Journal of Experimental Psychology:
Learning, Memory, and Cognition
2006, Vol. 32, No. 4, 734–748

0278-7393/06/$12.00 DOI: 10.1037/0278-7393.32.4.734

Cued Recall From Image and Sentence Memory: A Shift From Episodic to Identical Elements Representation

Timothy C. Rickard and Daniel Bajic
University of California, San Diego

The applicability of the identical elements (IE) model of arithmetic fact retrieval (T. C. Rickard, A. F. Healy, & L. E. Bourne, 1994) to cued recall from episodic (image and sentence) memory was explored in 3 transfer experiments. In agreement with results from arithmetic, speedup following even minimal practice recalling a missing word from an episodically bound word triplet did not transfer positively to other cued recall items involving the same triplet. The shape of the learning curve further supported a shift from episode-based to IE-based recall, extending some models of skill learning to cued recall practice. In contrast with previous findings, these results indicate that a form of representation that is independent of the original episodic memory underlies cued-recall performance following minimal practice.

Keywords: episode, identical elements, cued recall, shift, practice, transfer

Episodic memory plays a crucial role in cued recall of recently acquired knowledge and experiences. If asked what you had for breakfast this morning, your answer likely reflects cued recall from episodic memory. On the other end of the learning curve, highly automatized cued recall (e.g., recall of a highly familiar name) typically involves no subjective experience of episodic retrieval. Identification of the cognitive and neural mechanisms that underlie this transition is fundamental to theory development in the domains of everyday memory, skill acquisition, and automaticity. From a cognitive perspective, this issue can be framed in terms of four basic questions that are addressed in this article: First, what changes do experience (practice) produce in the representations and processes that underlie cued-recall performance? Second, if there is a qualitative shift in representation, does this shift require a high level of practice (automaticity)? Third, if there is a shift, how is the knowledge represented afterward, and how does that representation compare with the original episodic representation? Fourth, are the answers to the questions above domain independent? These questions are addressed through empirical tests of the two candidate models that are described below.

The simplest model, which we term the *holistic strengthening* model, posits that practice produces no major changes in the representations and processes that underlie performance. Instead, it simply improves access to, or strengthens, the originally encoded episode, resulting in speedup and, potentially, even a drop of episodic retrieval from conscious awareness. Closely related ideas are embodied in several recent proposals in the literature (Anderson, Fincham, & Douglass, 1997; Crutcher & Ericsson, 2000; Pirolli & Anderson, 1985; Rabinowitz & Goldberg, 1995). As a concrete example, consider the case of semantically based mental imagery that is studied in Experiment 1. The subject is first presented with a word triplet for study, such as *boy, gift, smile*, under instructions to form an interactive image. On the first cued-recall trial (e.g., *boy gift,* ______), the subject must access this episode to recall the missing element: “smile.” According to the holistic strengthening account, the underlying representation that supports performance does not fundamentally change with additional cued-recall practice. Rather, access to that representation simply becomes more efficient and faster.

In studies conducted to date, positive transfer of learning to altered versions of practice items has been treated as a diagnostic consequence of holistic strengthening (Anderson et al., 1997; Rabinowitz & Goldberg, 1995). Anderson et al. (1997) found that speedup with practice retrieving answers from one set of cues transferred positively to items in which the former responses constituted the cues. Similarly, Rabinowitz and Goldberg (1995) gave subjects practice on an alphabet arithmetic task (e.g., $D + 3 = ?$) in which the answer (for this example) is three letters down the alphabet from the presented letter (i.e., *G*). They concluded in favor of positive transfer to inverted problems (e.g., $G – 3 = ?$), again suggesting that practice strengthens a holistic representation that is symmetrically, and flexibly, accessible. Applied to the current experiments, the model advanced by these researchers predicts that speedup with repetition practice in retrieving the response “smile” when presented with *boy, gift,* ______ should transfer positively and substantially to the task of recalling “boy” when presented with ______, *gift, smile*.

Pirolli and Anderson (1985) used a different approach to address a related issue in the domain of sentence recognition memory. They observed that speedup in mean response time (RT) with repetition practice follows the classic power law (Newell & Rosen-

---

Timothy C. Rickard and Daniel Bajic, Department of Psychology, University of California, San Diego.

This research was supported by National Institute of Mental Health Grant R29 MH58202 to Timothy C. Rickard. We thank Jon Chiu for programming the experiments and Christine Hyjek and Brandon Walker for help with data collection.

Correspondence concerning this article should be addressed to Timothy C. Rickard, Department of Psychology, 0109 University of California, San Diego, 9500 Gilman Drive, La Jolla, CA 92093-0109. E-mail: trickard@ucsd.edu

bloom, 1981), and they argued that power law speedup is most consistent with the hypothesis that practice results only in increased availability of the target sentence memories, with no qualitative change in the underlying representation. Note also that holistic strengthening appears to be consistent with at least some global memory models that accommodate cued recall (e.g., Diller, Nobel, & Shiffrin, 2001), as those models currently have no mechanism that would produce a qualitative change in the nature of representation with cued-recall practice.

In diametric opposition to the holistic strengthening model is the possibility that cued-recall practice results in a shift to a new and independent representation that is somehow tailored to support the specific task at hand (i.e., the particular cued-recall item). The originally encoded episodic representation remains intact, but a new representation develops and provides the basis for recall performance following practice. The possibility of such a shift is generally consistent with theoretical frameworks that specify qualitative shifts in processing with practice in cognitive skill acquisition (Fitts & Posner, 1967; Logan, 1988; Palmeri, 1997; Rickard, 1997, 2004).

Though no model of representational shift for the general case of cued recall has been proposed, Rickard et al. (1994; Rickard, 2005; Rickard & Bourne, 1995, 1996) proposed an identical elements (IE) model of memory-based performance on simple arithmetic problems that may be applicable. Arithmetic fact recall differs from episodic cued recall in many respects, of course, so there is no strong a priori reason to expect a model of arithmetic fact representation to generalize to the current experiments. Nevertheless, the effects of recall practice on representation may well be the same in the two domains, and the IE model makes predictions for the current tasks without modification. The core principle of the model is that cued-recall practice results in an independent representation in memory for each unique combination of stimulus-response elements. The model specifies, for example, that three unique arithmetic facts exist for each number triplet that is related by an arithmetic operation, as depicted in Figure 1 for the number triplet (4, 7, 28). Each fact takes the form of a paired associate.

The IE model predicts no positive transfer of learning between arithmetic items that access different representations in Figure 1. For the triplet 4, 7, 28, for example, speedup with practice recalling the answer to *4 × 7* should not facilitate subsequent performance on *28/7*, or vice versa, because independent representations support performance on those two problems. Furthermore, speedup on *28/7* should not transfer to its inverse problem, *28/4*. For the case of multiplication-to-division and vice versa, these predictions have been confirmed in a series of experiments in which there is either no positive transfer or only slight and transient positive transfer that is best explained by a mediational strategy that is likely idiosyncratic to arithmetic (Campbell, 1997, 1999; Campbell, Fuchs-Lacelle, & Phenix, in press; Hittmair-Delazar, Semenza, & Denes, 1994; LeFevre & Morris, 1999; McCloskey, Aliminosa, & Sokol, 1991; Rickard, 2005; Rickard et al., 1994; Rickard & Bourne, 1995, 1996).

$$(4,7,x) \rightarrow 28$$

$$28 \div 4 \rightarrow 7$$

$$28 \div 7 \rightarrow 4$$

Figure 1. The identical elements (IE) model specifies that three unique arithmetic facts exist for each number triplet that is related by an arithmetic operation, as depicted here for the number triplet (4, 7, 28).

For the case of practicing a division problem and then being tested on its inverse, there is actually negative transfer. Rickard and Bourne (1996) observed that after extensive practice on a problem like *28/4*, subjects had a significantly higher than chance rate of erroneously responding "7" on a subsequent posttest on the complementary problem (e.g., *28/7*). Rickard (2005; Experiment 2) had subjects solve complementary division problems on alternating practice blocks. That is, if a subject solved *28/7* on Blocks 1, 3, 5, and so forth, he or she solved *28/4* on Blocks 2, 4, 6, and so forth. Subjects' division performance was significantly slower on even-numbered than on the immediately preceding odd-numbered blocks, despite the fact the even-numbered blocks had a one-block practice advantage—at the triplet level—relative to each proceeding odd-numbered block.

These negative transfer effects suggest an interrelated network of cued-recall facts that can, under some circumstances at least, be subject to interitem interference. When presented as a stimulus such as *28/4*, the dividend *28* may activate both of the division problem representations, (*28/7*) and (*28/4*), because of element overlap. If one problem, like *28/7*, has been recently practiced, then when the transfer problem *28/4* is presented, the node for *28* might send a large enough activation signal to the problem node (*28/7*) that that node is erroneously selected for retrieval. Alternatively, subjects may be able to suppress that error response, likely resulting in an RT delay, as observed in Rickard (2005). Although the details remain to be worked out, this mechanism is quite similar to that in a number of other network models of retrieval interference. Thus, although the simplest version of the IE model predicts no transfer (in either the positive or negative direction) between items that have different stimulus-response combinations, a finding of negative transfer is consistent with the model and, in fact, would buttress evidence of a shift to an IE-like fact organization (i.e., for the simplest case holistic representation that was defined above and that is descriptive of related ideas as developed in the literature to date, there is no mechanism by which negative transfer could be produced).

Rickard (2005) advanced a simple elaboration of the IE model that allows for positive transfer for the case of pure reversals in arithmetic and other paired-associate tasks. That revised IE model specifies that the associations for the stimulus-response pairings in Figure 1 are bidirectional, and that practicing an item in one direction strengthens the associative link both from the stimulus to the response and from the response back to the stimulus, albeit somewhat weaker strengthening in the latter case. The revised model predicts, for example, that practicing multiplication (e.g., *4 × 7 = ?*) will facilitate subsequent factoring for the same item (e.g., *28 = ? × ?*), and vice versa, and that similar RT transfer effects will occur for more traditional paired-associate tasks. As detailed in Rickard (2005), about 60% of speedup with practice in both types of tasks does indeed transfer to pure reversals. Thus, in the case of arithmetic at least, there is a stark contrast in transfer effects when comparing pure reversals—in which both forward and reverse items access the same paired-associate representation in Figure 1—versus cases in which different representations in Figure 1 would, according to the model, underlie performance on practice and transfer items. To date only the IE model (Rickard, 2005) is able to explain this set of transfer patterns.

The IE model as developed to date is not a model of representational or process shifts themselves, but instead focuses on the organization of facts at high practice levels. However, it implies a shift, because initial representations are not likely to have an IE structure. In arithmetic, for example, students often use counting or adding algorithms during initial learning to do addition and multiplication, respectively (e.g., Siegler, 1988). Children might also have holistic representations of number triplets (e.g., 4, 7, 28) that support arithmetic fact retrieval prior to practice, although there is no evidence for such representations to date.

Applied to the case of semantically based interactive imagery, the IE model predicts a shift with cued-recall practice from episodically based recall, in which a single episodic representation supports retrieval of any missing element, to a unique representation for each practiced stimulus-response combination. That is, whereas study yields a holistic (episodic) representation of the triplet, practice on, say, *boy, smile,* ______ results in an independent IE representation, (*boy, smile*) → *gift*, that supports recall performance for only that particular stimulus-response combination. Once that IE structure has formed, speedup with subsequent practice on *boy, smile,* ______ will not transfer positively to either of the other two stimulus-response combinations, *boy,* ______*, gift* or ______*, smile, gift*. Instead, retrieval of the answer for each of these two combinations requires access to the originally encoded episodic image. If subjects practice simultaneously on all three possible dual-word cued-recall items for a triplet (e.g., *boy, gift,* ______; *gift, smile,* ______; and *boy, smile,* ______), then an independent IE associative structure will develop for each item: (*boy, gift*) → *smile*; (*gift, smile*) → *boy*; and (*boy, smile*) → *gift*. In this case also, further practice on one of these items will not transfer positively to the other two because practice will strengthen only the IE representation corresponding to the practiced items.

The IE model does not specify how much recall practice from a newly formed episodic memory is required before the shift to an IE representation occurs. For adult arithmetic there was extensive real-world practice prior to the laboratory practice-transfer experiments, providing abundant opportunity for a preexperimental shift. In the present experiments, initial coding of episodes occurred at the outset of the roughly 1-hr session, and subjects received 20 practice repetitions on each cued-recall item prior to the final transfer phase. Transfer findings in support of the IE model would thus suggest both that the model generalizes beyond arithmetic, encompassing practice on episodic recall generally, and that the shift to the IE structure can occur with minimal cued-recall practice.

To summarize, the holistic strengthening and IE models yield different answers to the first three research questions that were outlined earlier. The holistic strengthening model assumes that practice produces no change in the representations and processes that underlie cued-recall performance (addressing Question 1), and as such Questions 2 and 3 are not applicable. The IE model, in contrast, assumes that practice does result in pronounced qualitative changes in the representations that underlie performance. It does not address the second question of how much practice is needed for this shift to occur, leaving that as an empirical matter. It does address the third question, however, stating that practice results in a shift from an episodic (holistic) to an item-specific, stimulus-response form of representation. We address the fourth question of model generality by comparing the results for semantically based interactive imagery (Experiments 1 & 2) and sentence memory (Experiments 3) with the earlier results for arithmetic.

The experimental design affords two independent approaches to addressing the representation issue. The first approach involves the transfer manipulation outlined above. The second is analysis of practice speedup curves, following logic analogous to that of a number of previous researchers (Anderson et al., 1997; Delaney, Reder, Staszewski, & Ritter, 1998; Pirolli & Anderson, 1985; Rickard, 1997, 1999, 2004). As we elaborate after discussion of the transfer results, if speedup reflects holistic or other simple strengthening processes, then mean RTs would be expected to follow a log-log-linear function of practice. If speedup reflects a shift to IE-based retrieval, however, systematic deviations from log-log linearity should be observed. Converging evidence from these two contrasting approaches should constitute strong empirical support for the corresponding theory.

Experiment 1

Subjects studied 16 word triplets under interactive imagery instructions, followed by a cued-recall pretest for all possible dual-cue, single-response items that can be formed from those triplets (3 items per triplet by 16 triplets). The pretest phase was designed to establish high accuracy prior to proceeding with the practice phase, and it also provided a prepractice transfer measure, as discussed below. Next a practice-transfer paradigm analogous to that used by Rickard and Bourne (1996) was used. Subjects practiced on one cued-recall item for 8 of the 16 word triplets. They then took a posttest on all 48 cued-recall items, yielding three transfer conditions: practiced items, response-change items (different items drawn from the practiced triplets), and unpracticed items (items from unpracticed triplets). The word triplets were designed to foster interactive imagery. Informal subject reports confirmed that subjects used interactive imagery to encode the great majority of triplets. To facilitate rapid learning, there was no overlap in words used across triplets.

Method

Subjects Thirty University of California at San Diego undergraduate students participated for course credit.

Materials, design, and procedure. Test stimuli consisted of 16 word triplets (see Appendix A). In the study, pretest, and posttest phases of the experiment, all 16 triplets served as stimuli. In the practice phase, only eight triplets were used.

Subjects were tested individually on IBM-compatible personal computers, with each subject seated approximately 50 cm from the computer screen and approximately 3 cm from a microphone. All experiments were programmed using E-Prime software (Psychology Software Tools, Pittsburgh, PA) and the accompanying voice key apparatus (Model 200A). Prior to each phase of the experiment, instructions were presented on the screen and were also read aloud by the experimenter.

The study phase introduced subjects to the word triplets. There was one block, which consisted of one randomly ordered trial for each of the 16 triplets. Each trial proceeded as follows: (a) the screen went blank for 1 s, and (b) there was a 5-s presentation of a triplet, with the words arranged in a column in the center of the screen, such that the middle word was three rows below the uppermost word and the lower word was three rows below the middle one. We also used this arrangement of stimuli in each trial of the following two phases of the experiment. Subjects were instructed to form a mental picture in which the objects corresponding to the words interacted.

In the pretest phase (and the remaining phases), one word in each presented triplet was replaced with a blank underline (i.e., ______). The subject was to recall the missing word and speak it into the microphone. In this phase, each block consisted of 48 trials, such that each triplet appeared in three separate trials, with each word in the triplet being the blanked element for one of these trials. In each successive set of 16 trials, one item (two words and a blank) was presented from each triplet. Each trial proceeded as follows: (a) the screen went blank for 500 ms; (b) the word "Ready" appeared in the center of the screen for 500 ms; (c) the screen went blank for another 500 ms; (d) the test item appeared on the screen, with these stimuli randomly assigned among the top, middle, or bottom positions on the screen on each trial; (e) the subject attempted to recall the word corresponding to the blank and spoke it into the microphone, under instructions to perform the task as quickly as possible while maintaining high accuracy.

After the subject responded and the voice key tripped, the experimenter entered the participant's response and recorded whether the voice key tripped properly. If the subject was in error, the correct response was presented for 3 s. At the close of each block, the screen presented accuracy and mean correct RT feedback. These blocks continued until the subject completed a block with at least 85% accuracy, at which point this phase concluded.

In the practice phase, the basic design of the trials and blocks matched those of pretest phase, with two exceptions. First, only eight of the word triplets were used in this phase. Second, for each triplet, the same word (randomly selected from the three) was blanked out on each practice block. Fifteen subjects received practice on items from Triplets 1–8, and 15 received practice on items from Triplets 9–16 (see Appendix A). For each set of eight triplets, three cued-recall subsets were created such that the blanked-out word for a given triplet was different in each subset, determined randomly. Five subjects received practice on each subset. The practice phase concluded after 20 blocks, with eight trials per block.

The design of the posttest matched that of the pretest, with the following exceptions. First, for the first time in the experiment, triplets were not arranged in a column but, rather, as a downward pointing triangle. The placement of the triplets consisted of two rows: an uppermost row—which consisted of two words (or a word and a blank) separated by the distance of eight blank spaces—and a lower row, which consisted of a single word or a blank centered relative to the upper row. The purpose of this change in stimulus presentation was to minimize perceptual learning during the practice phase as a factor in the transfer results. Locations of the two words and the blanked out space on each trial were again randomly determined. Second, end-of-block feedback on RT and accuracy were no longer provided, though trial-level feedback was still provided. Third, this phase concluded after three blocks.

Results and Discussion

Pretest. Twenty-two subjects reached or exceeded the 85% accuracy criterion on the first pretest block, 6 subjects required two pretest blocks, and 2 subjects required three or four pretest blocks. The mean accuracy rate on the last pretest block was .92, and, as expected, accuracy did not depend on which condition of the posttest an item would later be in, either in this experiment or in the subsequent experiments. Voice key failures occurred on 3.7% of trials. These trials, as well as error trials, were removed prior to RT analyses in all experiments. As for errors, pretest RTs did not depend of which condition of the posttest an item would later be in.

The experimental design affords two independent measures of transfer performance. The first measure addresses the possibility of positive transfer across items of a triplet following only one or two cued-recall repetitions and involves comparison of mean accuracy and RTs for Trials 1–16, 17–32, and 33–48 of the last pretest block. Over the first set of 16 trials, one item was presented from each triplet; over the second set of 16 trials, a second item was presented from each triplet; and over the third set of 16 trials, the third item was presented from each triplet. Thus, if there were enhanced access to a single holistic representation for each triplet with practice, there should be increases in accuracy and/or decreases in mean RT from the first to the third set of trials. On the other hand, if practice effects were entirely item specific from the beginning, then there should be no accuracy improvement or speedup across these sets of trials. Mean accuracy was .91, .92, and .94 on the first, second, and third sets of trials, respectively, suggesting a small amount of accuracy transfer. These accuracy differences were not analyzed for significance, however, because 33 of the 90 mean accuracy proportions (30 subjects by 3 sets) had values of 1.0.

The means of the subject-level mean RTs were 1,982, 1,961, and 1,950 ms for the first, second, and third pretest sets, respectively. In a within-subjects analysis of variance (ANOVA) on the subject-level means, the single factor of set (first, second, or third) did not approach significance, $F(2, 58) < 1.0$, $MSE = 243{,}620$. (The alpha level for all statistical tests is .05). Thus, by this measure at least, there is no evidence that one or two practice repetitions for a triplet yields RT transfer to response-change items (i.e., new items from practiced triplets).

Practice and posttest. Accuracy during the practice phase increased from .96 on the 1st block to .99 on the 20th block. Mean correct RT decreased markedly, from 1,611 ms on the 1st block to 881 ms on the 20th block, following a typical decelerating curve. Mean RT on the first practice block was significantly faster than on the last pretest block, $t(1, 29) = 2.88$, $p < .01$, indicating that substantial learning took place on the pretest trials even though there was no positive transfer across the three sets of pretest trials. These results suggest that the pretest learning, with respect to RT at least, was entirely item specific.

Mean accuracy rates on the posttest, averaged over the three blocks, were high, having values of .994, .963, and .969 for the practiced, response-change, and unpracticed items, respectively. The small difference in accuracy between the response-change and unpracticed conditions was not significant (Wilcoxon's sign-rank test, $p = .16$). Thus, there is no evidence of transfer to response-change items as indexed by accuracy. However, an orthogonal test comparing the average accuracy of the response-change and unpracticed conditions with that of the practice condition was highly significant (sign-rank test, $p < .001$), demonstrating accuracy learning for the practiced items.

The means of the subject-level mean RTs on the posttest are shown in Figure 2 as a function of condition and transfer block.

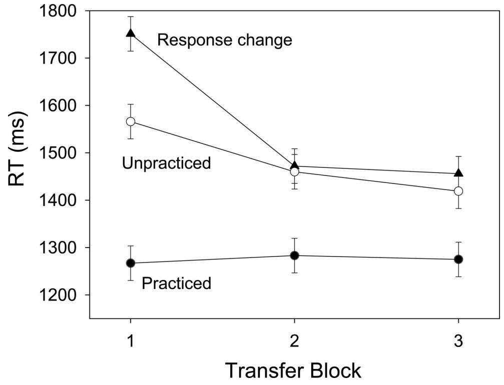

Figure 2. Mean response times (RTs) on the posttest of Experiment 1 as a function of condition and block. Error bars are standard errors based on the interaction error term of a within-subjects analysis of variance on the subject mean RTs.

Here and elsewhere, the same ordinal results were obtained for the mean of the subject median RTs. Performance was clearly fastest in the practiced condition.[1] The slower performance in the response-change than in the unpracticed condition suggests a negative RT transfer effect similar to that observed by Rickard (2005) for complementary division problems. To explore the significance of these patterns, we performed two orthogonal ANOVAs on the subject-level mean RTs. The first compared RTs in the practiced condition with the average RTs in the response-change and unpracticed conditions, using a 2 (transfer condition: unpracticed vs. average of response-change and unpracticed) × 3 (transfer block) factorial design. There were significant effects of block, $F(2, 58) = 3.4$, $p = .041$, $MSE = 49{,}115$; condition, $F(1, 29) = 69.9$, $p < .001$, $MSE = 36{,}423$; and their interaction, $F(2, 58) = 6.2$, $p = .004$, $MSE = 33{,}398$. The second ANOVA compared RTs in the response-change condition with those in the unpracticed condition, again using a 2 (transfer condition: response-change vs. unpracticed) × 3 (transfer block) factorial design. There was a significant effect of block, $F(2, 58) = 8.2$, $p < .001$, $MSE = 105{,}724$; a marginally significant effect of condition, $F(1, 29) = 4.1$, $p = .053$, $MSE = 66{,}787$; and a significant interaction, $F(2, 58) = 3.97$, $p = .024$, $MSE = 33{,}205$.[2]

Note that the decreasing RT differences among the three transfer conditions over blocks reflects a greater opportunity for speedup with practice in the response-change and unpracticed conditions relative to the practiced condition (i.e., speedup generally follows a nonlinear function, and RTs in the response-change and unpracticed conditions were further from asymptote than were RTs in the practice condition). The best estimate of the magnitude of the transfer effects, uncontaminated by the practice on all items over transfer blocks, is thus obtained by inspection of data from the first transfer block.

The above results are inconsistent with the holistic strengthening model outlined in the introduction. On the pretest there was no evidence of RT transfer from Set 1 to Set 3. On the posttest there was no transfer to response-change items as measured by accuracy, and there was negative transfer to those items as measured by RT. The findings are consistent, however, with the IE model. The negative transfer on the posttest can be accommodated by considering the IE representation within a broader associative network of items. Practice in the current experiment would create and strengthen a new and independent IE association for the practiced item, such as (*boy, gift*) → *smile*. When the subject sees a response-change item, such as *boy, smile,* ______ on the posttest, the stimulus word *boy* may partially activate the IE representation that formed with practice, (*boy, gift*) → *smile*. That activation may on some trials lead to a partial or complete retrieval of the practiced response, "smile." This interference could result either in the subject erroneously stating the response "smile" on that transfer trial (analogous to the error patterns discussed earlier for division) or in the subject being delayed in completing correct retrieval of the response "gift" by way of the holistic image. Given the high accuracy in this experiment, the interference would manifest primarily as slowed RTs, as was observed. The lack of negative transfer over trial sets on the pretest can then be understood based on the assumptions that (a) the IE representation may not yet have formed for some items, and (b) the IE representations that have formed would be relatively weak and thus less able to interfere with episodic retrieval.

In this experiment, columnar word order (first, second, or third row) was randomly determined on each trial during the pretest and practice phases, and the format was changed to an inverted triangle during the posttest (with word assignment to spatial position again determined randomly on each trial). This design precluded perceptual learning of the relative spatial locations of the words and the blank from playing a role in posttest performance. It is possible that positive transfer to response-change items would have been observed had the spatial arrangement of the words been held consistent. That is, it is possible the holistic learning and consequent positive transfer is possible for the interactive imagery task but only at a perceptual level of stimulus representation. In Experiment 2 we tested this possibility by holding columnar ordering and spatial arrangement of words constant.

Experiment 2

This experiment was identical to Experiment 1, with the exception that during all phases the words were represented in columnar format and each word (or corresponding blanked space) for a given

---

[1] Note that RTs for the practiced items on the transfer test are several hundred milliseconds slower than for the same items on the last practice block. This finding is routine in this type of design and is generally attributed to contextual interference among the transfer items. In particular, it is likely that the introduction of the more difficult response-change and unpracticed items on the posttest resulted in a global shift in response criterion, resulting in slowed RTs for the practiced items. For discussion, see Rickard et al. (1994).

[2] To investigate the possibility that the differing number of items in each condition on the pretest and the transfer test (8, 16, and 24 in the practiced, response-change, and unpracticed conditions, respectively) had an impact on relative performance in the conditions, we conducted a follow-up experiment with 30 subjects. That experiment was identical to Experiment 1 in all respects except that only 8 of the 16 possible items in the response-change condition and only 8 of the possible 24 items in the unpracticed condition, randomly selected, were presented on the pretest and the transfer test. Transfer results were nearly identical to those of Experiment 1, with significant negative transfer to response-change items on the first transfer block.

triplet was always presented in the same row (first, second, or third). An additional change in this experiment and in Experiment 3 was that subjects were required to achieve 90% accuracy on the pretest before proceeding to the practice phase.

Method

Subjects. Thirty University of California at San Diego undergraduate students participated for course credit.

Materials, design, and procedure. Materials, design, and procedures were identical to those of Experiment 1, with the exceptions noted above.

Results and Discussion

One subject was removed and replaced because of very slow mean RTs on the posttest (> 3,600 ms and far outside the distribution of other subjects) and high error rates. Eleven subjects reached the 90% accuracy criterion on the first pretest block, 13 required two blocks, 5 required three blocks, and 2 required four or five blocks. The mean accuracy rate on the last pretest block was .928.

Pretest. Mean accuracies on the first, second, and third set of 16 trials on the last pretest block were .931, .945, and .907, respectively. The mean correct RTs for these three sets of trials were 1,683, 1,792, and 1,694 ms for the first, second, and third sets of 16 trials, respectively. As in Experiment 1, there was no positive RT transfer, $F(2, 58) = 0.83$, $MSE = 129{,}618$.

Practice and transfer phases. Accuracy during the practice phase increased from .96 on the 1st block to 1.00 on the 20th block. Mean correct RT during the practice phase decreased from 1,308 to 806 ms. Mean RT on the first practice block was again significantly faster than on the last pretest block, $t(1, 29) = 7.34$, $p < .001$.

Mean accuracy rates on the posttest, averaged over the three blocks, were .968, .926, and .928 for the practiced, response-change, and unpracticed items, respectively. As in Experiment 1, the small difference in accuracy between the response-change and unpracticed conditions was not significant (Wilcoxon's sign-rank test, $p = .55$), whereas the orthogonal test comparing the mean accuracy of the response-change and unpracticed conditions to that of the practice condition was highly significant (sign-rank test, $p < .001$).

The means of the subject-level mean RTs on the posttest are shown in Figure 3 as a function of condition and transfer block. RTs were generally faster than in Experiment 1, possibly reflecting faster learning and performance because of the consistent word order and format of presentation throughout the experiment. Despite this evidence of a perceptually specific component of speedup, the transfer results closely mirrored those of Experiment 1. In the ANOVA comparing RTs in the practiced condition to the mean RTs in the response-change and unpracticed conditions (identical in design to that used in Experiment 1), there was no effect of block, $F(2, 58) = 1.96$, $p = .15$, $MSE = 19{,}601$, but there was a strongly significant effect of condition, $F(1, 29) = 51.5$, $p < .001$, $MSE = 60{,}550$, and a significant interaction, $F(2, 58) = 17.5$, $p < .001$, $MSE = 14{,}668$. In the second ANOVA, comparing RTs in the response-change condition to those in the unpracticed condition, there was a significant effect of block, $F(2, 58) = 12.9$, $p < .001$, $MSE = 29{,}845$, but no effects of either condition, $F(1, 29) = 1.4$ $p = .24$, $MSE = 58{,}210$, or the block by condition interaction, $F(2, 58) = 1.2$, $p = .32$, $MSE = 22{,}763$. There were, however, significant effects, as suggested in Figure 4, of both condition, $F(1, 29) = 10.6$ $p = .003$, $MSE = .002129$, and the block by condition interaction, $F(2, 58) = 3.2$ $p = .048$, $MSE = .001022$, when this second ANOVA was performed on the log transformed RTs. Inspection of the distribution of raw RTs revealed several outlier RTs of greater than 10 s that appear to have disproportionately influenced the analysis on mean RTs in this case. The above results are essentially the same as those for Experiment 1, with the caveat that the negative transfer effect on the posttest might be less pronounced. The randomized word order and the change in presentation format on the posttest in Experiment 1 can therefore be eliminated as factors underlying the lack of positive transfer to response-change items for the case of interactive imagery.

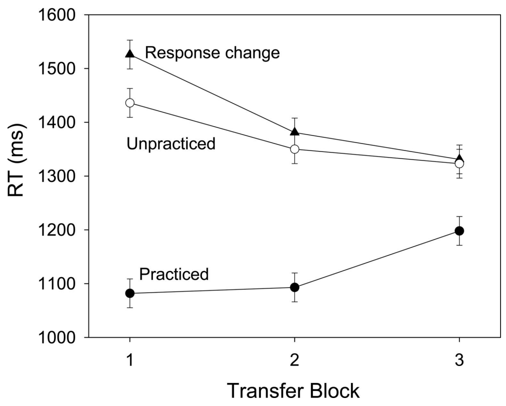

Figure 3. Mean response times (RTs) on the posttest of Experiment 2 as a function of condition and block. Error bars are standard errors based on the interaction error term of a within-subjects analysis of variance on the subject mean RTs.

Experiment 3

The results of Experiments 1 and 2 led us to consider what conditions, if any, might support positive within-triplet transfer. One candidate is triplet knowledge that is acquired, practiced, and tested in the form of a sentence structure. Consider the sentence, "Snow falls gently." After initial study and pretest, subjects were given cued-recall practice on, for example, *Snow* _____ *gently.* On the posttest, response-change sentences, ______ *falls gently* and *snow falls* ______ were tested, using the same design as in Experiment 2. Presentation of stimuli in sentence form does not preclude other types of representation in memory, such as imagery. However, that possibility was not problematic given the goal of this experiment, which was to determine whether existence of a linguistic form of representation can support positive transfer.

Method

Subjects and materials. Thirty University of California at San Diego undergraduate students participated for course credit. Stimuli were 16 three-word sentences (see Appendix B).

Design and procedure This experiment was identical to Experiment 2, with two exceptions. First, throughout the experiment, word triplets were

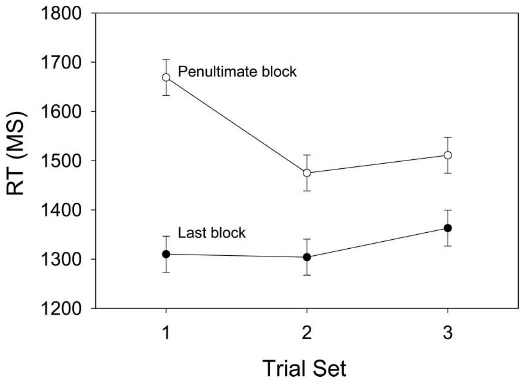

Figure 4. Mean response times (RTs) as a function of pretest block and item set in Experiment 3. Error bars are standard errors based on the error term of a within-subjects analysis of variance on the subject mean RTs.

presented in a standard sentence form, arranged sequentially on a single row, always with the same (grammatically correct) word order and with first word capitalization and a closing period. During the study phase, subjects were simply instructed to study the sentences. On all cued-recall trials (in pretest, practice, and posttest phases), the sentence structure was exactly the same as at study, with the exception that a blank underline replaced the to-be-recalled word. Second, a minimum of two pretest blocks were required for all subjects. This design allowed for a balanced analysis of pretest errors and RTs over both set and practice block. On the basis of the first two experiments, we expected to see significant speedup over the two pretest blocks but no speedup over trial sets within pretest blocks.

Results and Discussion

Pretest. Twenty-seven subjects reached the 90% accuracy criterion within the first two pretest blocks, one subject required three pretest blocks, and two other subjects required four blocks. Mean accuracies on the first, second, and third set of 16 trials on the penultimate pretest block were .77, .87, and .89, respectively. On the last pretest block, these mean accuracies were .94, .96, and .96. The mean RTs for the last two pretest blocks are shown in Figure 4. A within-subjects ANOVA with factors of set and block revealed a significant effect of block, $F(1, 29) = 10.1$, $p < .02$, $MSE = 227{,}633$, confirming the expected item repetition effect. There were no significant effects, however, of either set, $F(2, 58) = 1.19$, $MSE = 126{,}078$, or the set by block interaction, $F(2, 58) = 2.5$, $p = .09$, $MSE = 80{,}382$. There was a trend toward reduced RTs between the first and second set of trials on the penultimate block. However, there was also a substantial increase in accuracy over those two sets, indicating that stable holistic representations were presumably not present for a disproportionate number of items on the first set of trials on the penultimate block. A strong interpretation of the trend toward an RT speedup effect between the first and second sets is thus difficult.

Practice and transfer phases. Accuracy during the practice phase increased from .93 on the 1st block to 1.00 on the 20th block. Mean correct RT during the practice phase decreased from 1,308 ms to 806 ms.

Mean accuracy rates on the posttest were .99, .96, and .97 for the practiced, response-change, and unpracticed items, respectively. As in the previous experiments, the small difference in accuracy between the response-change and unpracticed conditions was not significant (sign-rank test, $p = .08$), whereas the orthogonal test comparing the mean accuracy of the response-change and unpracticed conditions with that of the practice condition was highly significant (sign-rank test, $p < .001$).

The means of the subject-level mean RTs on the posttest are shown in Figure 5 as a function of condition and transfer block. The results mirror those of the first two experiments. In the ANOVA comparing RTs in the practiced condition to the mean RTs in the response-change and unpracticed conditions, there were highly significant effects of block, $F(2, 58) = 8.3$, $p < .001$, $MSE = 10{,}489$; condition, $F(1, 29) = 76.7$, $p < .001$, $MSE = 8911$; and their interaction, $F(2, 58) = 15.9$, $p = < .001$, $MSE = 4{,}175$. In the second ANOVA, comparing RTs in the response-change condition with those in the unpracticed condition, there were significant effects of block, $F(2, 58) = 31.68$, $p < .001$, $MSE = 10.742$, and condition, $F(1, 29) = 13.9$ $p < .001$, $MSE = 24.471$, and a marginal interaction, $F(2, 58) = 3.08$, $p = .054$, $MSE = 6{,}107$.

The negative transfer effect survived all three posttest blocks in this experiment, whereas it dissipated after the first posttest block in Experiments 1 and 2. We can only speculate about this effect, but one possibility is that the vertical word presentation format and/or the imagery instructions of Experiments 1 and 2 encouraged a configural learning strategy in which the pair of presented words was used as a compound cue for retrieval of the missing word. The left-to-right word presentation format in Experiment 3, on the other hand, may have encouraged a strategy of simply associating the left-most presented word with the missing word. For response-change items on the posttest, the configural learning strategy would likely be more resistant to interference, because the pair of words presented during practice would not be recreated. In contrast, if subjects were more likely to form associations between single words and responses in Experiment 3, then those single words, when seen as part of a response-change item, may have been more likely to draw subjects into initiating retrieval of a response that was correct for the corresponding practiced item.

Our prior work on arithmetic (Rickard, 2005; Rickard & Bourne, 1996) lends some support to the hypothesis that failure to

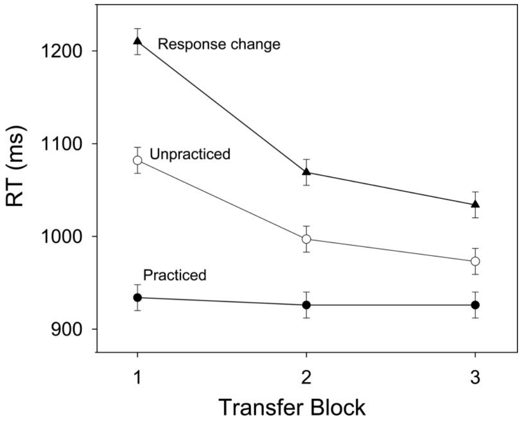

Figure 5. Mean response times (RTs) on the posttest of Experiment 3 as a function of condition and block. Error bars are standard errors based on the interaction error term of a within-subjects analysis of variance on the subject mean RTs.

represent stimulus elements as a configural cue is responsible for negative transfer effects on the posttest. In those studies, practice on one division problem, such as *28/7*, produced negative transfer to its inverse, *28/4*, whereas there was no negative transfer from multiplication to division or vice versa. It is possible that in the case of division practice, subjects focused on the dividend for each problem (e.g., *28*) because it uniquely occurred for only one problem in the stimulus set. Focus on the dividends may have led to strengthening of simple associations from each dividend to its response during practice, rather than (or in addition to) strengthening of the association from the entire division problem to the response. On the posttest, when the inverse division problem was presented (e.g., *28/7*) the simple association from *28* to *4* that formed during practice may have been responsible for the negative transfer. In the current experiment, subjects may have focused on the left-most word cue during practice simply because it was the first cue that was perceptually identified.

The results so far demonstrate that for episodic memories based on both interactive imagery and sentences, cued-recall practice on an item from a triplet results in either no RT transfer to response-change items from that triplet (in the case of the pretest) or negative transfer to those items (in the case of the posttest).

Practice Curve Analyses

The transfer results eliminate the simplest case holistic strengthening model that we set out in the introduction, instead supporting the IE model. It is conceivable, however, that some elaborated version of a strengthening model might be able to accommodate our results. For example, it is possible that the holistic representation continues to support retrieval throughout practice, with no shift in representation, but that access to the holistic representation itself does not become faster. Instead, the speedup observed for practiced items could reflect peripheral factors. Consider the possibility that the observed speedup reflects only improved response availability (e.g., long-term lexical and articulatory priming) for the required responses or, similarly, in associative strengthening between the holistic representation and the response, with no change in the holistic representation itself. Improved response availability with practice would occur only for the eight response words corresponding to the practiced items. If improved response availability is the only factor responsible for the speedup, it would follow that RTs on the posttest should be fastest for the practiced items, with no RT difference between response-change and unpracticed items. The negative transfer to the response-change items on the posttest might also be accommodated within this response availability model in the following way. Consider the case in which a subject practices *boy, gift,* ______. The response "smile" becomes, by hypothesis, more available and more quickly accessed following practice. On the posttest subjects see a response-change item, such as *gift, smile,* ______. Now one of the cue words, *smile*, is the former response and may serve to prime or enhance the availability of "smile" as a response on that trial. Thus, for response-change problems but not for unpracticed problems, there may have been interference because the highly available response was one of the cues.

Common to both the holistic model outlined in the introduction and the response availability model outlined above (as well as hybrids of these models) is the assumption that there is no qualitative shift in task processing with practice. Rather, both models can be understood as reflecting strengthening of the representations that were formed by the initial study phase (or that were preexisting, in the case of the response availability account). Ubiquitously in the skill literature, mean RTs for tasks for which a single process becomes faster with practice—with no qualitative shift in representation—are shown to closely follow a power function, the simplest form of which is

$$\mu_{\mathrm{RT}} = bN^{-c}, \tag{1}$$

where μ_{RT} is the population mean RT, N is the practice block, the parameter b corresponds to the predicted initial RT, and c is the nonlinear rate of speedup with practice (Heathcote, Brown, & Mewhort, 2000; Newell & Rosenbloom, 1981; Rickard, 1997; Rickard & Bourne, 1996). This version of the power function predicts a linear relation between RT and practice block when data are plotted in log-log coordinates. Rickard and Bourne (see Figure 1), for example, demonstrated essentially exact log-log-linear speedup in adults for simple arithmetic (e.g., $4 \times 7 =$ ______), a task for which practice should only result in strengthening of already existing IE associations. Similarly, speedup with practice in executing multistep algorithmic tasks that reflects only strengthening of, and perhaps gradually improved fluidity in execution of (Carlson & Stevenson, 2003), the component steps also follows a power function (e.g., Anderson & Fincham, 1994; Rickard, 1997, "no-transition participant" in Figure 11).

In contrast, several articles have demonstrated that the power function does not describe speedup for tasks that exhibit a shift with practice from algorithm- to retrieval-based performance (Delaney et al., 1998; Rickard, 1997, 1999, 2004). The power function describes speedup for each of these strategies considered in isolation, in accordance with the work summarized above, but overall mean RTs exhibit clear and predicable deviations from power function speedup, reflecting the gradual transition over items and subjects from use of the algorithm to use of memory retrieval to solve the problems. To date, these systematic deviations from log-log linearity have been observed only when there has been independent reason to believe that practice resulted in a strategy shift. For mean RT data, a four-parameter strategy mixture model often fits data from these tasks well (Rickard, 2004). The equation is as follows:

$$\mu_{\mathrm{RT}} = \mu_{\mathrm{alg}}(p) + b(\mathrm{N} - 1)^{-c}(1 - p), \tag{2}$$

where μ_{alg} is the algorithm (or more generally, the initially used strategy) mean RT; $p = \mathrm{e}^{-r * (N-1)}$ is an empirically motivated, simplest case function governing the proportion of trials on which the algorithm is used as a function of practice block (Rickard, 1999); and b and c are power function parameters describing speedup in execution of the memory retrieval strategy with practice. In this version of the mixture equation, the algorithm (initial strategy) RTs are assumed to not speed up with practice, although in the more general model that need not always be true (e.g., Rickard, 1997, Experiment 1).

Thus, the form of the practice speedup curves in the current experiments can be used as a basis for discriminating between two broad classes of models: those that predict no qualitative changes with practice in the representations and processes that support performance and those that predict a qualitative change with

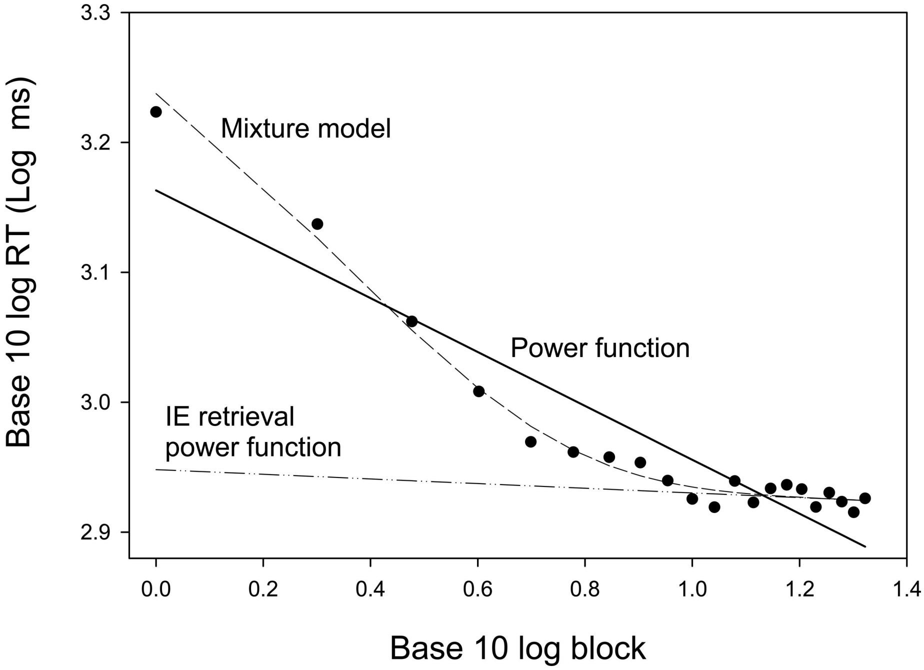

Figure 6. Mean log response times (RTs) plotted against log block for the practice data from Experiment 1. The best fitting two-parameter power and four-parameter mixture functions are shown, along with the predicted retrieval power function based on the mixture model fit. IE = identical elements.

practice, such that a faster process or strategy eventually overtakes a slower process or strategy. (For examples of other work in which the functional form of the speedup curve has been used to explore whether qualitative changes occur with practice, see Anderson and Fincham, 1994; Dulaney et al., 1998; and Pirolli and Anderson, 1985). The holistic strengthening and response availability models are members of the former class, and the IE model is a member of the later class. Here, holistic (episodic) retrieval plays the role that the algorithm strategy does in the skill studies outlined above, and IE-based retrieval plays the role that the memory retrieval strategy does in the skill studies.

To explore this issue, we log transformed the raw RTs from the last pretest block and the 20 practice blocks of Experiments 1 through 3, then averaged them across items within subject, and then averaged across subjects within each experiment.[3] We then plotted these mean data against the log of practice block, as shown in Figures 6 through 8.

The power function fits in all three figures are easily rejected by a Wald-Wolfowitz runs test for randomness of residuals, *p*s $<$ .001.[4] In contrast, the fits of the mixture equation (Equation 2) are good in all cases and are not rejected by the runs test (all *p*s $>$ .30). The mixture model fits better in all three cases even when Akaike's information criterion is applied to adjust for its greater number of free parameters.[5] The parameter estimates for p in the mixture fits imply that the shift to IE-based performance occurred for about half of the items by about the second or third practice trial for both the interactive imagery and sentence experiments.

The foregoing speedup analyses speak against pure strengthening models as a general class (including the holistic, response availability, response strengthening, and hybrid models), at least for the current tasks, and instead point to a qualitative shift in the cognitive basis of performance following minimal cued-recall practice. These results further buttress the IE model in the domains of both interactive image and sentence memory.

The RT curve analyses also suggest that skill models (Logan, 1988; Palmeri, 1997; Rickard, 1997) that posit a strategy shift from

---

[3] To encompass the maximum range of cued-recall practice, we included the last pretest block for each subject, along with the subsequent 20 practice blocks, in the speedup curve analyses. The pretest blocks differed from the practice blocks only in that there were 48 instead of 8 items (trials) per block. This fact did not bias the mean RT estimate for the first block in Figures 6 and 7, however, based on the pretest analyses, which showed that RTs did not change from the first to the third set of trials within the final pretest block.

[4] The generalized three-parameter version of the power function, $RT = a + bN^{-c}$, includes a nonzero asymptote parameter, a. Previous research has shown that that parameter can usually be ignored (particularly when there are only 20 item repetitions) with negligible loss in quality of fit (e.g., Heathcote, Brown, & Mewhort, 2000; Newell & Rosenbloom, 1981). Nevertheless, we also performed nonlinear regression fits of the three-parameter function on the mean of the raw RTs, using block instead of log block as the predictor variable. As expected, the fits were not meaningfully better, and the p values of the runs test were $<$ .001 for all experiments.

[5] The formula used to apply Akaike's information criterion (AIC) is $\ln(SSE/N) + 2 * (k + 1)/N$, where SSE is the sum of squared errors, N is the number of data points, and k is the number of model parameters. Lower values correspond to better model fits. For the power function fits, the AIC values were −6.78, −6.62, and −7.477 for the data from Experiments 1, 2, and 3, respectively. The corresponding values for the mixture model fits were −9.18, −9.12, and −10.24.

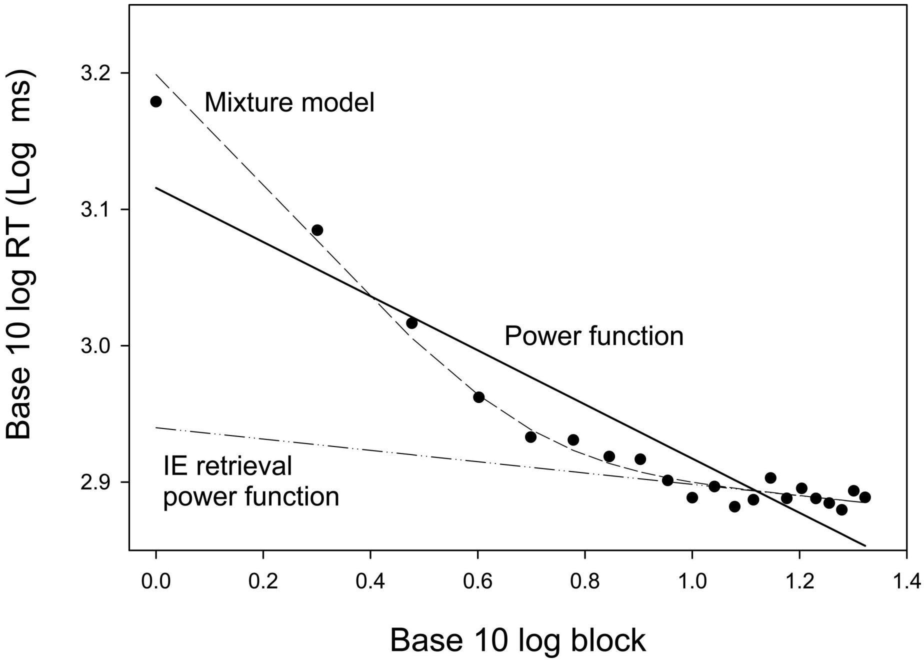

Figure 7. Mean log response times (RTs) plotted against log block for the practice data from Experiment 2. The best fitting two-parameter power and four-parameter mixture functions are shown, along with the predicted retrieval power function based on the mixture model fit. IE = identical elements.

algorithmic to retrieval-based performance may apply well beyond the class of tasks for which they have been developed to date. It appears that even generic memory recall, for which there is no algorithm as traditionally defined in the skill literature, exhibits qualitative process and representation shifts with practice. Instead of a shift from algorithm to retrieval, there was in the current experiments a shift from episode-based to IE-based recall. Rickard's (1997) component power laws (CMPL) mixture model embodies the IE representation in its associations between the item representation and the response. Alternative instance-based models (Logan, 1988; Palmeri, 1997) of strategy shifts are silent to date on the transfer issues addressed here.

General Discussion

Two sources of evidence converge to support the hypothesis that cued recall through recently acquired episodic memory results in a new, item-specific representation that is independent of the study episode and supports subsequent cued-recall performance. First, on both the pretest and the posttest, speedup attributable to cued recall for one item from a triplet (e.g., *boy, gift,* ______) did not transfer positively to response-change items taken from the same triplet (e.g., *boy,* _____, *smile*). On the posttest, there was in fact negative transfer, apparently resulting from interference from the practiced item. Second, the speedup curves from practice exhibited deviations from log-log linearity that are empirically diagnostic of a qualitative shift in processing. Both results are consistent with the IE model of arithmetic fact organization and extend that model to the general case of cued recall from recently acquired episodic memories.

The data further suggest that the holistic, episodic representations that are formed during study and that logically must support performance on the first cued-recall trial are not strengthened or enhanced at all by the act of cued recall. Before drawing this strong conclusion, however, the alternative possibility of a population mixture effect on the pretest should be considered. For some triplets, IE representations might not have formed during the pretest. For those triplets, there may have in fact been holistic strengthening, and consequent speedup, from the first to the third sets. For other triplets, it is probable that item-specific IE representations were formed on the first or second set of 16 trials on the last pretest block (or on an earlier pretest block for subjects who had one). Given the interference and negative RT transfer observed for response-change items on the posttest, these IE representations might have interfered with performance on the corresponding (response-change) items on the second or third set of trials on the pretest. If slowing attributable to this interference more than offset the speedup attributable to any concurrent holistic strengthening for the corresponding triplets, then RTs may have been slowed from Set 1 to Set 3 for those triplets. The consequence for overall mean RTs could, by happenstance, be zero net speedup in mean RT over the three pretest sets. This hypothesis of a mixture of interference (slowing) and facilitation effects for different subsets of items predicts that the *SD* of the overall mean RTs should increase from the first to the third pretest sets. Although subject-level *SD*s based on a relatively small number of trials (up to 16 correct trials per subject per pretest set) can be highly variable, by averaging these data over the 90 subjects from the three experiments, reasonably stable population *SD* estimates may be obtainable. These grand mean subject-level *SD*s, along with the grand

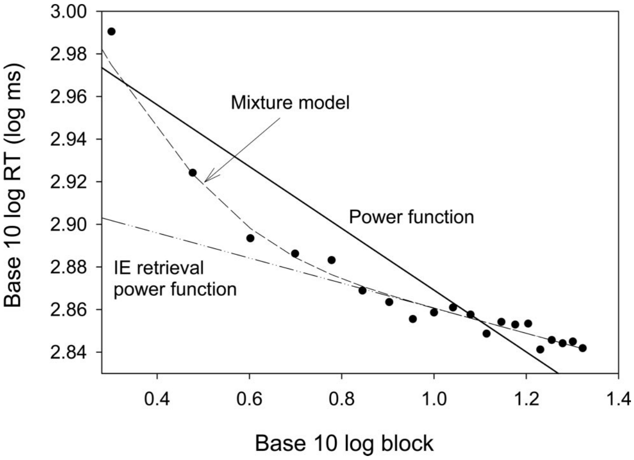

Figure 8. Mean log response times (RTs) plotted against log block for the practice data from Experiment 3. The best fitting two-parameter power and four-parameter mixture functions are shown, along with the predicted retrieval power function based on the mixture model fit. IE = identical elements.

mean RTs, are shown in Figure 9. As expected, the mean RT results mirror the experimentwise results, indicating no change from Set 1 to Set 3, $F(2, 178) < 1.0$. The same results hold for the *SD*s, $F(2, 178) < 1.0$. It therefore appears safe to conclude that there was no substantive holistic strengthening effect, even for a subset of the items, on the pretest. The substantial decreases in RT between the last pretest block and the first practice block (and also between the last two pretest blocks in Experiment 3), however, clearly show that there was substantial learning taking place on the pretest trials. That learning appears to reflect solely formation of item-specific IE representations that speeded subsequent item repetition performance. The finding that there was no negative transfer over the three sets of pretest items can then be accounted for under the assumption that the IE associations that formed during the pretest were new and relatively weak and thus did not interfere much with retrieval by way of the holistic study episode. Following 20 blocks of practice, however, the IE representations were stronger and thus able to interfere with retrieval through the holistic representation for response-change items on the posttest.

Note that the conclusions above do not necessarily imply that repeated study exposures for a triplet will not facilitate RT on the first cued-recall trial. Repeated study opportunities may well strengthen or otherwise enhance the episodic or holistic memory representation that was encoded on the first study trial, in turn speeding RTs on the first cued-recall trial, though that remains as an empirical question.

These conclusions contrast with those of Pirolli and Anderson (1985), Anderson et al. (1997), and Rabinowitz and Goldberg (1995), who advanced the hypothesis that practice enhances access to the originally encoded representations. Pirolli and Anderson studied the effects of practice on sentence recognition. Evidence favoring holistic strengthening came both from preservation of the fan effect (Anderson, 1976) at all practice levels and from a clear power function speedup curve for mean RTs when plotted by practice session. It may be that, because the same sentence stimuli—with no missing element to be recalled—were represented during both study and practice in their experiments, the holistic strengthening model holds for their task and more generally for recognition memory practice. In principle, however, any memory task that requires an overt response can be described as a cued-recall task following practice. In the case of sentence verification, subjects learn to make a "target" (key press) response for each target sentence, setting up a stimulus-response distinction just as for any other cued-recall task. Thus, a shift to an IE-like form of representation is conceivable for the sentence verificiation task. Despite Pirolli and Anderson's demonstration of power function speedup, this shift hypothesis remains viable, because Anderson and Pirolli analyzed mean RTs at the level of each practice session instead of at the level of each repetition. As pointed out by Rickard (1999), and as clearly evident in the current experiments, deviations from log-log linearity can be contained within the first several item repetitions of a single session. Plots of mean RT results at the level of each practice session may thus mask the important shift phase of practice, yielding approximate power function speedup even if there is a shift. It remains an open question, then, whether a qualitative process shift similar to that demonstrated here occurs in the case of recognition memory practice.

Anderson et al. (1997) found evidence for transfer of cued-recall retrieval practice to inverted items using a very different task than ours. They had subjects study items such as "Skydiving was practiced on Saturday at 5 p.m. and on Monday at 4 p.m." For each sport there was a rule to be discovered. In the example above, the rule is that the second practice is two days (minus 1 hr) after the first practice. Then subjects were given practice responding with the second day and time when given the first for an item, or the reverse. Anderson et al. found that following sufficient repetition practice, speedup in responding in one direction transferred positively to responding in the other direction, and they interpreted this in terms of a transition to memory encoding for the practiced items, which was symmetrically accessible. RTs following repetition practice in their task were, however, in the 4 to 6 s range, as opposed to the 800–900 ms range by the end of practice in the current experiments. Error rates in their experiment were also much higher (about 17% initially and about 7% by the end of practice). A great deal of cognition can occur in 4 s, so it seems unlikely that the mental processes in their tasks were at any point

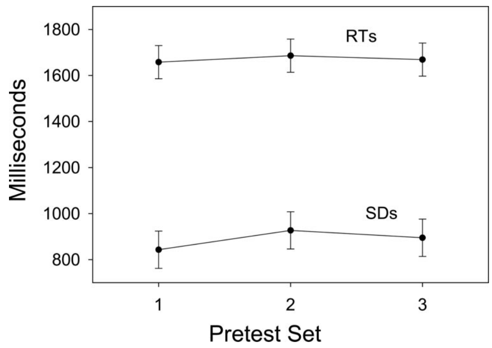

Figure 9. Mean response times (RTs) and standard deviations (*SD*s), averaged over all 90 subjects from the three experiments, on the last pretest block. Error bars are standard errors based on the error term of the within-subjects analyses of variance.

during practice directly comparable to those in ours. Because of the complexity of their items, along with the possibly high inter-item interference among them, their subjects may have had delayed onset of episodic retrieval for repeated items (Anderson et al.'s data indicated that their subjects used the rules for most items throughout most of practice). It would not be surprising from our perspective for those episodic (holistic) representations to support positive transfer to reversed items. There may not have been enough practice opportunity in their experiment for a subsequent shift to IE-based recall to occur.

There is ample evidence that cued recall for most tasks can be completed in 1 s or less after sufficient practice. The empirical evidence suggests that at that performance speed, the predicted shift to IE representation is likely to have occurred for all or at least most items. It is therefore imperative that in future tests of holistic versus IE representation, RTs at the end of the practice phase be in the range of about 1 s.

Anderson et al. (1997) proposed a multistage model of skill learning in which the last stage involves retrieval of holistic-type, or symmetric, memories of previous exposures to an item. Our results indicate that their model is incomplete. In at least some cases, the last stage instead reflects retrieval from asymmetric, stimulus–response associations that appear to have an IE structure.

Rabinowitz and Goldberg (1995; Experiment 2) concluded that practice on alphabet arithmetic (e.g., $D + 3 = ?$) transferred positively to inverted problems (e.g., $G – 3 = ?$). This transfer condition is analogous to our response-change condition, in which we observed no positive transfer. Inspection of Rabinowitz and Goldberg's transfer results, however, raises concern about their conclusion. At the end of 36 blocks of practice, the mean RT for their repeated-item subject group was about 1,300 ms. On the posttest, mean RTs for inverted items for their repeated-items group jumped up to about 4,500 ms, in the range expected for performing alphabet arithmetic using the alphabet recitation algorithm and far slower than that typically observed in alphabet arithmetic for memory-based performance (e.g., Rickard, 1997). By this measure, then, it appears that Rabinowitz and Goldberg's subjects reverted to an algorithm for inverted problems on the posttest, being unable to access the memory representation that supported the fast performance on the practiced items, just as the IE model predicts.

Rabinowitz and Goldberg (1995) did not discuss the evidence favoring a shift to IE representation that is outlined above, instead focusing on the fact that RTs on the posttest on inverted items were much slower, around 9,700 ms, for a different group of subjects who had practiced on a larger set of alphabet arithmetic items in which there was minimal repetition of specific items (we refer to this group as the *non-repeated-items group*). They used the transfer RT for the non-repeated-items group as, in effect, the equivalent of our unpracticed condition, and they focused on the fact that RTs for inverted problems for the repeated-items group were much faster than for the non-repeated-items group. That approach to interpretation does not acknowledge, however, the possibility that repeated solving of problems in the forward direction (e.g., $D + 3 = ?$) in the repeated-items group during practice might have facilitated algorithm execution on inverse problems like $G – 3 = ?$ on the posttest. Note that alphabet knowledge is primarily forward associative; that is, it is much easier to recite forward than backward (Klahr, Chase, & Lovelace, 1983). In fact, to recite a certain number of letters backward in the alphabet, as in $G – 3 = ?$, a possible strategy is to first recite forward from an earlier point in the alphabet, perhaps from *A* in this case, loading letters into working memory and so facilitating backward recitation from *G*. In the process of that forward recitation strategy, the subject arrives at *D*, and then, if in the repeated-items condition, has all cues necessary to retrieve the previously learned IE association, $(D + 3) \rightarrow G$. Thus, without finishing the recitation to *G*, the subject can in principle use the memorized fact from practice to respond "*D*" to the problem $G – 3 = ?$, shortcutting full algorithm execution for inverse problems on the posttest (relative to performance of subjects in the non-repeated-items condition, who would not have any IE representations from practice to use for algorithm shortcutting). Under this scenario, RTs for inverse problems on the posttest would be faster for the repeated-items group, with no need to assume that practice resulted in strengthening of holistic memory representations of the items. Using a design and novel arithmetic task that was not subject to this idiosyncratic strategy, Rickard (1997, see Figure 5) found no positive transfer to inverted (type-change) items following the shift from algorithm- to retrieval-based performance with practice.

In light of the possible idiosyncratic strategy for inverse alphabet arithmetic outlined above, a better way to test the IE model for alphabet arithmetic would be to first practice subjects on the entire set of forward and inverse problems to be tested until they shift from algorithm to memory retrieval for all items (for discussion of the diagnostics of a shift to retrieval, see Rickard, 2004) and then give them further practice on a subset of the items. The IE model makes the strong prediction that speedup with practice on that subset of items will not transfer to inverted items. Such a design reduces to something very similar to the arithmetic transfer studies already discussed (e.g., Rickard & Bourne, 1996).

Crutcher and Ericsson (2000) also concluded in favor of a strengthening account for the case of the keyword method of foreign vocabulary learning (Atkinson & Raugh, 1975). In that task, a foreign word (e.g., *perro*) is mapped to a phonologically similar English keyword (*pear*) and an interactive image is formed between the keyword and the English translation of the foreign word (e.g., a dog eating a pear). Crutcher and Ericsson concluded that practice results in a drop-out of conscious use of the keyword and interactive image mediator, but that these mediating representations remain a permanent, or at least very long term, component of the retrieval process. They suggested that no shift to a direct associative link between the foreign and English words occurs with practice. Their main evidence came from an interference manipulation. After extensive practice on the vocabulary task, there was an interference phase in which the keyword was associated with a new response (e.g., *pear* was associated with *cotton*) for half of the items. Subsequent performance on the vocabulary task was slower for items for which keywords had been associated with a new response, suggesting that the keyword and imagery mediators were still an integral part of the retrieval pathway even though their use had dropped out of working memory.

Rickard and Bajic (2003) challenged the generality of their conclusions, however, on several grounds. They pointed out that the keyword task is an unusual mnemonic mediator in that its first step, keyword extraction, involves extraction of a phonological subset of the foreign word stimulus. It may be that keyword extraction continues to be part of the retrieval pathway for an unusually long period of practice even if there is an eventual shift to direct retrieval. Furthermore, because their interference manipulation was at the stage of keyword extraction, their results have no bearing on whether the interactive image continued to be a part of the retrieval pathway or instead was replaced early in practice by a direct link between the keyword and the English response word.

Using a design analogous to Crutcher and Ericsson's, (2000) but with a task that did not have the idiosyncratic property of the keyword task that is outlined above, Rickard and Bajic (2003) showed that the interference manipulation had no effect on performance following practice, indicating that a qualitative shift to a direct retrieval pathway had occurred. In their reply, Crutcher and Ericsson (2003) pointed out that Rickard and Bajic's task did not involve semantic memory, which they suggested might be the crucial issue. The initial representations in the current experiments clearly did involve semantically based processing, however, and qualitative shifts in processing were observed. In summary, evidence from triplet transfer, the shape of the speedup curve when analyzed at the level of each repetition, and an interference manipulation all converge on the conclusion that cued-recall practice can result in a qualitative change in the mnemonic pathway underlying performance.

Other Theoretical Considerations

What is the representational substrate of the IE memory, and how does it relate to the substrates of the triplet representations that were formed by initial study? We start with the assumption that the study of triplets in domains such as imagery and sentences results in at least partially distinct representational forms for each domain, at both cognitive and neural levels. One possibility is that these initial representational substrates continue to support performance following practice, but that somehow there is a shift from an episodic to an IE representational organization within each substrate. It is difficult for us to imagine, however, a mental image that would support cued recall of the specific stimulus-response combination *boy, gift,* ______ and that is also separate and independent of the original study representation for that triplet. This difficulty is perhaps compounded for the case of sentences.

An alternative account is that IE-based representations are completely independent of memory substrates that may be specialized to support, say, imagery or linguistic structures. The IE representations may be part of a more generic stimulus-response memory architecture. This architecture may preserve representations of the stimuli and responses but nothing else of the original holistic, episodic structure that was laid down by study. Thus, the same system could support IE-based performance in all knowledge domains. In all experiments, visually presented word stimuli might be represented visually as a pair of words or a chunk of two words for each practiced item, and that visual representation may, because of the vocal response, become associated with a verbal representation of the required word response. Alternatively, the visual word stimuli may be recoded into a verbal form in all experiments, leading to verbal-verbal associative structure. Yet another possibility, assumed in the development of the IE model to date, is that the IE representations are part of a more abstract semantic system that is also divorced from the domain-specific episodic substrates of imagery and sentence memory. In any of these cases, practice would constitute a shift from domain specific (e.g., imagery or linguistic) episodic representation to a common associative system that has an IE organization.

It is important to consider the current results in the context of traditional memory distinctions, such as explicit-implicit and declarative-procedural (for review, see Eichenbaum, 2003). The case of cued-recall practice from newly encoded episodes is not easy to categorize a priori into one of these two types of memory systems. Cued recall involves performance, and repeated recall constitutes skill learning, processes that are generally believed to engage the implicit or procedural system. From this perspective, the shift to IE representation might be understood as a shift to a procedural representation. Because the procedural system is generally believed to be spared in medial temporal (i.e., declarative) amnesia, this perspective implies that, whereas the encoding of the original study episode is dependent on the medial temporal system, the shift to IE representation and subsequent speedup with practice is not. Recent work raises the possibility, for example, that IE-based retrieval is mediated through the basal ganglia (Myers et al., 2003). These researchers found that the basal ganglia support rapid associative learning that transfers poorly to items that are similar but not identical to the practiced items, in line with the learning observed here.

Alternatively, the medial temporal/declarative system may be crucial to the shift to IE representation. Consider the possibility that the first cued-recall event creates a new declarative episode that is independent of the study episode and that encodes as the major part of its trace the temporal/causal dimension of performance, that is, the occurrence of stimuli followed by the execution of a response (a form of representation that is consistent with the IE model). The next time the same item is presented, that cued-recall episode is likely more suitable than is the originally encoded study episode to support recall, because it is a better match to the stimulus-response requirements of the recall task. That is, recall through the study episode would require using the word stimuli to activate the full mental image (Experiments 1 and 2) and then transcoding of the image of the response element into a verbal response. In contrast, the episode laid down by the first cued-recall event might emphasize only a verbal, visual, or semantic representation of the stimulus words and an association between those words and the required verbal response, encoded perhaps as a temporal state change in the neural network. No mental image or sentence representation need logically be encoded as part of this cued-recall episode. The cued-recall episode may thus provide a faster, less effortful pathway to recall. Repeated recall practice might then gradually result in consolidation of this cued-recall episode into neocortex, perhaps as a sparse representation of only what is absolutely consistent from trial to trial and thus absolutely necessary for performance of the cued-recall task.

Finally, one might ask why, from an evolutionary perspective, separate types of representation result from study and cued recall. Why would there evolve a representation for cued recall that yields no positive transfer (and in some cases negative transfer) when new recall items are encountered from the same knowledge unit? We can only speculate here, but consider the possibility of three (nonexhaustive) adaptive goals of a memory system, in decreasing order of importance: (a) one-trial learning and flexible use of newly encoded knowledge, (b) optimal efficiency in performance for information that we retrieve frequently, and (c) optimal transfer on learning to retrieval tasks closely related to those in which we frequently engage. It may not be possible to achieve optimality for the first two goals within a single type of memory representation (for a related point, see McClelland, McHaughton, & O'Reilly, 1995). An episodic memory system that provides for flexible use of memory may require retrieval of too much information to also be optimal for speeded performance on more specific retrieval tasks, such as highly practiced cued recall. By similar reasoning, the third goal of optimal transfer of speedup with practice to related information, as in the case of transfer to response-change items in the current experiments, may be incompatible with the goal of optimal performance on what we do most often (Goal 2). Achieving the second goal may require that only information that is absolutely necessary to support performance be encoded. The hypothesis here is that the simpler the representation, the faster the access to and retrieval from that representation, all other factors held constant. The IE representation can be seen as an example of knowledge simplification for specific stimulus-response recall tasks. Thus, speedup with practice does not transfer to related recall items, such as the response-change items in the current experiment. Conceivably, a third memory system might have evolved, tailored to support optimal performance on transfer items. The neural resources required for such a system, however, may not have been worth the cost of inevitably degrading some other aspect of cognition, and the evolutionary pressure toward development of such a system might have been relatively low.

Conclusions

The foregoing experiments provide solid evidence pertaining to each of the four basic questions outlined in the introduction. Regarding the first question (Is there a qualitative shift with practice?), the answer is affirmative. Evidence comes both from the transfer results, which demonstrate a surprising degree of stimulus-response specificity in learning, and from the speedup curve analyses, which appear to rule out simple strengthening models as accounts of the learning effect. The second and third questions addressed the rate of the shift and its form of representation relative to the original episodic encoding. Results showed that the shift can happen quite quickly, often with only one or two cued-recall repetitions, and that knowledge representation following the shift has stimulus-response specificity consistent with the IE model. A final question addressed the domain generality of the shift process. The results indicate that the shift to IE-based performance occurs in a variety of cued-recall domains, regardless of the properties of the representations that support initial performance.

References

Anderson, J. R. (1976). *Language, memory and thought*. Hillsdale, NJ: Erlbaum.

Anderson, J. R., & Fincham, J. M. (1994). Acquisition of procedural skills from examples. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 20,* 1322–1340.

Anderson, J. R., Fincham, J. M., & Douglass, S. (1997). The role of examples and rules in the acquisition of a cognitive skill. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 23,* 923–945.

Atkinson, R. C., & Raugh, M. R. (1975). An application of the mnemonic keyword method to the acquisition of a Russian vocabulary. *Journal of Experimental Psychology: Human Learning & Memory, 104,* 126–133.

Campbell, J. I. D. (1997). On the relation between skilled performance of simple multiplication and division. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 23,* 1140–1159.

Campbell, J. I. D. (1999). Division by multiplication. *Memory & Cognition, 18,* 245–275.

Campbell, J. I. D., Fuchs-Lacelle, S., & Phenix, T. L. (in press). Identical elements model of arithmetic memory: Extension to addition and subtraction. *Memory & Cognition*.

Carlson, R. A., & Stevenson, L. M. (2003). Information acquisition and the cognitive structure of arithmetic. *Memory & Cognition, 31,* 1249–1259.

Crutcher, R. J., & Ericsson, K. A. (2000). The role of mediators in memory retrieval as a function of practice: Controlled mediation to direct access. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 26,* 1297–1317.

Crutcher, R. J., & Ericsson, K. A. (2003). The effects of practice on mnemonic encodings involving prior knowledge and semantic memory: Reply to Rickard and Bajic (2003). *Journal of Experimental Psychology: Learning, Memory, and Cognition, 29,* 1387–1389.

Delaney, P. F., Reder, L. M., Staszewski, J. J., & Ritter, F. E. (1998). The strategy-specific nature of improvement: The power law applies by strategy within task. *Psychological Science, 9,* 1–7.

Diller, D., Nobel, P. A., & Shiffrin, R. M. (2001). An ARC-REM model of accuracy and response time in recognition and recall. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 27,* 414–435.

Eichenbaum, H. (2003). Memory systems. In M. Gallagher & R. J. Nelson (Eds.), *Handbook of psychology: Biological psychology* (Vol. 3, pp. 543–559). New York: Wiley.

Fitts, P. M., & Posner, M. I. (1967). *Human performance.* Oxford, England: Brooks/Cole.

Heathcote, A., Brown, S., & Mewhort, D. J. K. (2000). The power law repealed: The case for an exponential law of practice. *Psychonomic Bulletin and Review, 7,* 185–207.

Hittmair-Delazar, M., Semenza, C., & Denes, G. (1994). Concepts and facts in calculation. *Brain, 117,* 715–728.

Klahr, D., Chase, W. G., & Lovelace, E. A. (1983). Structure and process in alphabetic retrieval. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 9,* 462–477.

LeFevre, J., & Morris, J. (1999). More on the relation between division and multiplication in simple arithmetic: Evidence for mediation of division solutions via multiplication. *Memory & Cognition, 27,* 803–812.

Logan, G. D. (1988). Toward an instance theory of automatization. *Psychological Review, 95,* 492–527.

McClelland, J. L., McHaughton, B. L., & O'Reilly, R. C. (1995). Why there are complementary learning systems in the hippocampus and neocortex: Insights from the successes and failures of connectionist models of learning and memory. *Psychological Review, 102,* 419–457.

McCloskey, M., Aliminosa, D., & Sokol, S. M. (1991). Facts, rules, and procedures in normal calculation: Evidence from multiple single-case studies of impaired arithmetic fact retrieval. *Brain and Cognition, 17,* 154–203.

Myers, C. E., Shohamy, D., Gluck, M. A., Grossman, S., Kluger, A., Ferris, S., et al. (2003). Dissociating hippocampal versus basal ganglia contributions to learning and transfer. *Journal of Cognitive Neuroscience, 15,* 185–193.

Newell, A., & Rosenbloom, P. S. (1981). Mechanisms of skill acquisition and the law of practice. In J. R. Anderson (Ed.), *Cognitive skills and their acquisition* (pp. 1–55). Hillsdale, NJ: Erlbaum.

Palmeri, T. J. (1997). Exemplar similarity and the development of automaticity. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 23,* 324–354.

Pirolli, L. P., & Anderson, J. R. (1985). The role of practice in fact retrieval. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 11,* 136–153.

Rabinowitz, M., & Goldberg, N. (1995). Evaluating the structure-process hypothesis. In F. Weinert & W. Schneider (Eds.), *Memory performance* *and competencies: Issues in growth and development* (pp. 225–242). Hillsdale, NJ: Erlbaum

Rickard, T. C. (1997). Bending the power law: A CMPL theory of strategy shifts and the automatization of cognitive skills. *Journal of Experimental Psychology: General, 126,* 288–311.

Rickard, T. C. (1999). A CMPL alternative account of practice effects in numerosity judgment tasks. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 25,* 532–542.

Rickard. T. C. (2004). Strategy execution in cognitive skill learning: An item-level test of candidate models. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 30,* 65–82.

Rickard. T. C. (2005). A revised identical elements model of arithmetic fact representation. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 31,* 250–257.

Rickard, T. C., & Bajic, D. (2003). Automatic mediation or absence of mediation? Commentary on Crutcher and Ericsson (2000). *Journal of Experimental Psychology: Learning, Memory, and Cognition, 29,* 1381–1386.

Rickard, T. C., & Bourne, L. E., Jr. (1995). An identical elements model of basic mental arithmetic skill. In A. F. Healy & L. E. Bourne, Jr. (Eds.), *Learning and memory of knowledge and skills* (pp. 255–281). Newbury Park, CA: Sage.

Rickard, T., & Bourne, L. E., Jr. (1996). Some tests of an identical elements model of basic arithmetic skills. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 22,* 1281–1295.

Rickard, T. C., Healy, A. F., & Bourne, L. E., Jr. (1994). On the cognitive structure of basic arithmetic skills: Operation, order, & symbol transfer effects. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 20,* 1139–1153.

Siegler, R. S. (1988). Strategy choice procedures and the development of multiplication skills. *Journal of Experimental Psychology: General, 117,* 258–275.

Appendix A

Word Triplets for Experiments 1 & 2

1.	cup	hand	tea
2.	girl	tree	swing
3.	car	pole	crash
4.	ball	foot	goal
5.	phone	ear	voice
6.	clock	meeting	rush
7.	vent	air	sweater
8.	rain	wet	cat
9.	gift	boy	smile
10.	bus	coins	slot
11.	room	key	enter
12.	cow	grass	milk
13.	trip	knee	hurt
14.	beat	drum	march
15.	wolf	hunt	meat
16.	spray	ants	trash

Appendix B

Word Triplets for Experiment 3

1.	Snow	falls	gently.
2.	Fire	burns	wood.
3.	Rabbits	run	fast.
4.	Wolves	eat	sheep.
5.	Pigs	like	mud.
6.	Rocks	break	glass.
7.	Balls	roll	away.
8.	Salesmen	talk	quickly.
9.	Gravel	feels	rough.
10.	Kids	hate	baths.
11.	Hippos	waddle	slowly.
12.	Trees	make	shade.
13.	People	want	friends.
14.	Cooks	bake	bread.
15.	Walls	stop	thieves.
16.	Bars	block	windows.

Received September 8, 2005
Revision received February 2, 2006
Accepted March 6, 2006 ■