leaderu.com/real/ri9404/bigbang.html>

Shaolin Temple

<http://shaolin-temple.gungfu.com/>

<http://www.shaolin-overseas.org/history.html>

<http://www.shaolin-overseas.org/buddhism.html>

<http://www.chinavoc.com/kungfu/shaolin/intro.asp>

<http://www.shaolintemple.org/buddhism.htm>

http://www.shaolin.com/shaolin_history.aspx

http://www.shaolin.com/shaolin_philosophy.aspx

History of Hong Kong

http://www.lonelyplanet.com/destinations/north_east_asia/hong_kong/history.htm

http://en.wikipedia.org/wiki/History_of_Hong_Kong

<http://www.askasia.org/frclasrm/readings/r000206.htm>

Appendix B

The following is an excerpt from the toucan article used in Experiment 1:

Everyone who has seen a Toucan has surely noticed its huge bill. It is assumed that there are approximately 40 Toucan species, all of them with more or less oversized bills. Scientifically the Toucan family is called *Ramphastidae*. Toucans get their name from 'tucano' given to them by the Tupi Indians of Brazil. The largest species, the Toco toucan, is about 25 inches (64 cm) long. The smallest toucan, the Aracar toucanet, is only about 14 inches (36 cm) long. While different species of toucans vary somewhat in longevity, the average lifespan of a well cared for toucan is at least ten years, and they can live to be approximately 15 years old. The toucans are poor fliers because of the size of their bill, and they rely on hopping from branch to branch in trees.

It's striking that the relation between the size of the Toucan's bill and the size of its body increases with the size of the species. Small Toucan species have rather "normal" bills, whereas the bills of large species like the Toco toucan (*Ramphastos toco*) are huge, even if the size of the body is taken into consideration. The feathers of most Toucan species is very conspicuous, too. Black is often the dominant color, but it's interrupted by colorful contrasting areas. Keel-billed toucans have a bright yellow chest, for instance. The area around the eyes is without feathers but is still very colorful in most species. Paradoxically the colorful feathers and the even more colorful bill are a perfect camouflage in the treetops, because when viewed from a distance they make Toucans look more like fruits than birds.

Toco toucans, a.k.a. *Ramphastos toco*, have the largest bill of any toucan. The toco toucan is mainly black, with white on the throat and upper breast. The bill is orange crimson, fading to greenish yellow. There's a large black oval blotch near the tip of bill, and a narrow black line at the base. The bills of the toucans are much lighter in weight than they appear. These large, boldly colored bills give them a distinctly out-of-balance appearance. The insides of their bills are shaped like honeycombs. It is hollow, made of protein keratin with thin rods of bone to support it - similar in consistency to a hard sponge. A thin outer sheath encloses a hollow which is crisscrossed by many thin, bony, supporting rods. Its tongue is like a feather which is used to catch food and flick it down its throat. It reaches to the tip of the bill.

The toucan's large bill can be used for many purposes. It's especially useful when foraging. Their bills enable them to perch inside the crown of a tree, where branches are thicker, and reach far outwards to pluck berries or seed from twigs too thin to bear their weight. Seized in the tip of the bill, food is thrown back into the throat by an upward toss of the head. When the fruits are too large, toucans use their bills (which have serrated edges) to tear these fruits into smaller pieces. Fruits dominate their diet, but Toucans are not pure vegetarians; they also hunt on insects and small reptiles - another field of activity for the large bill. The bill also plays an important part concerning communication, especially during courtship. In part thanks to their bill Toucans are well-fortified birds that are able to defend their young against some predators. The varied patterns of toucans' bills may also help these birds to recognize each other or attract a mate. During their nuptial display, both partners play a game in which they throw berries to each other or toss them against each other with their bills.

But there is one thing they can't use the large bill for: the construction of tree holes. Toucans need such holes for nesting and raising their young. Although they are closely related to woodpeckers, they are not able to construct such holes. Toucans depend on natural tree holes or holes constructed by other animals. When toucans sleep, they turn their head so that their long bill rests on their back and their tail is folded over their head. The bird becomes a ball of feathers. Often found in abandoned tree hollows or old woodpecker holes, five or six adults may sleep in one hole!

The following is a sample of the questions used in Experiment 1 for the toucan article. Answers are underlined.

Questions in Set A

The largest toucan species is the Toco toucan.

The feather of most toucan species is dominant by the color black.

The inside of a toucan's bill is shaped like honeycombs.

When a toucan swallows its food, it uses its tongue to flick the food down its throat.

What is the usage of a toucan's bill during nuptial display? To toss barriers between partners.

What other bird species is the toucan related to? Woodpeckers

Questions in Set B

The most colorful toucan species is the Keel Billed toucan.

Toucans are covered entirely by feathers except the bill and around the eyes.

The consistency of the inside of a toucan's bill is similar to a hard sponge.

The texture of a toucan's tongue is like a feather.

What is the usage of a toucan's bill when it encounters fruits that are too large to swallow? A toucan tears the fruits into smaller pieces with their bill.

Where do toucans sleep? Tree holes.