The Role of Mediators in Memory Retrieval as a Function of Practice: Controlled Mediation to Direct Access

Robert J. Crutcher  
Metropolitan State College of Denver

K. Anders Ericsson  
Florida State University

The role of prior knowledge in retrieval of Spanish–English vocabulary pairs learned using keyword mediators was examined in 4 experiments. Retrieval was tested immediately after learning and after 1-week and 1-month no-practice intervals (Experiment 1), after moderate retrieval practice (Experiment 2), and after extended retrieval practice (Experiments 3 and 4). Using accuracy, latency, and verbal report data, a detailed account of memory retrieval processes was developed. Initial retrieval is an explicit mediation process that involves retrieving keyword mediators into working memory and using them as retrieval cues to access the English equivalents of the Spanish words. After extended vocabulary retrieval practice, this sequential mediation process qualitatively changed to a direct retrieval process in which the English equivalent was accessed in a single working memory step. However, direct retrieval was still influenced by a covert mediation process.

An important issue in cognitive research is the role of prior knowledge and experience in cognitive processes. Evidence from a number of different research domains, including the study of expertise (Chase & Simon, 1973; Ericsson & Lehmann, 1996), exceptional memory (Chase & Ericsson, 1982; Ericsson & Kintsch, 1995; Wilding & Valentine, 1997), and mnemonics (Bellezza, 1981; Ericsson, 1985; Pressley, Levin, & Delaney, 1982), suggests that differences in prior knowledge and experience are associated with rather large differences in memory performance. However, the specific role of prior knowledge in mediating improved learning and memory processes is still not well understood. An important question, and the concern of the research reported here, involves determining the role played by the knowledge used initially to encode new material in subsequent memory retrieval processes and how that role changes as a function of repeated access of the memory system. We briefly review previous research on the role of mediation in memory and identify two general perspectives regarding the role of mediation in memory. We then describe a new process-tracing method for studying mediation issues and outline four experiments designed to examine the role of keyword mediators in memory retrieval processes of foreign–English vocabulary following different amounts of practice.

**Background**

Although he acknowledged the importance of prior knowledge in influencing memory processes in everyday life, Ebbinghaus (1885/1964) attempted to eliminate the influence of knowledge in studying basic associative memory processes by using nonsense syllables and fast presentation rates of material. Other researchers challenged Ebbinghaus's claim that only basic association processes were involved (e.g., Muller & Pilzecker, 1900; Muller & Schumann, 1894) and provided evidence for the influence of prior knowledge on memory storage and retrieval processes. Reed (1918a, 1918b) reported that memory for unrelated paired associates was much improved if the participants could access associative aids instead of relying on rote memorization. A large body of subsequent research has shown that when participants report having used meaningful mediators at encoding, retention of the items is improved (see Montague, 1972, and Richardson, 1998, for reviews). In addition, many studies have shown that instructing participants to use specific memory strategies improves recall relative to using rote rehearsal. However, instructing participants to use mnemonic strategies does not always increase performance compared to that of controls, because most adults appear to spontaneously use mnemonic strategies for memorization of arbitrary lists of words (e.g., Bower, 1972). In summary, there is general agreement that the use of mediators is related to ease of learning and improved retention in learning unfamiliar material (for reviews, see Montague, 1972; Prytulak, 1971; and Richardson, 1998).
An important issue, however, is the specific role mediators play in subsequent memory retrieval processes. As some researchers have argued, the use of mediators might influence initial learning processes without becoming a functional part of the memory representation or retrieval process, for example, by simply inducing participants to study the material more actively (Bellezza & Poplawsky, 1974; Bellezza, Poplawsky, & Aronovsky, 1977). The critical issue is whether or not the specific mediators used in learning unfamiliar material are subsequently involved in retrieval of this material from memory.

From earlier research, two general accounts of the role of mediators in memory retrieval processes can be identified. According to the association hypothesis (Adams & McIntyre, 1967; Bellezza et al., 1977; Underwood, 1972), or what we call the direct hypothesis, the mediators that individuals access during encoding do not become part of the functional memory trace that relates the items. Thus, learning an A–B pair using a mediator (e.g., the German–English pair Fleisch–meat using the mediator flesh) involves learning a simple association between A and B. In learning the pair, the mediator is simply a context for learning the A–B association but is not part of the memory representation relating A and B or the memory retrieval process that recovers B from A. The individual does not use the mediator flesh to retrieve meat from Fleisch.

According to an alternative mediation hypothesis (Adams & McIntyre, 1967; Bellezza, 1986; Bellezza & Poplawsky, 1974; Bellezza et al., 1977), mediators play an important functional role in the memory representation connecting A and B. Therefore, to retrieve B from A, the mediator must be accessed and then used to retrieve B. If the mediator cannot be accessed from A or if B cannot be accessed from the mediator, then it should not be possible to retrieve the A–B pair (Bellezza & Poplawsky, 1974; Bellezza et al., 1977; Montague, 1972; Montague, Adams, & Kiess, 1966). Retrieval can break down on either of the links, A–mediator or mediator–B.

To date, the strongest evidence for the mediation account is the correlation between recall of target material and recall of mediators (Bellezza & Poplawsky, 1974; Montague, 1972; Richardson, 1998). Higher recall of target material when the associated mediators are reported has been interpreted as evidence that mediators are functionally important in retrieval: In other words, the mediators are accessed first and then they are used to access the target material. However, because of concerns with the reactive influence of requesting verbal reports about mediators, this information is generally collected after recalling all of the items. Consequently, the observed correlation between recall of target material and mediators could be explained in alternative ways (e.g., see Richardson, 1998). First, participants may be biased to explain or report mediation for only those items they actually recall. Second, other shared factors may be responsible for the correlation; for example, the target material and mediators are studied together under instructions to use the specific mnemonic technique. Third, it is entirely possible that rather than the mediators cuing recall of the target material, it is the target material that cues retrieval of the mediators. In other words, the target B might be directly retrieved from A and then be used to retrieve the mediator M. This latter possibility would support the direct model, not the mediation model. To discriminate among these various possibilities, more detailed information about the sequence and time course of retrieval processes is needed. A method is needed that provides experimental control over encoding and a way to independently assess the strengths of the associations between A and M, between M and B, and between A and B.

Although the mediation hypotheses appear to best explain initial learning and retention, no studies have compared the mediation and direct accounts over time as material is repeatedly accessed from memory. Even if memory processes are initially mediated, the role of mediators may change over time. Everyday experience suggests, for example, that much of what we retrieve from memory is accessed directly from long-term memory without any intermediate working memory steps. Direct access is most parsimoniously accounted for by a direct model in which there is only a simple associative link between A and B, as proposed by the association hypothesis. In fact, it is often assumed that mediators serve only as a temporary bridge to maintain information in memory until a direct link between A and B is formed and strengthened (e.g., Atkinson & Raugh, 1975).

Alternatively, the mediation model could explain direct access by assuming that the original mediators remain integral to the memory trace and the retrieval process, with only the observable role of the mediators changing. Retrieval of material after initial study would require deliberately accessing mediators into working memory to use as retrieval cues to access the target material, whereas after extended practice the same mediators would be implicitly or covertly involved in the retrieval process (e.g., by semantic activation in long-term memory).

In the research reported here we examined the structure and duration of memory processes after initial learning as well as after varying amounts of retrieval practice to evaluate the role of mediators in memory retrieval.

Current Investigation

The first step in the current investigation was to select a naturally occurring task that permitted the study of mediation processes with a high degree of experimental control over encoding and retrieval. After specifying direct and mediated retrieval models from the association and mediation hypotheses, we developed a set of convergent methods to measure the duration and structure of retrieval processes to discriminate the two models. Finally, we designed a series of studies to examine the role of mediation in retrieval after initial learning and after retrieval practice.

The Task: Learning Vocabulary Using the Keyword Method

The task selected was learning Spanish–English vocabulary pairs using the keyword method (Atkinson & Raugh, 1975; Raugh & Atkinson, 1975), a two-step mnemonic technique for learning foreign vocabulary items and their English equivalents. The first step requires relating the foreign word to a keyword, a concrete English word that sounds or looks like the foreign word or to a salient part of the foreign word but is unrelated to its meaning. The second step requires relating the keyword and the English equivalent by forming an interactive visual image between the referents of the two words. For example, the Spanish–English pair perro–dog may be learned using the keyword pear by noticing that pear sounds like perro and then imagining a dog eating a pear.
We selected this task for a number of reasons. First, foreign vocabulary learning is a naturally occurring activity with which most adults have experience. Second, evidence from our own studies (Crutcher, 1993), as well as those of other researchers (Hall, Wilson, & Patterson, 1981; Lawson & Hogben, 1996; McDaniel & Tilman, 1987; Pressley, Levin, Digdon, Bryant, & Ray, 1983; Pressley, Levin, Kuiper, Bryant, & Michiner, 1982), suggests that people spontaneously use vocabulary learning strategies quite similar to the keyword method. Third, the keyword method is an effective mediation technique for learning unfamiliar vocabulary (for reviews, see Gruneberg & Jacobs, 1991; Levin & Pressley, 1985; Pressley, Levin, & Delaney, 1982). Current controversies about the generalizability of the keyword method (Gruneberg, 1998; Thomas & Wang, 1996; Wang & Thomas, 1999; van Helle & Mahn, 1997; Ellis & Beaton, 1993) do not question its benefits when there are opportunities for rehearsal and when the words in the two target languages are perceptually dissimilar and no natural mediators are available. Fourth, using the keyword method to learn unfamiliar material provides experimental control over the mediation process because the same keyword mediators are used by all participants, yet participants can still rely on individual knowledge in generating the interactive image relating the mediator and the English translation. Finally, by studying memorization of vocabulary items that are relatively difficult to learn because they are perceptually dissimilar and lack cognates, we expected that participants would find the presented keyword mediators helpful and thus use them in memorizing the items.

Retrieval Models of Vocabulary Items Learned Using Keyword Mediators

Once vocabulary items have been memorized with the keyword method, the English translation can be retrieved in two different ways. According to the direct model, which is based on the association hypothesis, the English equivalent is retrieved directly from the foreign word through a simple associative link connecting the foreign word and the English equivalent. In contrast, according to the mediation model the English equivalent can only be retrieved by using the foreign word to retrieve the keyword mediator, which, in turn, cues the interactive image linking the keyword with the English translation. Current accounts of the retrieval processes in the keyword method are more or less detailed versions of this basic mediation model (Atkinson & Raugh, 1975; Clark & Paivio, 1987; Desrochers & Begg, 1987; Paivio & Desrochers, 1981; Turnure & Lane, 1987).

Previous studies of the keyword method have provided evidence for a mediated retrieval account. Sweeney and Bellezza (1982), for example, showed that recalling the English translation depended on recalling the keyword. Additional support has been provided by studies in which verbal report of keywords or images correlate with recall (Delaney, 1978; Ellis & Beaton, 1993; Pressley, Levin, Kuiper, et al., 1982; Pressley et al., 1983; Pressley et al., 1980). Unfortunately, in many of these studies, participants are asked to report strategies they used to learn the vocabulary items rather than what they were aware of in retrieving them. In addition, the reports have been gathered after multiple trials or delays of more than a few seconds, jeopardizing the reliability and validity of the reports (see Ericsson & Simon, 1980, 1993; Richardson, 1998). Finally, most of these reports have not been validated by any convergent measures, such as response accuracy or latencies. In summary, previous keyword studies have provided support for the mediation model; as suggested earlier in our earlier discussion of the mediation and direct hypotheses, however, alternative accounts of the results are possible.

The case for the mediation model could be strengthened by providing evidence for the details of the retrieval process such as the intermediate cognitive structures and individual processing steps that mediate retrieval of the English equivalent. Retrospective verbal reports collected immediately after individual retrieval trials, for example, would provide a more reliable and valid measure of what was processed in retrieving a specific vocabulary item. Convergent evidence for any reported intermediate structures and processes would ensure the validity of these verbal report data.

Finally, the critical issue is the continued involvement of keyword mediators in memory processes, because even if the retrieval process is initially mediated, repeated access of the memory trace may alter the process. Some keyword researchers have argued that keyword mediators provide a temporary bridge until a direct link between the foreign word and its English equivalent can be strengthened (Atkinson & Raugh, 1975). Alternatively, the keyword mediator may remain involved in memory retrieval processes long after initial learning. Some have suggested, for example, that information learned mnemonically remains linked to the original mnemonic (see Higbee, 1978; Levin & Pressley, 1985). Levin and Pressley noted that this “mnemonic dependency” issue has not been investigated directly. Furthermore, they cited previous studies that have found dramatic increases in retrieval speed with practice (e.g., Corbett, 1977; Levin et al., 1983), but they noted a need for studies that examine the role of mediators as the retrieval process becomes automatic.

General Method for Assessing the Structure and Duration of Retrieval Processes

To distinguish between the mediation and direct models, we developed a set of convergent methods to provide evidence for the structure and duration of the underlying retrieval processes and any component steps of this process. The first part of the method was a task decomposition analysis. In all of the experiments reported here, participants learned Spanish–English pairs using the keyword method. After learning the vocabulary items, participants were tested on the items using a cued-recall procedure, in which the Spanish word was presented and participants responded with the English equivalent. This procedure is called the vocabulary retrieval task (or vocabulary task) and corresponds to the general testing procedure used in many vocabulary learning studies. In addition, participants were tested on two additional retrieval subtasks. The first was the keyword retrieval subtask (or keyword subtask), in which participants saw the Spanish word and responded with the appropriate keyword. The second was the English retrieval subtask (or English subtask), in which participants saw the keyword and responded with the corresponding English word.

According to the mediation model, the keyword and English subtasks together rely on the same cognitive structures and processing steps involved in the vocabulary retrieval task. It follows that if the vocabulary task is successfully carried out, both subtasks should be successfully carried out. Conversely, if the vocabulary
task cannot be successfully carried out, then at least one of the subtasks should be unsuccessful as well. On the other hand, according to the direct model, accessing the English equivalent from the Spanish word does not depend on accessing the keyword or interactive image but involves only the single step of accessing the English equivalent from the Spanish word. Therefore, performance on the vocabulary task would not relate to performance on the two subtasks. In fact, with sufficient practice retrieving the vocabulary pairs, cued-recall accuracy on the vocabulary task should be greater than that on the two subtasks.

Response latencies for the three retrieval tasks provide an additional means to test the two retrieval models. The mediation model assumes that processing times for the vocabulary task should relate systematically to processing times for the keyword and English retrieval subtasks. Specifically, latencies for the vocabulary task should be longer than for either of the subtasks, because each subtask involves only a subset of the retrieval steps necessary to retrieve the English equivalent given the Spanish word. Furthermore, anything that influences one of the component steps should influence the overall task and vice versa. Thus, practice on the vocabulary task would be expected to influence retrieval latencies for the two subtasks and vice versa. The direct model, on the other hand, predicts no relationship between task latencies for the vocabulary task and the subtasks. In fact, with practice, the latencies for the vocabulary task should be shorter than the latencies for either of the subtasks because practicing the vocabulary task strengthens only the direct link between the Spanish word and the English equivalent. In addition, practice on the vocabulary task should not affect latencies on the subtasks, nor should practice on one or both of the subtasks influence latencies on the vocabulary task.

This task decomposition method provides a finer grained analysis of retrieval processes than simply analyzing the correlation between recall of English translations and keyword mediators because it examines the individual processing steps that may mediate retrieval. In particular, none of the previous studies of mediation have used latency measures to describe the number or sequence of steps. A convergent source of evidence for the underlying retrieval processes studied here was verbal report data. During testing of the vocabulary items, on a randomly selected subset of retrieval trials, participants gave a retrospective report immediately after an individual retrieval trial of what they could remember thinking from the moment they saw the Spanish word until they came up with a response. According to the mediation model, these reports should contain references to the keyword or interactive image, as well as to the Spanish word and English equivalent. According to the direct model, the reports were expected to contain only references to the Spanish word and the English equivalent, without mention of intermediate steps or cognitive products.

Finally, an interference paradigm was developed to assess the role of keyword mediators in retrieval, regardless of whether the keywords are actually retrieved into working memory and thus available for verbal report. This paradigm is useful after extended practice when the English equivalents might be retrieved from the Spanish words directly—without any intermediate working memory steps. The paradigm involves participants learning the vocabulary pairs using keyword mediators followed by assessment of retrieval using the task decomposition and verbal report methods. This testing is followed by an interference phase, in which participants learn and practice retrieving new keyword–English pairs for half of the items (interference condition) but not for the other half of the items (noninterference control condition). Afterwards, the original Spanish–English vocabulary pairs are retested on the vocabulary task only. If retrieval is accomplished through a direct link between the Spanish word and English equivalent, there should be no effect of the interference manipulation on retrieval of the pairs. On the other hand, if retrieval of the vocabulary pairs still relies on the original keyword mediators, the interference manipulation should affect retrieval of the pairs.

Overview of the Experiments

All of the studies reported here followed the same general design. Participants first learned a set of Spanish–English vocabulary pairs using the keyword method and were tested to ensure mastery of all items. An initial test then examined whether retrieval was mediated or direct. Retrieval performance was then manipulated in different ways in each experiment, and afterwards a final test re-examined whether retrieval was mediated or direct. In Experiment 1 the experimental manipulation was a 1-week or 1-month delay to induce substantial forgetting of items. In Experiment 2, participants received moderate practice on the vocabulary task or on the keyword and English subtasks. In Experiments 3 and 4 participants received extended practice on the vocabulary task or on the English subtask. Experiment 4 contained an additional phase after the final tests, in which the interference paradigm was used to further assess retrieval processes following extended retrieval practice.

Experiment 1

Immediately after learning a set of Spanish–English vocabulary items with the keyword method, both English translations and keywords are likely to be accurately retrieved, making it difficult to distinguish the direct and mediation models. However, the two models make very different predictions about performance after a delay, when forgetting has occurred. In the first experiment, we tested participants' memory for the vocabulary items and keywords after 1 week or 1 month. The mediation model predicted that vocabulary recall would be limited by the speed and accuracy of the component processes involving the keyword. In contrast, the direct model predicted no such dependency. In addition, the mediation model predicted that verbal reports would include references to the Spanish word, keyword mediator, and the English equivalent, whereas the direct model predicted references only to the Spanish word and the English equivalent.

Method

Participants. A total of 24 undergraduate students at the University of Colorado received course credit for participating in this experiment. The experiment was restricted to individuals who had no previous training in Spanish.

Apparatus. The study and test materials were presented to each participant individually using a 12-in. monochrome monitor controlled by an IBM XT PC. Latencies for the study trials were recorded using a Carnegie Mellon University button box connected to the computer. For the cued-recall trials, participants answered orally, and response latencies were
registered through a microphone and a voice-actuated relay connected to the button box and computer. In addition, a cassette tape recorder was used to record participants' responses and verbal reports.

**Materials and design.** The stimuli were 42 Spanish–English pairs and their associated keywords (e.g., perro–bear–dog). Twenty-nine of the items were taken from two previous keyword studies (Hall et al., 1981; Pressley, 1977). An additional 11 items were selected using criteria similar to those used by Hall et al. and Pressley. The stimuli were selected so that both the keyword and English word were concrete, imaginable nouns. In many cases, the keyword corresponded to the first syllable of the Spanish word. A complete list of the stimuli appears in Appendix A.

Three lists were constructed, using the same 42 three-word stimuli for each list. The first six items on each list served as filler items that were always presented before the other 36 items. The responses for the filler items were not analyzed.

Participants performed three retrieval tasks for each item within a block of testing. These were the vocabulary, keyword, and English retrieval tasks, in which participants saw the Spanish word and retrieved the English equivalent, saw the Spanish word and retrieved the keyword, or saw the keyword and retrieved the English equivalent. Task order was counterbalanced for the 36 target items and the 6 filler items on each list. As there were six task orders, each task order was used for six items and one control item. Task order assignment was random, but, across the three lists, no item was assigned to the same task order twice. Half the items on each list were verbal (i.e., participants concurrently or retrospectively reported their thoughts); the other half were silent. Assignment of items to the verbal and silent conditions was also counterbalanced across participants. Two different randomized presentation orders were used.

**Procedure.** The experiment was conducted in two sessions. The first session consisted of four phases: an introductory phase, an acquisition phase, a dropout phase, and a test phase.

In the introductory phase, the experimenter gathered background information concerning each participant's previous foreign language experience. The participants then received instructions and practice giving think-aloud and retrospective verbal reports (see Crutcher, 1990, for the detailed procedure). Following the verbal report instructions and practice, the experimenter explained the details of the keyword method, and participants practiced learning and retrieving sample items. The practice tasks were the same as those used for the acquisition and dropout phases described below.

In the acquisition phase, the vocabulary items were presented individually with a computer display consisting of the Spanish word on the left, the keyword in the middle, and the English word on the right. Participants were instructed to study each vocabulary item as follows: “First, look at the Spanish word and pronounce it as best you can. Then look at the keyword and try to relate it to the Spanish word by noticing any similarity in sound or appearance. Then form an interactive image between the keyword and English equivalent.” For half the items, a cue instructing participants to think aloud while studying the items preceded the onset of the three words. Each vocabulary item was presented only once and study of the items was self-paced, with a maximum study time of 20 s before the next vocabulary item was presented.

A dropout procedure followed the acquisition procedure to ensure that participants were able to recall the English equivalents for all vocabulary items. For each trial, a Spanish word appeared on the left and a question mark appeared to the right. Participants responded by saying aloud the correct English equivalent as soon as they thought of it. For items not recalled correctly within 15 s, feedback was provided for 5 s, consisting of the original three words, displayed as before. Missed items were retested until each item had been correctly recalled once. No verbal reports were collected during the phase.

Following completion of the acquisition and dropout procedures, the test phase began. After instructions on the new tasks (the keyword and English subtasks), participants practiced performing the new tasks and giving retrospective reports. Throughout the test phase, participants performed three different retrieval tasks for each item: the vocabulary task, the keyword subtask, and the English subtask. For each retrieval task, the appropriate cue (Spanish word or keyword) was displayed where it had originally appeared during study and a question mark appeared in the location where the target word (keyword or English equivalent) had previously appeared. Participants had 15 s to make a response. No feedback was provided concerning accuracy. For the verbal items (half the trials), a visual cue followed the offset of the stimulus display, asking participants to give a retrospective report of their thoughts during the retrieval of the response.

The test phase for Session 1 consisted of two test blocks with three subblocks of testing within each block.1 For each subblock, all 42 vocabulary items were tested, but for a given item only one of the three possible retrieval tasks was performed. Half the participants returned after 1 week and half after 1 month to complete two additional blocks of testing.

**Results**

Only the 36 target items on each list were analyzed. Voice key and other equipment-related errors occurred on 2.75% of the trials and were excluded from further analysis. The analyses of variance (ANOVAs) for the recall and latency analyses were performed on participant means computed by averaging across the task orders. For the recall analyses, a *data point* was the proportion correct score based on 36 observations for a given Task X Verbal Condition X Test Occasion cell; for the latency analyses, a data point was a mean RT computed using the correct retrieval trials from the 36 available observations. The immediate results (Blocks 1 and 2) and delayed results (Blocks 3 and 4) were analyzed separately.

**Proportion of words recalled.** As expected, retrieval accuracy at immediate test was at ceiling (above .93 for all three retrieval tasks for all levels of the other factors), and therefore no further analyses are reported here. For the delayed test, mean proportion of items correctly recalled as a function of retrieval task and delay interval is displayed in Figure 1.2 The analysis of primary interest here was the comparison of recall performance on the three retrieval tasks. The pattern of results was clear-cut: Recall accuracy on the vocabulary retrieval task was significantly less than the average accuracy on the subtasks, based on an a priori comparison of the vocabulary task performance to the two subtasks, $F(1, 22) = 56.22, p < .001, MSE = .0113$. Comparing the vocabulary task to each of the subtasks separately showed that vocabulary task performance was reliably less than performance on the keyword subtask, $F(1, 22) = 56.62, p < .001, MSE = .0316$; but it was not significantly different from performance on the English subtask, $F(1, 22) < 1.0$. For the overall analysis, in addition to the main effect of task, $F(2, 44) = 52.79, p < .001, MSE = .0220$, there were the expected main effects of delay (1 week vs. 1 month), $F(1, 22) = 30.50, p < .001, MSE = .0996$, and verbal report (silent vs. discussed).

---

1 Two blocks of testing were used to obtain as stable an estimate as possible of the reaction times (RTs) for the three retrieval tasks.

2 The error bars in figures represent the standard error of the mean. For the latency data these error bars are different heights above and below the mean because they are based on the log-transformed data reconverted to the original scale by antilogging the values and adding back a constant of 610 ms. These error bars are included as a visual aid in comparing the graphically displayed means. For all analyses (graphed or not), we have reported the mean squared error in the text.
verbal), $F(1, 22) = 41.01, p < .001, \text{MSE} = .0063$, as well as a Task $\times$ Delay interaction, $F(2, 44) = 27.51, p < .001, \text{MSE} = .0220$, and Task $\times$ Verbal Report interaction, $F(2, 44) = 26.95, p < .001, \text{MSE} = .0020$. The main effect of verbal report as well as the Verbal Report $\times$ Task interaction were anticipated because generating a retrospective verbal report requires an additional retrieval of the item in question and thus strengthens the associated memory trace for the item. Moreover, these effects require no qualification of the primary results, because the vocabulary versus subtask differences were the same as those reported above for every level of the verbal report and delay factors—vocabulary task recall was significantly less than keyword subtask recall but not significantly different than English subtask recall.

Recall analyses for individual items. A more stringent test of the mediation account is to compare performance on the three retrieval tasks for each vocabulary item for a given participant within each test block. If retrieval of the English equivalent for a specific Spanish word (e.g., perro) depends on successfully retrieving the keyword (pear) and then using that keyword to retrieve the English translation (dog), it follows that if the vocabulary retrieval task is successful, both subtasks will be successful; if the vocabulary retrieval task is unsuccessful, the likelihood of success on both subtasks will be quite small. Small deviations from this pattern of results would be consistent with the mediation account; large deviations from this pattern of results would be more consistent with the direct model. Thus, consistently correct retrieval of a vocabulary item’s English equivalent but incorrect retrieval accuracy on one or both of the subtasks would support the direct model.

To test these predictions we compared performance on the three retrieval tasks for each vocabulary item in a test block for each participant, scoring whether both, neither, or one of the two subtasks was correctly performed. Looking at both correctly retrieved and incorrectly retrieved vocabulary task items separately, we then computed in each case the proportion of items for which both subtasks were correctly performed. Given a ceiling effect on the immediate retention test, only the results for the delayed tests were analyzed. In addition, overall recall was high for the 1-week delay test, so these results should be interpreted somewhat cautiously.

When the vocabulary retrieval task was successful, the proportion of items for which both subtasks were successful was quite high: .91 for the 1-month retention test (.97 for the 1-week retention test). On the other hand, when the vocabulary retrieval task was unsuccessful, the proportion of items for which both subtasks were successful was low: only .05 for the 1-month retention test (.19 of the items for the 1-week retention test). In other words, whenever an error was made retrieving the English translation of the Spanish word, an error was made on at least one of the subtasks as well. Further analyses of the incorrectly retrieved vocabulary items revealed that these errors were predominantly associated with errors made on the English subtask (.85 of the incorrect items tested at 1 month and .62 of the incorrect items at 1 week) rather than errors made on the keyword subtask.

Retrieval times. Another way to evaluate the mediation and direct models is to compare latencies for the vocabulary retrieval task and the two subtasks. According to the mediation model the vocabulary retrieval task should always take longer than the subtasks, whereas the direct model predicts no such difference. This especially should be true in comparing the vocabulary task to the English subtask, because the English subtask uses the keyword mediator as a retrieval cue, which, according to the direct model, is assumed not to be involved in the vocabulary retrieval process.

In general, the distribution of response latencies is skewed when participants are permitted to respond beyond several seconds (Ratcliff, 1978); however, this distribution can often be closely approximated by a three-parameter log-normal distribution (Ericsson, 1974). The distribution assumes a fixed response component (estimated to be 610 ms in the current studies). This fixed response component was subtracted from each latency, and the resulting number was then log transformed (base 10), generating a symmetric distribution (Crutcher, 1990). All of the analyses reported here (and in Experiments 2, 3, and 4) are for the transformed data; however, analyses were performed on the transformed as well as the untransformed latency data, and the pattern of results in all cases was the same unless reported otherwise in the text. To facilitate interpretation of means in tables and graphs, we reconverted the reported values to the original scale by applying the antilog transformation and adding back a constant of 610 ms. As with the recall data, separate analyses were performed for the immediate and delayed latency results. For the delayed results, a few data points were missing (this occurred when none of the 18 observations used to compute the RTs were correct). These missing values constituted less than 2% of the data and were replaced with participant means.

The primary analysis for the latency data compared performance on the vocabulary task and the two subtasks. Retrieval latencies as a function of retrieval task and verbal report condition (silent vs. retrospective) are presented in Figure 2. An a priori comparison of the vocabulary task to the average of the subtask latencies showed that vocabulary task latencies were significantly greater than the average of the keyword and English subtask latencies, $F(1, 23) = 219.27, p < .001, \text{MSE} = .0029$. Additional comparisons revealed that the vocabulary task latencies were greater than the keyword subtask latencies, $F(1, 23) = 77.02, p < .001, \text{MSE} = .0064$, and greater than English subtask latencies, $F(1, 23) = 152.84, p <
The Task X Verbal Report interaction was reliable for the log-transformed times, $F(2, 46) = 5.30, p < .01$, $MSE = .0030$. In addition, the results of an overall ANOVA with task and verbal report as the within-subjects factors produced a significant main effect of task, $F(2, 46) = 48.81, p < .001$, $MSE = .0065$, but no main effect of verbal report, $F(1, 23) < 1$. The Task X Verbal Report interaction was reliable for the log-transformed times, $F(2, 46) = 3.41, p < .042$, $MSE = .0015$, but not for the raw times, $F(2, 46) = 2.83, p > .05$, $MSE = 21,771$.

Results for the analysis of the delayed tests are presented in Figure 3. The a priori comparison of the vocabulary task to the average of the subtask latencies again showed that vocabulary task latencies were significantly greater than the average of the subtask latencies, $F(1, 22) = 75.93, p < .001$, $MSE = .0118$. Additional comparisons of the vocabulary task latencies to the individual subtask latencies also revealed that vocabulary task latencies were greater than the keyword subtask latencies, $F(1, 22) = 87.06, p < .001$, $MSE = .0154$, and greater than the English subtask latencies, $F(1, 22) = 31.27, p < .001$, $MSE = .0073$. In addition, the results of the overall ANOVA with task and verbal report as the within-subjects factors and delay (1 week vs. 1 month) as the only two between-subjects factor produced significant main effects of task, $F(2, 44) = 61.72, p < .001$, $MSE = .0109$, and delay, $F(1, 22) = 13.34, p < .001$, $MSE = .0527$, as well as a main effect of verbal report that was reliable for the log-transformed times, $F(1, 22) = 5.30, p < .05$, $MSE = .0081$, but not for the raw times, $F(1, 22) = 2.90, p > .05$, $MSE = 745,087$. The only reliable interaction was for the untransformed latencies: The Task X Delay interaction was significant, $F(1, 22) = 3.92, p < .05$, $MSE = 578,514$.

Verbal reports. The retrospective reports for the vocabulary retrieval task provided an alternative means of assessing whether the vocabulary retrieval process was mediated or direct. An encoding scheme was developed prior to collecting and transcribing the verbal reports, and two encoders were employed to test the reliability of the encoding scheme. Agreement was 92% or higher for each of the five variables encoded. After a general discussion of the encoding rules with the experimenter, with no reference to specific protocols, the encoders recoded the protocols on which they disagreed. Agreement was 99%.

For the correct vocabulary task trials, participants reported keyword mediators 85% of the time and reported no mediation of any kind 6.4% of the time. For a small proportion of retrievals (7.9%), the verbal reports were classified as ambiguous, which meant that the participant reported the keyword and the English equivalent but mentioned additional information that could not be unambiguously categorized. For the correct vocabulary task trials, participants reported keyword mediators 85% of the time and reported no mediation of any kind 6.4% of the time. For a small proportion of retrievals (7.9%), the verbal reports were classified as ambiguous, which meant that the participant reported the keyword and the English equivalent but mentioned additional information that could not be unambiguously categorized.

Briefly the encoding scheme was as follows: For each retrospective report, five variables were coded, corresponding to the Spanish word, the Spanish–keyword link, the keyword, the keyword–English link, and the English equivalent. The encodings for the Spanish word, the keyword, and the English word (the three relevant elements for the current analyses) were straightforward: if the word was mentioned in the retrospective report or had appeared as part of the computer display (e.g., the Spanish word was a cue in the vocabulary retrieval task), it was encoded as present ("1"); if not mentioned, it was encoded as absent ("0"). If there was not enough information to decide or if it was ambiguous (e.g., participant reported thinking of the keyword but did not actually mention the keyword), it was coded as a "9."

Using the codings for the Spanish word, keyword, and the English equivalent, the verbal reports were categorized as follows: (a) mediated: Spanish word and English equivalent present and keyword mentioned; or Spanish word and English equivalent present and keyword mentioned along with a clear link between the keyword and English equivalent (e.g., “Saw the Spanish word perro/Remembered pear/Then and the dog holding a pear in its mouth); and (b) unmediated: Spanish word and English equivalent present, but keyword not mentioned and no other information reported (e.g., “Saw perro/And thought dog”). A more complete description of these categories may be found in Crutcher (1990).³

For the correct vocabulary task trials, participants reported keyword mediators 85% of the time and reported no mediation of any kind 6.4% of the time. For a small proportion of retrievals (7.9%), the verbal reports were classified as ambiguous, which meant that the participant reported the keyword and the English equivalent but mentioned additional information that could not be unambiguously categorized.

³ In the initial analysis of these protocol data, two categories of mediation were generated: (a) keyword plus mediating link and (b) keyword mediator only. For ease of comparison across the three studies described here, these were combined into a single category: keyword mediation.
vocabulary retrieval should strengthen the mediating cognitive
structures between the foreign word and translation. The goal of
the second study, then, was to examine retrieval processes for
vocabulary pairs following moderate vocabulary retrieval practice.
After initial mastery of the vocabulary items, participants practiced
groups of vocabulary items differently. For one third of the items
participants practiced retrieving only the English equivalent from
the Spanish word (full practice). For a second third of the items,
they practiced retrieving the keyword from the Spanish word and
the English equivalent from the keyword (subtask practice). This
served as a control condition to maintain mediated retrieval,
against which to compare the effects of vocabulary retrieval (full
practice). Finally, the last third of the items received no practice at
time (no practice). This served as an additional control condition
against which to compare the effects of full and subtask practice.
These control conditions are important in differentiating the me-
diation and direct accounts following practice.

The mediated and direct models made different predictions
concerning the pattern of latency results for the various retrieval
tasks after the practice phase. According to the direct model, the
pattern of latencies for the vocabulary task and the two subtasks
was expected to depend on the type of practice (full, subtask, or no
practice). Specifically, following subtask practice or no practice
the latencies for the vocabulary task were expected to be greater
than those for the subtasks, but following full practice the opposite
pattern was expected; namely, the vocabulary task would be faster
than the subtasks. In contrast, the mediation model predicted
longer latencies on the vocabulary task after all three types of
practice. In summary, the direct model predicted a crossover
interaction of task and practice condition, whereas the mediation
model did not.

The two models also made different predictions about retrieval
performance on the unpracticed tasks. According to the direct
model, practicing the vocabulary task strengthens only the direct
link between the Spanish word and its English translation, whereas
subtask practice strengthens only the Spanish-to-keyword and
keyword-to-English translation links, neither of which functionally
influences vocabulary retrieval. Therefore, no nonspecific practice
effects were predicted by the direct model—full practice was not
expected to facilitate performance on the unpracticed subtasks, nor
was subtask practice expected to facilitate performance on the
unpracticed vocabulary task. In contrast, according to the media-
tion model, either type of practice strengthens the same set of
cognitive structures and was predicted to produce improvement in
retrieval performance on the unpracticed tasks.

For the verbal report data, the mediation model predicted key-
word mediation on vocabulary retrieval following all three types of
practice, whereas the direct model predicted an absence of reported
keyword mediation following full practice but considerable re-
ported mediation for the other two practice conditions (subtask and
no practice). Given that retrieval performance was assessed im-
mediately after acquisition and practice, retrieval accuracy was
expected to be very high and of little use in distinguishing the
retrieval models.

Method
Participants. The participants were 18 undergraduate students at the
University of Colorado, recruited by an ad placed in the campus news-
paper. The students were paid for their participation in the experiment. The
experiment was restricted to individuals who had no previous training in Spanish.

Materials and apparatus. The stimuli were 45 Spanish-English pairs along with their keywords. Forty-one of the Spanish-English pairs and keywords from the first experiment, along with four new Spanish-English pairs and keywords, were used to create the set of 45 stimuli used in the second experiment (see Appendix B for the complete list of items). The four new items were selected using the same criteria described in Experiment 1. The same IBM PC computer and tape recorder setup described in the first experiment was used to collect responses in this study.

Because there were three practice conditions (full practice, subtask practice, and no practice), three lists were created, so that each item appeared in each of the practice conditions across the three lists. In addition, three of the six possible task orders were used, with a third of the items on each list assigned to each task order and assignment of items to specific task orders counterbalanced across the three lists so that each item in the experiment appeared in each practice condition and task order equally often.

Procedure. The procedure was similar to that in Experiment 1, but with a practice phase interposed between the dropout and test phases, during which subsets of the 45 items were practiced as described below. The final test phase consisted of a single test block of three subtests across which each item was tested on all three retrieval tasks. Presentation order of the items was randomized within each subtest.

During the practice phase, subsets of the 45 vocabulary items were practiced differently. For the full-practice items, participants performed only the vocabulary-retrieval task; for the subtask-practice items, they practiced only the two subtasks; the no-practice items were not practiced at all during this phase. Thus, 30 items (15 full-practice items and 15 subtask-practice items) were practiced during this phase.

Because it was necessary to practice both subtasks in the subtask-practice condition, two trials were required for a subtask-practice item within each practice block. Therefore, the 15 subtask-practice items were blocked and counterbalanced, so that for half the items the subtask occurred first followed by the English subtask, whereas for the other half the task order was reversed. Presentation order of the items within each practice block was randomized for each participant. There were six blocks of practice.

A few other minor changes were made to the procedure. First, the computer display of the vocabulary items was modified so that instead of presenting and testing items in a horizontal linear format, as in the first experiment, the three words, and the symbols that replaced the various words during the dropout, practice, and test phases, were displayed in a stepwise format. Second, participants had 10 instead of 15 s to respond on each trial in the dropout, practice, and test phases. Third, participants gave verbal reports on all acquisition and test trials; no reports were collected during the dropout and practice phases.

Results

Voice key and other equipment-related errors occurred on 3.8% of the trials and were excluded from the analyses. As in Experiment 1, participant means were computed for each cell. There were three types of practice and three retrieval tasks, yielding 9 scores per participant. Thus the main design was completely within-subjects, with practice condition and retrieval task as the two within-subjects factors. In addition to the omnibus tests, specific comparisons (vocabulary vs. subtasks and vocabulary vs. English subtask) were performed to test the predictions of the direct and mediation models. As expected, accuracy of recall was uniformly very high (.89 or greater in all cases), and thus no statistical analyses of the accuracy data are reported.

Retrieval times. Mean retrieval times for the different retrieval tasks as a function of practice condition (full, subtask, and no practice) are displayed in Figure 4. The pattern of results suggests longer latencies for the vocabulary task relative to the two subtasks for each of the three practice conditions. The a priori comparison of vocabulary to subtask latencies was reliable, $F(1, 17) = 96.56, p < .001, MSE = .0078$, but qualified by an interaction of the contrast with practice condition, $F(2, 34) = 38.99, p < .001, MSE = .0026$. Testing this contrast for each level of practice showed that vocabulary task latencies were greater than subtask latencies for each practice condition: full practice, $F(1, 17) = 7.39, p < .015, MSE = .0049$; subtask practice, $F(1, 17) = 173.54, p < .001, MSE = .0038$; and no practice, $F(1, 17) = 61.52, p < .001, MSE = .0040$.

Rather than comparing the vocabulary task to the average of the two subtasks, a more appropriate comparison is to compare the vocabulary task to the English subtask, which required the same response as the vocabulary task. Comparison of the vocabulary task to the English retrieval subtask alone was significant, $F(1, 17) = 36.98, p < .001, MSE = .0082$, but again produced an interaction with practice condition, $F(2, 17) = 13.96, p < .001, MSE = .0026$; nevertheless, vocabulary task latencies were greater than the English subtask latencies for the full-practice items, $F(1, 17) = 4.36, p = .052, MSE = .0042$; the subtask-practice items, $F(1, 17) = 51.47, p < .001, MSE = .0051$; and the no-practice items, $F(1, 17) = 23.58, p < .001, MSE = .0040$, although the effect for the full-practice items was marginal.

Comparison of the vocabulary task to the keyword subtask produced a similar pattern of results. The vocabulary vs. keyword subtask contrast was reliable, $F(1, 17) = 70.35, p < .001, MSE = .0129$, again qualified by an interaction with practice condition, $F(1, 17) = 43.79, p < .001, MSE = .0029$. Additional contrast tests for each level of practice condition showed greater vocabulary task latencies compared to the keyword subtask for full practice, $F(1, 17) = 6.75, p = .019, MSE = .0057$; subtask practice, $F(1, 17) = 141.67, p < .001, MSE = .0057$; and no practice, $F(1, 17) = 42.73, p < .001, MSE = .0071$.

To test for nonspecific transfer effects, we compared the subtask retrieval latencies for the full and no-practice conditions and the vocabulary task latencies for the subtask and no-practice condi-

![](https://example.com/figure4.png)
tions. English subtask latencies were significantly faster following full practice than after no practice (1,750 vs. 2,176 ms), $F(1, 17) = 27.03, p < .001, MSE = .0063$. Similarly, keyword subtask latencies were significantly faster after full practice than after no practice (1,697 vs. 1,910 ms), $F(1, 17) = 17.82, p < .001, MSE = .0031$. In addition, the analysis of vocabulary task latencies showed that vocabulary task latencies were significantly faster after subtask practice than after no practice (2,074 vs. 2,595 ms), $F(1, 17) = 45.14, p < .001, MSE = .00349$.

Finally, an overall ANOVA produced main effects of task, $F(2, 34) = 34.34, p < .001, MSE = .0003$, and practice condition, $F(2, 34) = 76.48, p < .001, MSE = .0071$, and an interaction of the two factors, $F(4, 64) = 21.61, p < .001, MSE = .0029$.

**Verbal reports.** The verbal reports for the vocabulary retrieval task were analyzed as described in Experiment 1. Less than 1% of the protocols were inaudible or missing. A very small number of reports (about 4%) were not captured by our encoding categories and were coded as ambiguous.

Keyword mediators were reported on the majority of vocabulary retrieval trials for all three practice conditions: .78, .81, and .86 for the full-practice, subtask-practice, and no-practice conditions, respectively. There were reliably fewer reports of mediation for the full-practice items than for the subtask- and no-practice items, $F(1, 17) = 11.16, p < .004, MSE = .0071$. However, the difference between the proportion of reported mediation for the subtask-practice and no-practice conditions was not reliable, $F(1, 17) < 1.0, p > .05$. The proportion of unmediated retrieval reports showed a similar pattern of results with .17, .07, and .09 for the full-practice, subtask-practice, and no-practice conditions, respectively. As a validity check on the content of the reports, for the participants who reported both mediated and direct retrievals, we again analyzed the response latencies for different categories of reported mediation. Mean retrieval latencies were longer for the mediated than the unmediated vocabulary task retrievals, $F(1, 15) = 18.72, p < .001, MSE = .0220$. Given that very few participants reported any unmediated retrievals for the subtask-practice and no-practice conditions, we did a second comparison of response latencies for the two categories of reports using only the items in the full-practice condition. For this analysis 15 participants had reports for both categories. Once again, the mediated latencies (1,981 ms) were longer than the unmediated latencies (1,395 ms), $F(1, 14) = 49.01, p < .001, MSE = .0093$.

**Discussion**

The overall pattern of results in this second experiment again supports the mediation model. First, vocabulary task latencies were greater than subtask latencies for all practice conditions, consistent with the results in the first experiment. Second, participants reported keyword mediators on the vast majority of trials as in the first experiment. Third, the nonspecific transfer effects of full practice on the subtasks and the corresponding facilitating effect of subtask practice on the vocabulary task latencies are consistent with only the mediation model. According to the direct model, retrieving the vocabulary pairs should have strengthened only the direct link between the Spanish and English words and produced no improvement on subtask performance after vocabulary practice. Likewise, subtask practice should have strengthened only the Spanish-keyword link and the keyword-English link without affecting the direct link between the Spanish word and the English translation.

A mediation model based on accessing mediators into working memory to cue retrieval of the English equivalents explains most but not all of the results in this experiment. First, a mediation model cannot explain why retrieval latencies for the vocabulary retrieval task were about 200 ms faster after full practice than after subtask practice, $F(1, 17) = 8.00, p < .01, MSE = .0047$. Practice on the vocabulary-retrieval task provided an additional benefit in retrieving the English equivalent of the Spanish word beyond simply practicing the corresponding subtasks. One potential explanation is that for some of these items, direct access of the English definitions occurred, making it unnecessary to retrieve the mediating keyword and image into working memory to use as cues in retrieving the English translation. This explanation is supported by the analysis of the retrospective reports for the vocabulary task: There was 10% less reported mediation in the full-practice condition than in the subtask-practice and no-practice conditions. In addition, latency analyses showed that vocabulary retrievals without any reported mediation were about half a second faster than retrievals with reported mediation. The additional benefit of vocabulary practice can, at least in part, be explained by an increased frequency of direct access of the English equivalents from the Spanish words. At issue is the mechanism responsible for the change to direct access.

In its original form the direct model provides the simplest account of direct access by rejecting the evidence for mediation as epiphenomenal and positing a direct association or link between the foreign word and its corresponding English equivalent, which is strengthened with repeated use. Unfortunately, the direct model cannot explain the considerable evidence on mediation reported here so far. In particular, it cannot account for the speed-up of both the vocabulary task and the subtasks as a result of practice on either type of task. The direct model could be modified to account for these results by assuming a continuous strengthening of both the direct and indirect (mediated) pathways with the strength of the direct pathway eventually exceeding the strength of the mediated pathway. However, there are potential problems with such an account. First, although practice on the vocabulary task potentially could strengthen either the direct or indirect pathway, it is difficult to imagine how practice on the keyword or English subtasks could strengthen the direct link. Second, although an account based on combining two different retrieval mechanisms is difficult to reject on purely empirical grounds, it can be challenged on the grounds of parsimony.

On the other hand, the mediation model easily accounts for the speed-up of the vocabulary task and the subtasks as a function of practice on either type of task as well as the pattern of latency results and the verbal report data. The emergence of direct access, without verbally reported mediation, could be explained by a covert mediation process in which the keyword mediator is activated in long-term memory but not retrieved into working memory. In summary, most of the results of the second experiment support the mediation model over the direct model. However, the results suggest changes in the retrieval process for a few of the vocabulary items such that the pairs were directly accessed from long-term memory without any intermediate working memory steps. Given these changes in retrieval after moderate practice, the
next logical step was to examine retrieval after extended retrieval practice.

Experiment 3

The goal of the third experiment was to look at the effect of extended practice on retrieval of vocabulary pairs learned using the keyword method. The experimental procedure was the same as that in Experiment 2, except that the amount of practice was increased by a factor of 10. In addition, a few minor modifications and improvements were made to the basic method. Whereas the previous two studies tested both the keyword and English subtasks as well as the vocabulary task, the current study used only the vocabulary task and the English subtask to assess retrieval performance. In addition, during the practice phase for the subtask practice condition, participants practiced only the English subtask rather than both the keyword and English subtasks.

The reasons for these changes were as follows. First, testing only the vocabulary task and English subtask eliminated any possibility of task interference caused by using the same retrieval cue for the two different retrieval tasks. In the earlier studies, the keyword and vocabulary tasks used the Spanish word as a retrieval cue, and which task to perform was cued by the location of the retrieval cue. Our concern was that after extended practice, the responses to the three retrieval tasks were likely to become automated, thus creating response competition interference between the vocabulary task and keyword subtask. By using only the vocabulary task and the English subtask, each task used a unique retrieval cue and the cues could be displayed in exactly the same screen location. This arrangement ensured that differences in task latencies for the vocabulary and subtask comparisons would be due only to differences in the underlying retrieval processing steps. In addition, making inferences about the mediated and direct models using the decomposition method does not require comparing the vocabulary task to both subtasks.

Despite the advantages of eliminating the keyword subtask, a concern was that not seeing the Spanish words for the items in the subtask-practice condition over the extended practice period might influence performance when participants were retested on both retrieval tasks, because repeated exposure to the Spanish words might be important to learning and increased retrieval speed. To assess whether differential exposure of Spanish words influenced performance, we tested the speed of reading the Spanish words before and after extended practice for each of the two practice conditions. In addition, given that in the full-practice condition the keywords would not be seen until the final test, we also assessed reading speed for the keywords before and after practice.

Method

Participants. The participants were eight undergraduate students, recruited through an advertisement placed in the University of Colorado campus newspaper. The experiment was restricted to individuals who had no previous training in Spanish.

Materials, apparatus, and design. Thirty-two vocabulary items selected from those used in the first two experiments (see Appendix C) were presented individually using a computer display controlled by an IBM PC as described in the previous experiments. The design was completely within subjects, with retrieval task (vocabulary vs. English subtask), practice condition (full vs. subtask practice), test occasion (initial vs. final), and verbal condition (verbal vs. silent) the factors of main interest.

For the initial and final tests, all vocabulary items were tested on both the vocabulary retrieval task and the English retrieval subtask. During the practice phase, half the items received practice on only the vocabulary task (full-practice items), and the other half of the items were practiced on only the English retrieval subtask (subtask-practice items). Finally, verbal reports were collected on half of the 32 vocabulary items (verbal), and no reports were collected on the other half of the items (silent). Initial assignment of items to the different task orders, practice conditions, and verbal report conditions was random and the conditions were counterbalanced so that each vocabulary item appeared in each combination of the three factors across participants.

Procedure. The experiment was conducted in five phases: an introductory phase, a learning phase, an initial test phase, a practice phase, and a final test phase. At the beginning of the introductory phase, the experimenter obtained background data concerning participants' previous foreign language experience. Participants were then given a reading task that required reading the Spanish words and keywords for all of the vocabulary items used in the experiment. On each trial, a fixation point first appeared in the center of the screen, followed by either a Spanish word or a keyword. Participants read each word aloud as soon as it appeared. Each participant completed five randomized blocks of testing on the reading task (for additional details concerning the procedure, see Crutcher, 1992). After the reading task, acquisition and dropout procedures occurred, as described in the previous study. The dropout procedure was modified slightly to make it more consistent with the procedure used at initial and final test. Participants were tested on all items using the vocabulary task and the English subtask. The task displays were modified so that all items appeared at the same location on the screen. For incorrect responses feedback was displayed for 10 s: for the vocabulary task, the feedback was the Spanish word, keyword, and English equivalent; for the English subtask, it was the keyword and English equivalent. Task order was blocked across the 32 items, so that for each half of the items, the vocabulary task occurred first followed by the English subtask and for the other half of the items the order of the two tasks was reversed. In all other respects the dropout procedure was identical to that used in the previous experiment.

After dropout, an initial test was given with each vocabulary item tested on the vocabulary and English retrieval tasks. The procedure was identical to that used for the dropout phase, but no response accuracy feedback was provided and retrospective verbal reports were collected after half of the trials.

Following initial testing was the practice phase in which the vocabulary items were divided into two groups so that for half the items participants practiced only the vocabulary retrieval task (full-practice condition), and for the other half they practiced only the English subtask (subtask-practice condition). Feedback on incorrect responses consisted of the correct response—the English equivalent—displayed for 3 s in the center of the screen. This practice phase lasted for 8 sessions, scheduled every other day, with 10 blocks of practice per session. One participant's training schedule deviated slightly, with a break of 3 days between Sessions 5 and 6.

After completion of the practice phase, a final test was given in which all vocabulary items were retested on both the vocabulary-retrieval task and the English retrieval subtask. Order of testing for the two retrieval tasks was counterbalanced across items in each of the practice conditions (full practice and subtask practice). As at initial test, there were two blocks of testing, so that each item was tested on each retrieval task twice. Cued recall and latencies were collected on every trial, and retrospective verbal reports were collected on half the trials. Immediately following final retrieval testing, participants were retested on the reading task.

Results

All analyses reported here are for the initial and final tests only. Voice key errors occurred on 1% of the trials and were excluded.
from the analyses. The recall results for the vocabulary task and English subtask on the initial test were .95 and .97 proportion correct for the subtask-practice items and .96 and .97 for the full-practice items. The corresponding scores on the final test were .96 and 1.00 for subtask-practice items and 1.00 and .99 for the full-practice items. No further analyses of the recall results were performed. The ANOVAs for the latency and the verbal reports were performed on participant means. For the retrieval task latency analyses, a data point was a mean RT computed from the correct retrieval trials of 32 observations (2 blocks of testing on 16 items). For the verbal report analyses, a data point was a proportion based on the number of correct trials out of 16 observations (2 blocks of testing on 8 items).

**Retrieval task latencies.** Mean retrieval times for the vocabulary task and English subtask as a function of test occasion (initial vs. final) and practice condition (full practice vs. subtask practice) are displayed in Figure 5. Inspection of the means suggests the pattern of results predicted by the direct model: Vocabulary task latencies were greater than English subtask latencies for all items before practice, but after retrieval practice, the pattern depended on the type of practice. For the subtask-practiced items, vocabulary task latencies were longer than English subtask latencies; for the full-practice items, however, vocabulary task latencies were shorter than the English subtask latencies. This pattern of results was supported by the statistical analysis, with main effects of test occasion, $F(1, 7) = 103.90, p < .001, MSE = .0543$; retrieval task, $F(1, 7) = 70.95, p < .001, MSE = .0143$; and practice condition, $F(1, 7) = 70.39, p < .001, MSE = .0029$. There was no main effect of verbal report condition, $F < 1$. There was the Retrieval Task $\times$ Practice Condition $\times$ Test Occasion interaction predicted by the direct model, $F(1, 7) = 101.43, p < .001, MSE = .0143$. There was also a Retrieval Task $\times$ Practice Condition interaction, $F(1, 7) = 146.40, p < .001, MSE = .0071$, and a Practice Condition $\times$ Test Occasion interaction, $F(1, 7) = 48.69, p < .001, MSE = .0029$. An interaction of verbal report condition and retrieval task was reliable for the log-transformed times, $F(1, 7) = 5.96, p < .05, MSE = .0086$, but not for the untransformed times, $F(1, 7) = 2.09, p = .19, MSE = 121.595$. Given the triple interaction of retrieval task, practice condition, and test occasion and the prediction by the direct model of different patterns of results before and after practice, further analyses of the simple effects for the initial and final tests were performed. These analyses confirmed that all of the vocabulary task and English subtask differences observed in Figure 5 were significant, with vocabulary task latencies greater than English subtask latencies in every case except for the full-practice items on the final test, where the pattern of latencies was reversed.

Because the nonspecific transfer results in Experiment 2 showed that practice on the vocabulary task (full practice) influenced latencies on the subtasks and vice versa, we performed a similar analysis here. Performance on the retrieval tasks before extended practice served as a no-practice control. For the items in the subtask-practice condition, there was a significant decrease in vocabulary task latencies before (1,646 ms) versus after (1,375 ms) extended subtask practice, $F(1, 7) = 25.62, p < .001, MSE = .0029$. Similarly for the full-practice condition, there was a significant decrease in the English subtask latencies before (1,290 ms) versus after vocabulary-retrieval task practice (963 ms), $F(1, 7) = 65.9, p < .001, MSE = .0043$.

**Verbal reports.** The encoding scheme for the retrospective reports was based on the encoding scheme used in the first two studies, with a few minor refinements. For a complete description of the encoding rules, see Crutcher (1992). Using this encoding scheme, the verbal reports were coded as follows. Each participant's verbal reports for the acquisition phase and for the initial and final tests were transcribed and segmented by trial. The reports for the initial and final tests of all 8 participants were then concatenated into a single file. An encoding program randomly presented the protocols, one at a time, to a coder. For each protocol presented, the coder was provided with information concerning the type of retrieval task (vocabulary or English subtask) and with the three words (Spanish, keyword, and English) associated with that retrieval trial. Protocols were presented in random order so that the coder did not know with which participant, condition, or test occasion a protocol was associated.

Two coders coded the protocols. The first coder, a paid assistant, encoded the entire set of 1,024 protocols using a set of rules (Crutcher, 1992). The second coder, the experimenter, coded a randomly selected subset (25%) of the 1,024 protocols as a reliability check on the primary coder's codings. Agreement across all five variables for the two coders was 96% (agreement was 99% for the three variables used in the present analyses). All analyses reported here were performed on the first coder's codings.

The verbal report analyses reported were restricted to correct vocabulary retrieval trial tasks only. Of 512 possible trials, 488 (95.3%) were available for analysis. Using the codings, the verbal reports were categorized as mediated (Spanish word and English equivalent present and keyword mentioned), and unmediated (Spanish word and English equivalent present, but keyword not mentioned and no reference to mediation of any other kind). For each participant, the proportion of trials for which mediation was reported was computed for the subtask-practice and full-practice items at initial and final test. Reported mediation as a function of practice condition (subtask vs. full) and test occasion (initial vs. final) was reliable for the log-transformed times, $F(1, 7) = 5.96, p < .05, MSE = .0086$, but not for the untransformed times, $F(1, 7) = 2.09, p = .19, MSE = 121.595$. Given the triple interaction of retrieval task, practice condition, and test occasion and the prediction by the direct model of different patterns of results before and after practice, further analyses of the simple effects for the initial and final tests were performed. These analyses confirmed that all of the vocabulary task and English subtask differences observed in Figure 5 were significant, with vocabulary task latencies greater than English subtask latencies in every case except for the full-practice items on the final test, where the pattern of latencies was reversed.

![Figure 5. Mean retrieval time (in milliseconds) as a function of retrieval task, practice condition, and test occasion in Experiment 3. The four bars on the left are for the initial test; the four bars on the right are for the final test. Solid bar = vocabulary task; open bar = English task.](image-url)
final) is displayed in Figure 6. A 2 × 2 within-subjects ANOVA on the proportions with practice condition and test occasion as the two factors showed a main effect of practice condition, $F(1, 7) = 74.42, p < .001$, $MSE = .02$, and test occasion, $F(1, 7) = 38.63, p < .001$, $MSE = .02$, as well as a Practice Condition × Test Occasion interaction, $F(1, 7) = 102.76, p < .001$, $MSE = .01$. Additional analyses confirmed that before practice the amount of mediation reported for all items was similar, $F(1, 7) < 1$, but that after practice, full practice resulted in dramatically less mediation than subtask practice, $F(1, 7) = 130.32, p < .001$, $MSE = .02$. Furthermore, the amount of mediation for the subtask-practice items did not change significantly following practice, $F(1, 7) < 1$, whereas mediation declined considerably for the items that received full practice, $F(1, 7) = 50.40, p < .001$, $MSE = .26$.

As before, latencies for the mediated and unmediated retrieval reports were compared and showed that mean RTs for the unmediated retrievals (892 ms) were faster than those for the mediated retrievals (1,629 ms), $F(1, 7) = 216.70, p < .001$, $MSE = .01$. A possible confound in this analysis was that before practice, when reported mediation was quite high, overall retrieval speed was slower, whereas following retrieval practice when reported mediation had declined considerably, overall retrieval speed was faster. Therefore, additional tests of the relationship between retrieval latencies and reported mediation were performed by analyzing the initial and final test latencies separately. The analysis of the initial test showed that latencies were faster for the unmediated reports (1,349 ms) than for the mediated reports (1,799 ms), $F(1, 6) = 41.27, p < .001$, $MSE = .001$. The analysis of the final test also showed that the unmediated retrieval reports were faster than those for the mediated reports (835 ms vs. 1,394 ms), $F(1, 7) = 92.79, p < .001$, $MSE = .01$.

Reading task latencies. The analyses of the reading task latencies were restricted to the final block of testing of the first reading test and the first block of testing for the second reading test. The reported analyses are based on the untransformed reading times, but the analyses based on the transformed latencies produced the same pattern of effects. A within-subjects analysis of variance on reading task (Spanish vs. keyword), practice condition (full vs. subtask), and test (initial vs. final) revealed a main effect of task, with reading times for the Spanish word greater than those for the keyword ($527 vs. 500 ms$), $F(1, 7) = 38.68, p < .001$, $MSE = 314.76$, qualified by the interaction of reading task and test occasion, $F(1, 7) = 6.72, p = .036$, $MSE = 166.26$. Subsequent post hoc analyses showed that reading times for the Spanish word were longer than those for the keyword before and after practice; however the difference was less before practice ($514 vs. 495 ms$), $F(1, 7) = 17.02, p < .004$, $MSE = 87.67$, than after practice ($540 ms vs. 504 ms$), $F(1, 7) = 33.75, p < .001$, $MSE = 153.37$. The analyses of the reading task latencies were restricted to the final block of initial testing and the first block of the postpractice testing to provide a more powerful analysis of the effect of practice on reading times. An analysis of reading task latencies for all five blocks before and after practice similarly revealed that reading times for the Spanish word were longer than those for the keyword: $553 versus 515 ms$, $F(1, 7) = 38.68, p < .001$, $MSE = 314.76$. None of these analyses found any reliable main effects or interactions with practice condition (full vs. subtask), indicating that the exposure to the keywords and Spanish words during the extended practice did not have an important influence on reading speed.

Discussion

Immediately following learning, both the decomposition analysis of task and subtask latencies and the retrospective report analysis of keyword mediation for the vocabulary task provided evidence that retrieval of the Spanish–English vocabulary pairs is mediated by a sequential process that required first accessing the keyword into working memory and then using that mediator to retrieve the English equivalent. After extended practice retrieving the vocabulary pairs, the retrieval process data were no longer consistent with this sequential mediation model. Instead the data strongly indicate direct access to the English equivalent from the Spanish word without any intermediate working memory steps. Participants were able to complete the vocabulary task reliably faster than the English subtask—a finding that is inconsistent with a sequential mediation model that must first access mediators into working memory. Further support for direct access after extended vocabulary practice was provided by the retrospective reports, which rarely included references to keyword mediators. Convergent evidence for this finding came from the latency analyses showing that trials with reported mediation were, on average, half a second longer than trials reported as unmediated.

In support of the validity of the retrospective reports, the latencies for the trials with reported mediation were, on the average, half a second longer than the trials with reports of direct retrieval. In summary, Experiment 3 provided strong evidence that extended vocabulary practice transforms a sequential retrieval process controlled by cues in working memory into a direct retrieval process that no longer relies on such working memory cues.

Two additional findings clarify the nature of direct access in response to extended vocabulary practice. First, reading speed for the Spanish words before and after extended practice provided no evidence that the large number of exposures had changed the perception or encoding of these words. Second, practice on the vocabulary task appeared to influence the speed of completing the...
English subtask, and practice on the English subtask appeared to influence the speed of retrieval on the vocabulary task. These important findings of transfer of practice beyond the specific task practiced, also observed in Experiment 2, are not consistent with a direct model that assumes that the link responsible for direct access is independent of the original mediated pathway. Rather, these results appear more consistent with a model that assumes continued dependence on the original mediation structure but without the need to access the structure into working memory to retrieve the English equivalent. Alternative accounts of the transfer of practice that would focus on other types of shared elements, such as perceptual encoding of the Spanish word, are not supported by our efforts to measure these processes directly by the reading task. Our results suggest that extended practice converts a sequential mediation process to direct retrieval in which no intermediate working memory steps are necessary to recover the translation. This direct access seems most consistent with a covert mediation account in which the original mediators are still involved in recovering the translation.

**Experiment 4**

Experiment 3 established that extended practice leads to direct access of the Spanish–English vocabulary pairs. At issue is the nature of this direct access. Although we have argued that the results of Experiments 3 are more consistent with a covert mediation model than a direct model, the goal of Experiment 4 was to explicitly compare the two accounts. The first part of this study was similar to that of Experiment 3, with vocabulary acquisition, vocabulary–subtask testing, extended retrieval practice, and retesting with the vocabulary–subtask procedure. Following these procedures an interference phase occurred in which new associations (interference condition) for half of the original 32 vocabulary items, participants learned paired associates (interference condition) where the first member of the pair was a previously used keyword and the second member was a new English word (e.g., for cabra–cab–goat the new keyword–English pair was cab–hospital). For the other half of the original 32 vocabulary items no new word pairs were learned (noninterference control condition).

The participants were told that this new experiment involved learning pairs of English words. Pairs of English words would appear on the screen and the participants were to study each pair in any way they wished so that if they later saw the first word of the pair they would be able to recall the second word. Participants were informed that they might encounter some words that had appeared in the first experiment and that they should simply learn these words in the new context.

For the interference acquisition procedure, each pair was presented individually on the computer screen with the first word (the keyword) in the center of the screen and the second word (the new English word) to the lower right. Study of the word pairs was self-paced with a maximum study time of 20 s before the next pair was presented. Participants thought aloud while studying each pair.

After learning the word pairs, participants went through a dropout procedure to ensure they had learned all the word pairs. For each trial, the first word of a pair appeared in the center of the screen, and a question mark appeared in the lower right of the screen where the second word had appeared previously. Participants responded by saying aloud the second member of the pair as soon as they thought of it. After an incorrect response, feedback (consisting of the original pair of words) was displayed for 10 s. All other aspects of the procedure were identical to the previously described dropout phase for the Spanish–English vocabulary items. No verbal reports were collected during this phase of the experiment. Following the dropout procedure, participants received 40 additional blocks of testing on the new word pairs using exactly the same testing procedure as for dropout, except that the feedback consisted of only the correct response presented in the center of the computer screen for 3 s.

Following this interference phase, participants were tested on the original 32 Spanish–English vocabulary items with the vocabulary retrieval

**Method**

**Participants.** The participants were eight undergraduate college students from the University of Colorado recruited through an ad placed in the campus newspaper. They were paid for their participation in the experiment. The experiment was restricted to individuals who had no previous training in Spanish.

**Materials, apparatus, and design.** The computer, recording, and voice key apparatus were the same as those used in Experiment 3. The stimuli were the same 32 Spanish–English pairs and associated keywords used in Experiment 3 (e.g., cabra–cab–goat). In addition, 32 new concrete English nouns were selected from the Paivio, Yuille, and Madigan (1968) list using the same criteria using to select the original keywords and English words (i.e., all words had concreteness, imagery, and meaningfulness ratings of approximately 6.0 or greater). The complete list of stimuli appears in Appendix D. The 32 new nouns were randomly paired with the old keywords to form paired associates used during the interference phase of the experiment (e.g., the keyword cab was paired with hospital to form cab–hospital). The design of the experiment was completely within-subjects, with retrieval task (vocabulary vs. English subtask), practice condition (full vs. subtask), and interference condition (interference vs. noninterference) as within-subjects factors. Assignment of items to interference condition, practice condition, and task order (vocabulary–subtask or subtask–vocabulary) was randomized and counterbalanced across participants.

**Procedure.** The general procedure was the same used in Experiment 3 through the final test of the vocabulary and English tasks, with a few minor modifications. First, there were 6 sessions instead of 10 scheduled over 6 days instead of over 2½ weeks. Session 1 consisted of acquisition, dropout, and initial testing, as in Experiment 3, and the first 20 blocks of retrieval practice, followed by Sessions 2–5, which consisted of 15 blocks of practice each. For the final session, participants were first tested on the items with the vocabulary task and the English subtask. Verbal reports were collected on all trials during acquisition and testing.

As soon as the final vocabulary–English test was over, participants were given a break and were told that the first experiment was over and that they were now going to participate in a new experiment. In fact, this "new experiment" was simply the interference phase of the experiment, in which for half of the original 32 vocabulary items, participants learned paired associates (interference condition) where the first member of the pair was a previously used keyword and the second member was a new English word (e.g., for cabra–cab–goat the new keyword–English pair was cab–hospital). For the other half of the original 32 vocabulary items no new word pairs were learned (noninterference control condition).

The participants were told that this new experiment involved learning pairs of English words. Pairs of English words would appear on the screen and the participants were to study each pair in any way they wished so that if they later saw the first word of the pair they would be able to recall the second word. Participants were informed that they might encounter some words that had appeared in the first experiment and that they should simply learn these words in the new context.

For the interference acquisition procedure, each pair was presented individually on the computer screen with the first word (the keyword) in the center of the screen and the second word (the new English word) to the lower right. Study of the word pairs was self-paced with a maximum study time of 20 s before the next pair was presented. Participants thought aloud while studying each pair.

After learning the word pairs, participants went through a dropout procedure to ensure they had learned all the word pairs. For each trial, the first word of a pair appeared in the center of the screen, and a question mark appeared in the lower right of the screen where the second word had appeared previously. Participants responded by saying aloud the second member of the pair as soon as they thought of it. After an incorrect response, feedback (consisting of the original pair of words) was displayed for 10 s. All other aspects of the procedure were identical to the previously described dropout phase for the Spanish–English vocabulary items. No verbal reports were collected during this phase of the experiment. Following the dropout procedure, participants received 40 additional blocks of testing on the new word pairs using exactly the same testing procedure as for dropout, except that the feedback consisted of only the correct response presented in the center of the computer screen for 3 s.

Following this interference phase, participants were tested on the original 32 Spanish–English vocabulary items with the vocabulary retrieval
task only. There were two blocks of testing. Retrospective verbal reports were collected after each trial.

Results

The analyses reported here are restricted to the vocabulary task and English subtask latencies at initial and final test and to the postinterference phase vocabulary task latencies. For the initial and final tests, voice key errors occurred on 1.4% of the trials and were excluded from the analyses. For the postinterference vocabulary retrieval test, voice key errors occurred on only .8% of the trials and were again excluded from the analyses. The recall results for the vocabulary task and English subtask at initial test were .97 and .98 for the subtask-practiced items and .96 and 1.00 for the full-practiced items. The corresponding scores on the final test were .97 and .99 for the subtask-practiced items and 1.00 and 1.00 for the full-practiced items. Recall results for the vocabulary task following the interference phase were as follows: following subtask practice, .98 versus .98 for the noninterference and interference conditions; following full practice, 1.00 versus .99 for the noninterference and interference conditions. Given the ceiling effects for all these tests, no further analyses of the recall results were performed.

Retrieval task latencies for the initial and final tests as well as for the postinterference vocabulary test were transformed using the same procedure described previously. This transformation resulted in the loss of less than 1% of the trials for the initial and final tests and less than 1% of the trials for the postinterference vocabulary test. The latency results for the initial and final tests are presented in Figure 7. The pattern of latencies was identical to that found in Experiment 3, with vocabulary task latencies greater than English subtask latencies at initial test, but at final test, the pattern of latencies depended on the type of prior practice. For the subtask-practice items, vocabulary task latencies were still greater than the English subtask latencies; but for the full-practice items, the vocabulary task latencies were less than those for the English subtask. The statistical analysis showed exactly the same main effects of test occasion, \( F(1, 7) = 32.55, p < .001, MSE = .06; \) retrieval task, \( F(1, 7) = 75.83, p < .0001, MSE = .01; \) and practice condition, \( F(1, 7) = 7.56, p < .05, MSE = .01. \) Once again, there was a Retrieval Task \( \times \) Practice Condition \( \times \) Test Occasion interaction, \( F(1, 7) = 274.67, p < .0001, MSE = .01, \) as well as a Retrieval Task \( \times \) Practice Condition interaction, \( F(1, 7) = 63.44, p < .0001, MSE = .01, \) and a Practice Condition \( \times \) Test Occasion interaction, \( F(1, 7) = 27.92, p < .001, MSE = .01. \) No other main effects or interactions were significant.

Given the triple interaction of retrieval task, practice condition, and test occasion, further analyses of the simple effects for the initial and final tests were performed. For the ANOVA on the initial test, the results were the same as those for Experiment 3, with only a main effect of retrieval task: Vocabulary task latencies were greater than the English subtask latencies, \( F(1, 7) = 65.11, p < .0001, MSE = .01. \) For the ANOVA on the final test there was a significant Retrieval Task \( \times \) Practice Condition interaction, \( F(1, 7) = 133.64, p < .0001, MSE = .01. \) Further comparisons confirmed that vocabulary task latencies were longer than English subtask latencies for the subtask-practice items (1,545 vs. 903 ms), \( F(1, 7) = 121.12, p < .0001, MSE = .01, \) but for the full-practice items, the reverse was the case (934 vs. 1,127 ms), \( F(1, 7) = 54.05, p < .0001, MSE = .01. \)

The postinterference vocabulary task latencies were faster for the items receiving full practice than for the subtask-practice items, \( F(1, 7) = 41.20, p < .0001, MSE = .01. \) More importantly, the latencies for the interference items were longer than those for the noninterference items (1,311 vs. 1,070 ms), \( F(1, 7) = 25.91, p < .001, MSE = .01, \) but the Practice Condition \( \times \) Interference Condition interaction was not reliable, \( F(1, 7) = 2.40, p < .166, MSE = .01. \) Additional comparisons confirmed that latencies for the interference items were greater than those for the noninterference items, both for the subtask-practice condition, \( F(1, 7) = 12.84, p = .009, MSE = .02, \) and the full-practice condition, \( F(1, 7) = 10.43, p = .014, MSE = .01. \)

Discussion

The pattern of results is clear-cut. First, as in Experiment 3, the results for the vocabulary task and English subtask analysis at final test are not consistent with a working memory mediation model but are consistent with either the covert mediation model or the direct model to explain direct access. However, the results for the postinterference vocabulary task latencies clearly support the mediation model over the direct model. If the keyword played no role in the retrieval process, as the direct model suggests, then the interference manipulation should not have influenced the full-practiced items. This was not the case: The interference manipulation significantly slowed retrieval times for the full-practiced items and the subtask-practiced items, as predicted by the mediation model. Therefore, it appears that keyword mediators continue to influence retrieval of the vocabulary definitions despite extended retrieval practice of the Spanish–English pairs.

General Discussion

The results of the four experiments reported here offer several important findings about the role of prior knowledge in memory
The results clearly show that the knowledge used in learning new material can play an important functional role in subsequent access of that material from memory. Following initial mastery of the vocabulary items using the keyword method, we found converging evidence for explicit retrieval of the keyword mediators in recalling the English equivalent of the Spanish word. First, success in recalling the English definitions of the Spanish words consistently related to performance on the keyword and English subtasks. When memory for the vocabulary items decreased after a no-practice delay, participants were able to retrieve the translation from the Spanish word only if they were able to carry out both the keyword and English subtasks. On the other hand, when participants were unable to retrieve the translation from the Spanish word they almost invariably made errors on one of the subtasks. Second, prior to extended practice, the time required to perform the vocabulary task always exceeded the time required to execute either of the subtasks. This difference was typically large (around 200–500 ms), consistent with the time necessary to complete one or more retrieval operations in working memory (see Ericsson & Kintsch, 1995, for a recent review). Finally, participants’ retrospective reports for the vocabulary retrieval task mentioned keyword mediators in retrieving the English equivalents of the Spanish words, and the content of these reports was validated by the 500 ms latency difference between the mediated and unmediated reports. These latency differences were again consistent with current estimates of the time necessary to perform one or more retrieval operations in working memory. It is worth highlighting the close correspondence of these two very different estimates of the mediational retrieval operations: the decomposition analysis comparing the vocabulary retrieval task to the subtasks and the latency analysis comparing vocabulary retrieval task trials reported to be mediated or unmediated. In summary, the pattern of results in four different experiments strongly argues that the initial retrieval process consists of a sequence of individual working memory steps: (a) retrieving the keyword mediator into working memory and then (b) using the keyword as a retrieval cue to access the memory structure containing the English equivalent.

The second major finding here concerns the changes in the retrieval process with practice. The most obvious change was the decrease in vocabulary retrieval task latencies, consistent with findings from other studies that have found improvements in performance with practice (e.g., Corbett, 1977; Levin et al., 1983). Three sources of data provide a unique perspective on assessing the structure of the speed-up in retrieving the vocabulary pairs: the component subtask latencies, the retrospective verbal reports for the vocabulary task, and the interference latency results. With moderate practice (Experiment 2), retrieval of the vocabulary pairs was still explicitly mediated by retrieval of keyword mediators, evidenced by the longer latencies for the vocabulary task compared to the subtasks as well as by the facilitative effect of vocabulary retrieval practice on the unpracticed subtasks and vice versa. This interpretation is supported by the retrospective reports for the vocabulary task. After moderate practice on the vocabulary task, participants continued to report keyword mediators on the majority of vocabulary retrievals.

However, this pattern of results changed completely when the amount of practice was increased by an order of magnitude in Experiments 3 and 4. After extended practice retrieving vocabulary pairs, participants were able to retrieve the English equivalent from the Spanish word faster than they were able to retrieve the English equivalent from the keyword, a result that is inconsistent with a retrieval process in which keyword mediators are explicitly retrieved into working memory and used to access the English equivalents. Instead, this result suggests the English equivalent was accessed directly from the Spanish word. Again, the verbal report analysis supports this interpretation. After 80 blocks of practice retrieving the Spanish–English pairs, participants reported almost no mediation at all. Their verbal reports referred only to perceiving the Spanish word and then thinking of the English translation without any intermediate processing steps (e.g., “I saw perro and thought dog”). In summary, there is strong evidence that a controlled mediation process in which mediators must first be retrieved into working memory and then used as explicit cues to retrieve the English equivalent has changed to a process in which the English equivalent is directly accessed from the Spanish word with no intermediate working memory steps.

The nature of this direct access is informed by the interference latency results in Experiment 4, which showed that when new responses to the original keyword mediators were learned, retrieval times for the Spanish–English vocabulary pairs declined. This important theoretical finding, along with the others, argues strongly that direct access continues to rely on the original mediators used to learn the vocabulary pairs.

The present findings have implications for research on memory and mnemonics as well as general issues related to change processes in cognition. First, these results have important implications for research on the functional role of mediation in memory processes. Proponents of the mediation hypothesis have argued that a contingency between recall of target material learned using mediators and recall of the mediators themselves constituted evidence for the role of mediation in memory. However, proponents of a direct hypothesis could still argue that recall of the mediator and target material were merely correlated: Both mediator and target were recalled from the stimulus, or the mediator was recalled from the target itself. The current results clearly rule out these accounts and firmly establish the role of mediation in memory retrieval processes when material is learned using simple mediators. Furthermore, these results strongly imply that such mediators continue to play a role in subsequent memory retrieval processes even when we are no longer aware of any mediating steps. This is an important theoretical finding because it suggests the continued influence of knowledge used at encoding on subsequent memory retrieval processes. In one sense, this finding could be viewed as yet another validation of the encoding specificity principle.

A second implication of our results is methodological. Previous keyword researchers have suggested the need for more direct
process measures in studying the keyword method and other mediation techniques (e.g., Pressley, Levin, & Delaney, 1982). The decomposition approach combined with trial by trial think-aloud and retrospective verbal reports, as well as the interference paradigm, constitute a powerful set of methods for studying underlying memory processes involving mediators. Moreover, these measures provide a means of assessing the role of mediators at different stages of the retrieval process before and after different amounts of practice.

A third implication of the current results concerns the mnemonic dependency issue raised in the introduction. Some have suggested that information learned using a mnemonic might never be remembered as an independent entity because it is permanently linked to the originally applied mnemonic device (see Higbee, 1978, and Levin & Pressley, 1985, for discussions). In other words, to recall that *perro* is *dog* an individual would forever have to recall the *pear* associated with *perro* and then the interactive image of "a dog eating a pear." The current results demonstrate that material learned using keyword mediators depends on the mediators initially: In fact, the keyword mediators serve as intermediate retrieval cues to access the English translations. However, this dependency changes with practice, as direct retrieval of the English translations of the foreign words becomes possible. After extended practice, retrieval of the vocabulary pairs is automatic, with participants not relying on the keyword mediators to cue recall of the translations; in fact, participants report no awareness at all of the keyword mediators or interactive images. Nevertheless, the mediators appear to remain covertly engaged in retrieval.

The mnemonic dependency issue relates to an important general theoretical issue in cognitive research: how processes and structures change as a function of practice (e.g., Strayer & Kramer, 1990). Does practice simply strengthen existing cognitive structures or processes, or are there qualitative changes in these processes or structures (Cheng, 1985)? A number of recent research efforts have focused on this issue (e.g., Ashcraft, 1982; Carlson, Sullivan, & Schneider, 1989; Logan, 1988, 1991, 1992; MacLeod & Dunbar, 1988; Neves & Anderson, 1981; Schneider, 1985; Schneider & Shiffrin, 1985; Siegler, 1987, 1988). However, most of this research has focused on skill acquisition. Relatively few studies have examined change mechanisms in memory processes over extended practice intervals. Two exceptions are the studies by Hayes-Roth (1977) and Pirolli and Anderson (1985) that examined changes in retrieval of facts from memory. These studies used a fact-fan paradigm in which new facts are learned that reuse concepts. Hayes-Roth found that RTs to targets increased as a function of the number of targets that were related to a target subject. However, with continued fact retrieval practice, this fan effect disappeared and participants were able to recognize all facts equally fast despite the number of related propositions, which Hayes-Roth interpreted as evidence for a qualitatively different retrieval process after extended practice. In a similar series of studies, Pirolli and Anderson found just the opposite: Fan effects persisted despite massive amounts of practice, and from this result they argued that there was no change in the nature of the retrieval process. In addition, Pirolli and Anderson found that these fan effects decreased in direct proportion to the amount of retrieval speed-up and argued that it is difficult to reconcile such a gradual improvement process with the kind of qualitative changes in the memory process suggested by Hayes-Roth. In summary, although Pirolli and Anderson found that interference effects were reduced with practice, the effects did not disappear completely, implying no qualitative change in the retrieval process but simple speed-up. The vocabulary mediation findings reported here suggest both quantitative and qualitative changes in memory retrieval processes. Moderate practice produced only quantitative changes in retrieval: faster execution of the same mediating steps. However, extended practice changed the process qualitatively, with the separate working memory steps of retrieving the mediator and using it to retrieve the English equivalent replaced by a single retrieval step. However, the initial mediating cognitive structures appear to remain engaged in direct retrieval.

Understanding the mechanisms underlying changes in memory process is important given current accounts of automaticity and how it develops. These accounts explain automaticity as the result of qualitative changes in how tasks are carried out—general procedures or algorithmic processing being replaced by memory based processing (e.g., Logan, 1988). Research on acquisition of arithmetic (Ashcraft, 1982; Siegler, 1987, 1988), for example, suggests that solving problems initially relies on general counting strategies that guarantee correct answers; however, with sufficient practice the sums of all digit pairs are memorized rather than computed. One of the leading accounts of automaticity as memory (Logan, 1988, 1991, 1992; Logan & Klapp, 1991) proposes that with repeated performance of the same tasks, the sequential generation of intermediate steps is completely replaced by direct retrieval of the answers based on memory of the previously generated answer. Given the importance of theories of automaticity in understanding skilled human performance and the role that memory retrieval may play in explaining automaticity, it is important to understand how memory retrieval itself changes and what "direct access" is.

To summarize, the studies reported here provide detailed process evidence for the functional role of prior knowledge in mediating the encoding and retrieval of material from memory. Specifically, retrieval of Spanish–English vocabulary pairs learned using keyword mediators is accomplished initially by a controlled mediation process that requires first retrieving the mediator and then using it as an explicit cue to retrieve the English equivalent. However, following extended retrieval practice, this controlled mediation process changes to direct access in which the English equivalent is retrieved directly from the Spanish word in a single working memory step. Finally, the results suggest a new theoretically interesting possibility that this direct access still depends on the original mediating knowledge engaged at initial encoding.

References


(Appendices follow)
### Appendix A

#### Stimuli for Experiment 1

<table>
<thead>
<tr>
<th>Spanish word</th>
<th>Keyword</th>
<th>English equivalent</th>
<th>Spanish word</th>
<th>Keyword</th>
<th>English equivalent</th>
</tr>
</thead>
<tbody>
<tr>
<td>babucha</td>
<td>baby</td>
<td>slipper</td>
<td>garra</td>
<td>garage</td>
<td>claw</td>
</tr>
<tr>
<td>balde</td>
<td>ball</td>
<td>bucket</td>
<td>lrowno</td>
<td>horn</td>
<td>oven</td>
</tr>
<tr>
<td>banjero</td>
<td>banjo</td>
<td>tray</td>
<td>juguetes</td>
<td>jug</td>
<td>toy</td>
</tr>
<tr>
<td>bandeja</td>
<td>band</td>
<td>flag</td>
<td>manchar</td>
<td>man</td>
<td>stain</td>
</tr>
<tr>
<td>barca</td>
<td>bark</td>
<td>boat</td>
<td>maton</td>
<td>mat</td>
<td>bully</td>
</tr>
<tr>
<td>beso</td>
<td>baseball</td>
<td>kiss</td>
<td>pantano</td>
<td>pants</td>
<td>swamp</td>
</tr>
<tr>
<td>bombero</td>
<td>bomb</td>
<td>fireman</td>
<td>panoelo</td>
<td>pan</td>
<td>handkerchief</td>
</tr>
<tr>
<td>borsaca</td>
<td>board</td>
<td>storm</td>
<td>pavo</td>
<td>paw</td>
<td>turkey</td>
</tr>
<tr>
<td>cabra</td>
<td>cab</td>
<td>goat</td>
<td>pena</td>
<td>pen</td>
<td>rock</td>
</tr>
<tr>
<td>campana</td>
<td>camp</td>
<td>bell</td>
<td>perro</td>
<td>pear</td>
<td>dog</td>
</tr>
<tr>
<td>candada</td>
<td>candle</td>
<td>padlock</td>
<td>pinzas</td>
<td>pizza</td>
<td>tweezers</td>
</tr>
<tr>
<td>carta</td>
<td>cart</td>
<td>letter</td>
<td>pollo</td>
<td>pole</td>
<td>chicken</td>
</tr>
<tr>
<td>charco</td>
<td>charcoal</td>
<td>puddle</td>
<td>reina</td>
<td>rain</td>
<td>queen</td>
</tr>
<tr>
<td>cielo</td>
<td>ceiling</td>
<td>sky</td>
<td>saco</td>
<td>sack</td>
<td>jacket</td>
</tr>
<tr>
<td>cocina</td>
<td>coke</td>
<td>elbow</td>
<td>sonrisa</td>
<td>smile</td>
<td></td>
</tr>
<tr>
<td>codo</td>
<td>code</td>
<td>twilight</td>
<td>soplete</td>
<td>soap</td>
<td>blowtorch</td>
</tr>
<tr>
<td>crepusculo</td>
<td>crepes</td>
<td></td>
<td>tacon</td>
<td>tack</td>
<td></td>
</tr>
<tr>
<td>dieronico</td>
<td>door</td>
<td>leopard</td>
<td>tarima</td>
<td>tar</td>
<td>platform</td>
</tr>
<tr>
<td>fingido</td>
<td>finger</td>
<td>dummy</td>
<td>tenedor</td>
<td>ten</td>
<td>fork</td>
</tr>
<tr>
<td>flecha</td>
<td>flesh</td>
<td>arrow</td>
<td>tierra</td>
<td>tea</td>
<td>dirt</td>
</tr>
<tr>
<td>forastero</td>
<td>forest</td>
<td>outsider</td>
<td>tonel</td>
<td>tunnel</td>
<td>keg</td>
</tr>
</tbody>
</table>

### Appendix B

#### Stimuli for Experiment 2

<table>
<thead>
<tr>
<th>Spanish word</th>
<th>Keyword</th>
<th>English equivalent</th>
<th>Spanish word</th>
<th>Keyword</th>
<th>English equivalent</th>
</tr>
</thead>
<tbody>
<tr>
<td>babucha</td>
<td>baby</td>
<td>slipper</td>
<td>lrowno</td>
<td>horn</td>
<td>oven</td>
</tr>
<tr>
<td>balde</td>
<td>ball</td>
<td>bucket</td>
<td>juguetes</td>
<td>jug</td>
<td>toy</td>
</tr>
<tr>
<td>banjero</td>
<td>banjo</td>
<td>tray</td>
<td>manchar</td>
<td>man</td>
<td>stain</td>
</tr>
<tr>
<td>barca</td>
<td>bark</td>
<td>boat</td>
<td>maton</td>
<td>mat</td>
<td>bully</td>
</tr>
<tr>
<td>bolsa</td>
<td>bowl</td>
<td>purse</td>
<td>muneca</td>
<td>money</td>
<td></td>
</tr>
<tr>
<td>bombero</td>
<td>bomb</td>
<td>fireman</td>
<td>pantano</td>
<td>pants</td>
<td>swamp</td>
</tr>
<tr>
<td>borrasca</td>
<td>board</td>
<td>storm</td>
<td>panoelo</td>
<td>pan</td>
<td>handkerchief</td>
</tr>
<tr>
<td>butaca</td>
<td>boot</td>
<td>armchair</td>
<td>pavo</td>
<td>paw</td>
<td>turkey</td>
</tr>
<tr>
<td>cabra</td>
<td>cab</td>
<td>goat</td>
<td>pena</td>
<td>pen</td>
<td>rock</td>
</tr>
<tr>
<td>campana</td>
<td>camp</td>
<td>bell</td>
<td>perro</td>
<td>pear</td>
<td>dog</td>
</tr>
<tr>
<td>candada</td>
<td>candle</td>
<td>padlock</td>
<td>pinzas</td>
<td>pizza</td>
<td>tweezers</td>
</tr>
<tr>
<td>carta</td>
<td>cart</td>
<td>letter</td>
<td>pollo</td>
<td>pole</td>
<td>chicken</td>
</tr>
<tr>
<td>charco</td>
<td>charcoal</td>
<td>puddle</td>
<td>reina</td>
<td>rain</td>
<td>queen</td>
</tr>
<tr>
<td>cielo</td>
<td>ceiling</td>
<td>sky</td>
<td>ropa</td>
<td>rope</td>
<td>clothing</td>
</tr>
<tr>
<td>cocina</td>
<td>coke</td>
<td>cuisine</td>
<td>saco</td>
<td>sack</td>
<td>jacket</td>
</tr>
<tr>
<td>codo</td>
<td>code</td>
<td>elbow</td>
<td>sonrisa</td>
<td>sun</td>
<td>smile</td>
</tr>
<tr>
<td>crepusculo</td>
<td>crepes</td>
<td>twilight</td>
<td>soplete</td>
<td>soap</td>
<td>blowtorch</td>
</tr>
<tr>
<td>dador</td>
<td>dad</td>
<td>dice</td>
<td>tacon</td>
<td>tack</td>
<td></td>
</tr>
<tr>
<td>dieronico</td>
<td>door</td>
<td>leopard</td>
<td>tarima</td>
<td>tar</td>
<td>platform</td>
</tr>
<tr>
<td>fingido</td>
<td>finger</td>
<td>dummy</td>
<td>tenedor</td>
<td>ten</td>
<td>fork</td>
</tr>
<tr>
<td>flecha</td>
<td>flesh</td>
<td>arrow</td>
<td>tierra</td>
<td>tea</td>
<td>dirt</td>
</tr>
<tr>
<td>forastero</td>
<td>forest</td>
<td>outsider</td>
<td>tonel</td>
<td>tunnel</td>
<td>keg</td>
</tr>
<tr>
<td>garra</td>
<td>garage</td>
<td>claw</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
### Appendix C

#### Stimuli for Experiment 3

<table>
<thead>
<tr>
<th>Spanish word</th>
<th>Keyword</th>
<th>English equivalent</th>
<th>Spanish word</th>
<th>Keyword</th>
<th>English equivalent</th>
</tr>
</thead>
<tbody>
<tr>
<td>balde</td>
<td>ball</td>
<td>bucket</td>
<td>horno</td>
<td>horn</td>
<td>oven</td>
</tr>
<tr>
<td>barca</td>
<td>bark</td>
<td>boat</td>
<td>jugueteca</td>
<td>jug</td>
<td>toy</td>
</tr>
<tr>
<td>bolsa</td>
<td>bowl</td>
<td>purse</td>
<td>manchar</td>
<td>man</td>
<td>stain</td>
</tr>
<tr>
<td>bombero</td>
<td>bomb</td>
<td>fireman</td>
<td>maton</td>
<td>mat</td>
<td>bully</td>
</tr>
<tr>
<td>cabra</td>
<td>cab</td>
<td>bell</td>
<td>muneca</td>
<td>money</td>
<td>doll</td>
</tr>
<tr>
<td>campana</td>
<td>camp</td>
<td>padlock</td>
<td>panselo</td>
<td>pan</td>
<td>handkerchief</td>
</tr>
<tr>
<td>candada</td>
<td>candle</td>
<td>letter</td>
<td>pavo</td>
<td>paw</td>
<td>turkey</td>
</tr>
<tr>
<td>carta</td>
<td>cart</td>
<td>padlock</td>
<td>pena</td>
<td>pen</td>
<td>rock</td>
</tr>
<tr>
<td>charco</td>
<td>charcoal</td>
<td>paddle</td>
<td>perro</td>
<td>pear</td>
<td>dog</td>
</tr>
<tr>
<td>codo</td>
<td>code</td>
<td>elbow</td>
<td>reina</td>
<td>rain</td>
<td>queen</td>
</tr>
<tr>
<td>dados</td>
<td>dad</td>
<td>dices</td>
<td>saco</td>
<td>sack</td>
<td>jacket</td>
</tr>
<tr>
<td>doronico</td>
<td>door</td>
<td>leopard</td>
<td>soplete</td>
<td>soap</td>
<td>blowtorch</td>
</tr>
<tr>
<td>fingido</td>
<td>finger</td>
<td>dummy</td>
<td>tacon</td>
<td>tack</td>
<td>heel</td>
</tr>
<tr>
<td>flecha</td>
<td>flesh</td>
<td>arrow</td>
<td>tarima</td>
<td>tar</td>
<td>platform</td>
</tr>
<tr>
<td>forastero</td>
<td>forest</td>
<td>outsider</td>
<td>tenedor</td>
<td>ten</td>
<td>fork</td>
</tr>
<tr>
<td>garra</td>
<td>garage</td>
<td>claw</td>
<td>tonel</td>
<td>tunnel</td>
<td>keg</td>
</tr>
</tbody>
</table>

### Appendix D

#### Stimuli for Experiment 4

<table>
<thead>
<tr>
<th>Spanish word</th>
<th>Keyword</th>
<th>English equivalent</th>
<th>New English equivalent</th>
<th>Spanish word</th>
<th>Keyword</th>
<th>English equivalent</th>
<th>New English equivalent</th>
</tr>
</thead>
<tbody>
<tr>
<td>balde</td>
<td>ball</td>
<td>bucket</td>
<td>flower</td>
<td>horno</td>
<td>horn</td>
<td>oven</td>
<td>priest</td>
</tr>
<tr>
<td>barca</td>
<td>bark</td>
<td>boat</td>
<td>umbrella</td>
<td>jugueteca</td>
<td>jug</td>
<td>toy</td>
<td>hammer</td>
</tr>
<tr>
<td>bolsa</td>
<td>bowl</td>
<td>purse</td>
<td>vest</td>
<td>manchar</td>
<td>man</td>
<td>stain</td>
<td>diamond</td>
</tr>
<tr>
<td>bombero</td>
<td>bomb</td>
<td>fireman</td>
<td>sugar</td>
<td>maton</td>
<td>mat</td>
<td>bully</td>
<td>candy</td>
</tr>
<tr>
<td>cabra</td>
<td>cab</td>
<td>dices</td>
<td>hospital</td>
<td>muneca</td>
<td>money</td>
<td>doll</td>
<td>butterfly</td>
</tr>
<tr>
<td>campana</td>
<td>camp</td>
<td>bell</td>
<td>girl</td>
<td>panselo</td>
<td>pan</td>
<td>handkerchief</td>
<td>book</td>
</tr>
<tr>
<td>candada</td>
<td>candle</td>
<td>padlock</td>
<td>tree</td>
<td>pavo</td>
<td>paw</td>
<td>turkey</td>
<td>rattle</td>
</tr>
<tr>
<td>carta</td>
<td>cart</td>
<td>letter</td>
<td>mountain</td>
<td>pena</td>
<td>pen</td>
<td>rock</td>
<td>flag</td>
</tr>
<tr>
<td>charco</td>
<td>charcoal</td>
<td>paddle</td>
<td>coffee</td>
<td>perro</td>
<td>pear</td>
<td>dog</td>
<td>cotton</td>
</tr>
<tr>
<td>codo</td>
<td>code</td>
<td>elbow</td>
<td>bird</td>
<td>reina</td>
<td>rain</td>
<td>queen</td>
<td>pipe</td>
</tr>
<tr>
<td>dados</td>
<td>dad</td>
<td>dices</td>
<td>potato</td>
<td>saco</td>
<td>sack</td>
<td>jacket</td>
<td>ticket</td>
</tr>
<tr>
<td>doronico</td>
<td>door</td>
<td>leopard</td>
<td>judge</td>
<td>soplete</td>
<td>soap</td>
<td>blowtorch</td>
<td>clock</td>
</tr>
<tr>
<td>fingido</td>
<td>finger</td>
<td>dummy</td>
<td>toast</td>
<td>tacon</td>
<td>tack</td>
<td>heel</td>
<td>pepper</td>
</tr>
<tr>
<td>flecha</td>
<td>flesh</td>
<td>arrow</td>
<td>truck</td>
<td>tarima</td>
<td>tar</td>
<td>platform</td>
<td>university</td>
</tr>
<tr>
<td>forastero</td>
<td>forest</td>
<td>outsider</td>
<td>poster</td>
<td>tenedor</td>
<td>ten</td>
<td>fork</td>
<td>beggar</td>
</tr>
<tr>
<td>garra</td>
<td>garage</td>
<td>claw</td>
<td>river</td>
<td>tonel</td>
<td>tunnel</td>
<td>keg</td>
<td>factory</td>
</tr>
</tbody>
</table>

Received July 3, 1998
Revision received December 8, 1999
Accepted March 23, 2000