Research Article

THE LAST SHALL BE FIRST: How Memory Strength Affects Children's Retrieval

C.J. Brainerd,[a] V.F. Reyna,[a] M.L. Howe,[b] and J. Kevershan[a]
[a]University of Arizona and [b]Memorial University

Abstract—*When items are read out of storage during unconstrained recall, how should their memory strengths affect recall order? Common-sense argues that items will present themselves to consciousness in the order of their relative memory strength and that, hence, the read out sequence will be stronger to weaker In experiments with 6- to 13-year-olds, however, this sequence was never obtained Dichotomous measures of memory strength (strong vs weak) showed that weak items were always recalled before strong items Continuous measures of memory strength showed that the sequence was weaker to stronger to weaker This pattern, which we term the* cognitive triage *effect, occurred at all age levels, for a variety of material, with different definitions of memory strength, and became more pronounced with age Three results suggested that the pattern was not due to a deliberate strength-based ordering strategy The pattern was observed in children as young as age 6, it was present at the earliest stages of recall, and adults (who also exhibited the pattern) were unaware of what they were doing*

Free recall is the most frequently used paradigm in the memory development literature, owing perhaps to its resemblance to naturalistic situations in which children must retrieve items from episodically-related collections These situations range from innocuous inquiries about the events of everyday life to formal courtroom interrogations about criminal acts (e g , see Ceci, Ross, & Toglia, 1987; Ceci, Toglia, & Ross, 1987). We have discovered that when children regenerate items in this manner, their output queues exhibit a certain property that violates our intuitive psychological theories of retrieval The property is a nonmonotonic relationship between the memory strengths of episodically-related items and the order in which children articulate those items during free recall. We call this relationship the cognitive triage effect because it is redolent of the familiar medical procedure of treating the most difficult cases first

Under just about any definition of memory strength, whether commonsensical or theoretical, it seems self-evident that items with stronger memory representations will appear in consciousness before items with weaker representations. Thus, if recall order is plotted against some monotonic measure of memory strength, the intuitive order of read out is stronger to weaker. This idea was first given quantitative expression in Marbe's Law (Marbe, 1901; Thumb & Marbe, 1901), which proposed that the activation latency of a word's memory representation is an inverse logarithmic function of its strength (see also, Osgood, 1953) The stronger to weaker ordering also falls out of numerous contemporary theories (e g , Bjorklund & Muir, 1988) We have found, however, that children *never* articulate items in this order when their recall is unconstrained. Instead, (a) departures from stronger to weaker are present at the earliest stages of free recall, (b) these departures become more marked as learning unfolds; (c) they increase with age; (d) they obey a simple polynomial rule, (e) they are conserved across forgetting intervals, and (f) they conflict with adults' introspective analyses of their recall We report data bearing on each of these points

GENERAL METHOD

The subjects in our experiments were children and adolescents, ranging from age 6 on the low end to age 13 on the high end (mostly first, second, fifth, and sixth graders) All experiments were standard free recall designs with trials consisting of a study phase during which items were presented individually (3–5 s rate), followed by an irrelevant buffer activity to empty short-term memory (20–30 s of letter shadowing or backward counting), followed by an unconstrained test of recall ("Tell me all the things you can remember") Tests terminated when 15 s had elapsed without item production. Depending on the experiment, the items were either words (concrete or abstract nouns) or pictures drawn from familiar norms (Battig & Montague, 1969, Paivio, Yuille, & Madigan, 1968; Toglia & Battig, 1978) The numbers of items on the lists of individual experiments were 12, 16, or 24

We summarize data from 11 experiments. The number of subjects who participated in the individual experiments ranged from a low of 50 to a high of 96. The experiments fell into two groups. (a) 6 *fixed-trials studies* in which the number of study-buffer-test trials was either 2 (Experiment 1), 4 (Experiment 2), or 5 (Experiments 3–6); and (b) 5 *criterion + retention* studies (Experiments 7–11). In the latter studies, the subjects received whatever number of trials was necessary to achieve a criterion of two consecutive errorless free recall tests. Two weeks later, they returned to the laboratory for a long-term retention session that consisted of five free recall tests, interspersed with 20 to 30 s of buffer activity to empty short-term memory.

THE TRIAGE EFFECT: ACQUISITION DATA

An index of memory strength is required to study the relationship between recall order and memory strength. Since recall

Correspondence and reprint requests to C.J. Brainerd, Division of Educational Psychology, University of Arizona, Tucson, AZ 85721

accuracy has been interpreted as a monotonic measure of memory strength from Ebbinghaus to the present day (e.g., compare Osgood, 1953, to Jahnke, Davis, & Bower, 1989), this was the natural index to choose. Accuracy has the further advantage of being sensitive to individual differences in the memory strengths of specific items. An item that is strong for one subject may be weak for another, but this will translate into an accuracy differential. For this reason, accuracy is an inherently less noisy strength index than other measures that might be considered.

To determine how items' memory strengths affect their output order in the earliest phases of free recall, we classified items that children recalled on Trial 2 of Experiments 1 to 6 as being "strong" or "weak" on the basis of Trial 1 performance. Strong items were those recalled and weak items were those not recalled. We were surprised to find that children recalled weak items before strong ones on Trial 2. Since these studies varied with respect to item difficulty and list length, an overall picture was sought by converting to standard scores (*Z*-scores). For children between 6 and 13, these analyses showed that the distance between the mean recall positions of weak items (earlier) and strong items (later) was in the neighborhood of .5 *SD*. This effect held up from early childhood to early adolescence, though it increased with age. The distance was roughly .3 *SD* and .6 *SD* for the youngest and oldest subjects, respectively.

We examined recall order on subsequent trials using the data of Experiments 2 to 6. Weak items continued to occupy earlier positions than strong items, and the positional spread increased steadily across trials. We classified items according to memory strength as before (strong = correct and weak = error) and then analyzed order of recall on Trials 3 to 5 as a function of the classifications on immediately preceding trials. Weak items were always recalled before strong items. The distance between the average positions of the two types of items increased to approximately .7 *SD* by Trial 5, and became more marked with age (about .9 *SD* for the oldest subjects versus about .5 *SD* for the youngest subjects, pooled across Trials 3–5). So, the counterintuitive relationship between recall order and memory strength was not a chimerical feature of early stages of recall.

So far, a dichotomous index of memory strength has been used. A more fine-grained index is possible that takes items' complete error-success histories into account. On any Trial i, previous recall errors for an item may vary between 0 (strongest items) and $i - 1$ (weakest items). When we applied this graded index to Trials 4 and 5 of Experiments 2 to 6, the children seemed to be implementing memory-strength discriminations that were more refined than simple dichotomies. Their protocols followed a continuous, nonmonotonic ordering, with the weakest items (lowest accuracy) being read out at the start and finish of tests and the strongest items (highest accuracy) being read out in between.

To verify this suggestion, we turn to Experiments 7 to 11. Remember that these studies were identical to Experiments 1 to 6, except for the perfect recall criterion and the retention tests. We plotted order of output on the two criterion tests (when all items were recallable) against two experimentwise measures of memory strength—total errors to criterion per item and trial number of the last error per item. The weaker to stronger to weaker ordering emerged at all age levels with both measures.

To illustrate the fundamental pattern, summary curves for second grade children (7-year-olds) and sixth grade children (11-year-olds) are exhibited in Figures 1 and 2. These curves were constructed by aggregating data from the two age levels across Experiments 7 to 11. (The plots for individual experi-

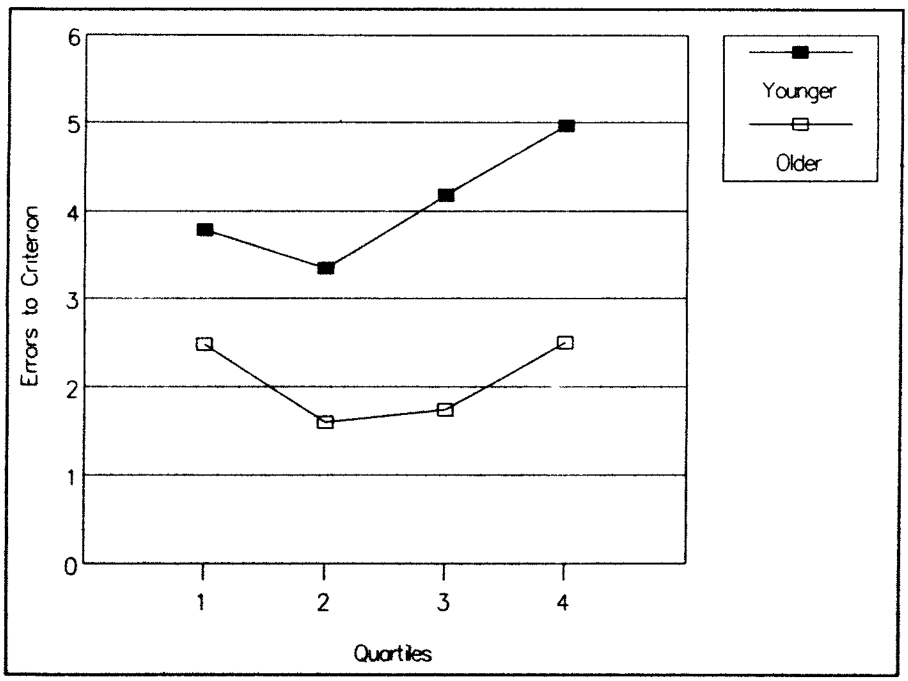

Fig. 1. Mean total errors to criterion per item as a function of the items' positions in criterion recall queues (Vincentized quartiles)

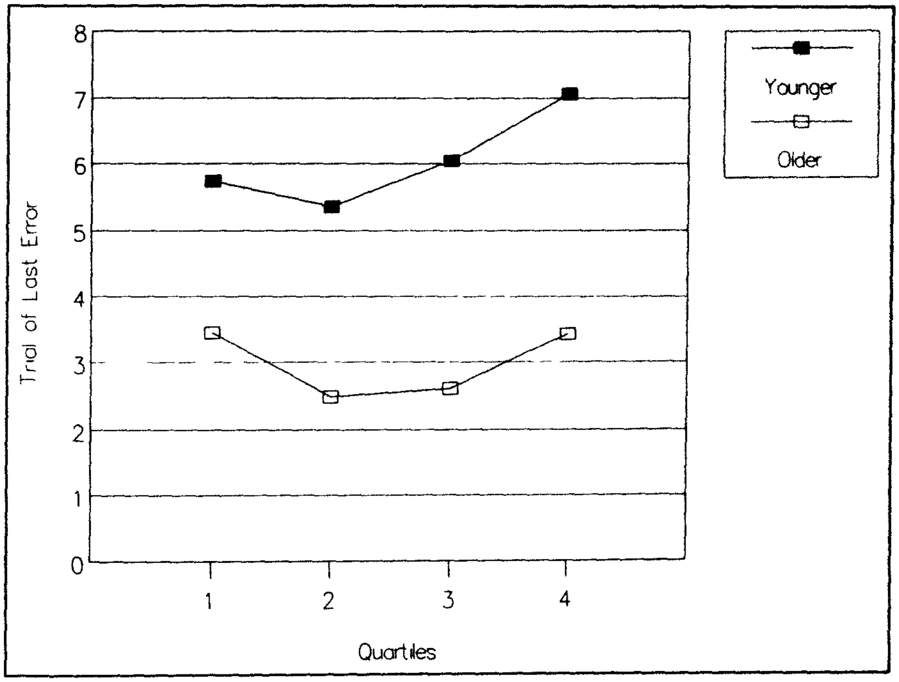

Fig. 2 Mean trial of last error per item as a function of the items' positions in criterion recall queues (Vincentized quartiles)

ments displayed the same pattern as the aggregate plots) Since list length ranged from 12 items to 24 items, the curves in Figures 1 and 2 are based on Vincentized quartiles for the individual studies

Three aspects of these curves are notable First and most important, the relationship between memory strength and recall order was weaker to stronger to weaker at both age levels for both measures of memory strength The children began criterion tests by reading out weaker items, then rotated in progressively stronger items, and finally rotated in the remaining weaker items As might be expected from the shapes of these plots, the curves were well fit by quadratic equations Typically, such expressions accounted for 80 to 95% of the variance

Second, age permuted the internal geometry of the weaker to stronger to weaker relationship For both the total errors measure (Figure 1) and the trial-of-last-error measure (Figure 2), by early adolescence the relationship was a symmetrical U curve of the sort that is routinely associated with serial learning. In young children, however, the weaker to stronger arm was shorter than the stronger to weaker arm; the curve was more like a J than a U

Third, although the usual age trend in recall accuracy was present, the trend was more marked in some segments of the curve than elsewhere The age difference was much greater in the stronger to weaker arm than in the weaker to stronger arm

These findings have been confirmed through reanalysis of published data. We have reported several prior criterion developmental studies of recall. The same patterns have been observed in the data of Brainerd (1983), Brainerd, Howe, Kingma, and Brainerd (1984), Brainerd, Kingma, and Howe (1986a, 1986b), and Brainerd and Reyna (1990, in press)

SPECIAL PROCESSING?

We interpret our data as favoring a weaker to stronger to weaker ordering There is another explanation, however, that might cover our results and retain the stronger to weaker ordering The explanation claims that (a) common-sense and classical theories are correct in supposing that recall accuracy and recall order are both monotonically related to memory strength *at the time that each is measured*, but that (b) processes intervene between the accuracy measurements taken on earlier tests and the order measurements taken on later tests such that (c) accuracy is nonmonotonically related to *current* memory strength by the time that order is measured

This explanation devolves from the familiar idea of error-driven special processing It assumes that on study trials, subjects use error feedback from previous tests to focus mnemonic activity on items that are giving them trouble. This tactic converts some weak items into strong items, and these formerly weak (now strong) items are read out first on the next test. If most of the items classified as "weak" by our dichotomous index were specially processed, this could explain why they were read out before the items classified as "strong." With our continuous index, the lower accuracy to higher accuracy to lower accuracy ordering could be explained on the ground that the lower accuracy items appearing at the start of a queue have become the strongest items via special processing. Finally, the developmental trends could be due to age improvements in special processing ability. Although this explanation is appealing, it is inconsistent with several aspects of our data. We list six such findings by way of illustration, summarizing them in Table 1 as an advance organizer.

Table 1. *Predictions and observations for the special-processing hypothesis*

Prediction	Observation
Forgetting rates are inversely related to output positions in criterion queues	Forgetting rates are nonmonotonically related to output positions in criterion queues
The triage effect should emerge in late childhood or early adolescence	The triage effect emerges in very early childhood
Output positions in consecutive criterion queues should be positively related	Output positions in consecutive criterion queues are nonmonotonically related
Output positions in consecutive precriterion queues should be positively related	Output positions in consecutive precriterion queues are nonmonotonically related
Triage curves should flatten across trials	Triage curves sharpen across trials
Error-success and preexperimental frequency measures of memory strength should be related to recall order in opposite ways	Error-success and preexperimental frequency measures of memory strength are related to recall order in the same way

1 *Forgetting rates.*

In Experiments 7 to 11, retention tests were administered two weeks after acquisition Since there were no further study opportunities after the acquisition criterion was achieved, the output sequence on the last criterion test was necessarily a *final memory strength ordering* from the standpoint of the special-processing hypothesis If the hypothesis is correct in assuming that this sequence was stronger to weaker, it is constrained to predict a negative relationship between criterion recall positions and forgetting rates across retention intervals Very strong items (early positions) should be forgotten more slowly than moderately strong items (middle positions), which in turn should be forgotten more slowly than very weak items (terminal positions) But if the hypothesis is wrong and the criterion output sequence was weaker to stronger to weaker, criterion recall positions will be nonmonotonically related to forgetting rates, with the largest amounts of forgetting for words that were recalled first and last The latter pattern was obtained in our experiments Aggregate curves from Experiments 7 to 11 appear in Figure 3 Since recall was perfect by the end of acquisition, error rate during the retention session is the forgetting measure in these curves

2 *Early emergence*

For the special-processing hypothesis to work, one must assume that error feedback causes subjects to activate mnemonic strategies and focus them on error-producing items Under this assumption, test trial errors will be more beneficial for subsequent learning than test trial successes, an outcome that occurs in adult recall (for a review, see Halff, 1977) But when the

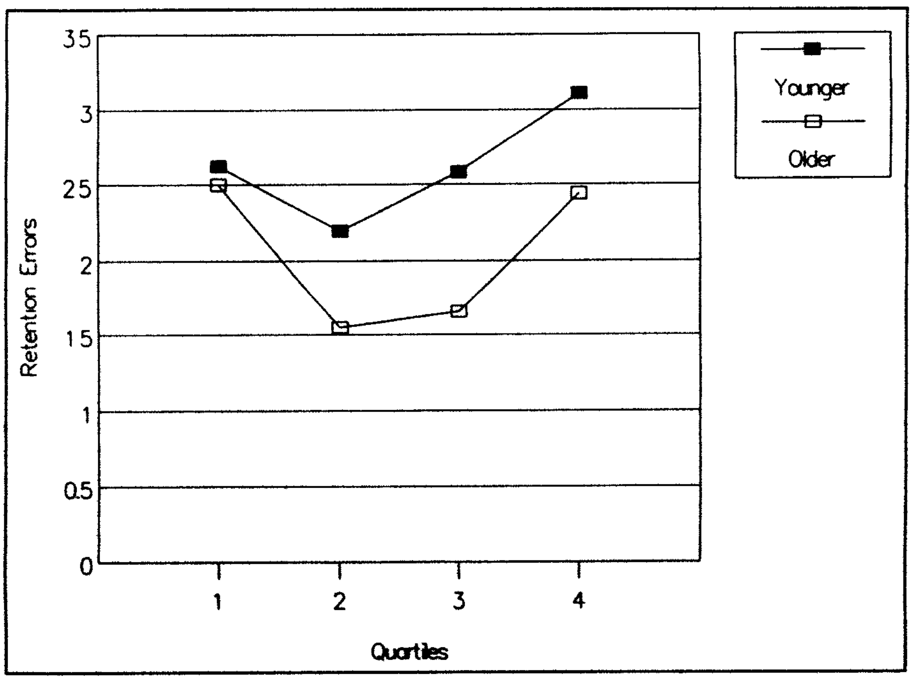

Fig. 3. Average recall error probability per item during long-term retention sessions (forgetting rate) as a function of the items' positions in acquisition recall queues (Vincentized quartiles).

assumption is studied with children, the results are different in the sense that error-driven special processing does not seem to occur before early adolescence (e g., Brainerd & Howe, 1980, Brainerd et al , 1984) This result is in good agreement with the larger literature on children's mnemonic strategies, which shows that spontaneous use of mnemonics such as rehearsal, elaboration, and subjective organization in recall is a post-childhood phenomenon (for a review, see Bjorklund & Muir, 1988) But the result is clearly not in agreement with the special-processing hypothesis because the nonmonotonic relationship between recall order and recall accuracy is observed in children as young as age 6

3 *Criterion recall orders*

Because all items are recallable at criterion, there is no error information to initiate special processing So, the output sequence on any criterion test will be a final ordering with respect to memory strength, which implies a monotonic, positive relationship between items' recall positions on adjacent criterion tests But if the output sequence is weaker to stronger to weaker, this relationship should be nonmonotonic, specifically an inverted U function Because items recalled at the start and finish of Test i are alike in memory strength, they will tend to exchange positions on Test $i + 1$, whereas items recalled in the middle of Test i will tend to stay there on Test $i + 1$ To examine these possibilities, we plotted Criterion #1 recall order against Criterion #2 recall order in Experiments 7–11 We obtained the inverted U predicted by the weaker to stronger to weaker hypothesis throughout the 6 to 13 age range The third author has observed the same pattern in the criterion protocols of aged subjects (data from Howe, 1988)

4 *Precriterion recall orders*

A similar difficulty arises with the relationship between recall orders on consecutive precriterion tests On adjacent Tests i and $i + 1$, suppose we restrict attention to items that are recalled on both tests For the same reason as in Point #3, special processing predicts a positive relationship between items' recall positions on the two tests Also for the same reason as in Point #3, weaker to stronger to weaker predicts the same nonmonotonic relationship as for adjacent criterion tests We plotted recall orders on adjacent precriterion tests using the Trial 2 to 5 data of Experiments 2 to 11 The inverted U relationship forecast by weaker to stronger to weaker was consistently observed

5 *Intertrial sharpening.*

If errors initiate special processing of error-producing items, then weaker items (higher error rates) are, in general, being strengthened more than stronger items (lower error rates). Hence, the "strength spread" between items recalled at different positions contracts as trials accumulate It follows that the relationship between recall order and any other measure of strength, whatever that relationship is, will become less pronounced across trials because strength differences are becoming harder to detect In fact, however, we have seen that the order/accuracy relationship sharpens across trials.

6 *Preexperimental memory strength.*

We conducted a control experiment with fifth and sixth grade children in which a preexperimental measure of memory strength was adopted The subjects were administered four study-test cycles with lists composed of half high-frequency nouns and half low-frequency nouns, the nouns having been equated on other difficulty factors The assumption, of course, was that high-frequency words have greater memory strength than low-frequency words The memory strengths of words recalled on Trials 2 to 4 could be classified in two ways. (a) success or error on the previous test and (b) high frequency or low frequency The special-processing hypothesis expects that these two strength measures will be related to recall order in opposite ways—words that produced an error on the previous test will again be recalled before words that produced a success, but high-frequency words will be recalled before low-frequency words On the other hand, the weaker to stronger to weaker ordering implies that the two strength measures will be related to recall order in the same way We observed the second result.

DELIBERATE STRATEGY?

If weaker to stronger to weaker is accepted as a working hypothesis, it is natural to ask whether this ordering is the result of some conscious, deliberate strategy or whether it arises from nonstrategic retrieval mechanisms We have examined this issue in adults because such subjects are both most likely to use deliberate strategies and most able to articulate the nature of their strategies We have studied two questions, in particular. Is the triage pattern present in adult recall and, if so, are adults aware of what they are doing?

The first question is pertinent because the triage pattern may be a by-product of immature memory systems, and it could be replaced by some other pattern (e g , stronger to weaker) in adults Consequently, we analyzed data sets from earlier free recall experiments with adult samples, some published (e.g., Howe, 1988) and others unpublished We also analyzed unpublished data supplied to us by C. Camp of the University of New Orleans The results were qualitatively similar to those we have reported for children The U curves that were obtained when recall order at criterion was plotted against graded memory-strength measures resembled the symmetrical curves for adolescents in Figures 1 and 2

The second question was studied by administering a memory questionnaire to 147 undergraduates The questionnaire described a typical free recall study in which the task was to remember as many of the studied items as possible. The subjects answered 10 questions about this hypothetical experiment. Seven of the questions were distractors that dealt with irrelevant factors (e.g., the physical appearance of a memory laboratory). The fourth, seventh, and tenth questions bore on the triage effect. Question #4 asked what strategies respondents might use to maximize output on a free recall test. Nearly all of the reported strategies were semantic (e g., associative or taxonomic relatedness), phonological (e.g., rhyming or other sound patterns), or visual (e.g., spatial positions or images). No subject reported a serial order strategy based on item difficulty. Question #7 asked whether recall order would be affected by prior knowledge that some items were more likely to produce errors than others Only a small proportion (19%) said that this knowledge would influence them Most subjects stated that they would recall words in whatever order they came to mind or that they would rely on semantic, phonological, or visual strategies Question #10 asked subjects to imagine that they possessed knowledge of relative item difficulty They were then asked which of the following orders would produce the best recall easy to hard, hard to easy, easy to hard to easy, hard to easy to hard They were also asked to explain their choice The results showed that 72% chose easy to hard, 21% chose hard to easy, 5% chose easy to hard to easy, and 2% chose hard to easy to hard Most of the subjects who selected easy to hard gave sensible rationales (e.g , "it's better to warm up with the easy ones to get your brain going"), and so did most of the subjects who selected hard to easy (e g , "get the hard ones out before they disappear") None of the subjects who selected the other sequences gave intelligible explanations

To sum up, adults exhibit the same sort of nonmonotonic relationship between recall order and measures of memory strength as children do The relationship does not seem to be under the control of a conscious strategy because adults are unaware of what they are doing Our best evidence, then, is that the weaker to stronger to weaker sequence is a fundamental attribute of retrieval rather than a curious consequence of strategic intervention The reader will have noticed that this conclusion comports well with two other findings that we have reported The triage effect first appears at age levels that are far below those at which children display conscious application of mnemonic strategies (early adolescence in most instances), and the effect appears at the very start of free recall, before there has been much opportunity for children to sample potentially helpful strategies

Acknowledgments—Preparation of this paper and the experiments that we report were supported by a grant to the first author from the Spencer Foundation Portions of the research were also supported by grants to the first and third authors from the Natural Sciences and Engineering Research Council

REFERENCES

Battig, W F , & Montague, W E (1969) Category norms for verbal items in 56 categories *Journal of Experimental Psychology Monograph, 80* (No 3, Pt 2)

Bjorklund, D F , & Muir, J E (1988) Children's development of free recall memory Remembering on their own *Annals of Child Development, 5* 79–123

Brainerd, C.J (1983) Structural invariance in the developmental analysis of learning In J Bisanz, G L Bisanz, & R V Kail (Eds), *Learning in children* (pp 1–36) New York Springer-Verlag

Brainerd, C J , & Howe, M L (1980) Developmental invariance in a mathematical model of associative learning *Child Development 51*, 349–363

Brainerd, C J , Howe, M L , Kingma, J , & Brainerd, S H (1984) On the measurement of storage and retrieval contributions to memory development *Journal of Experimental Child Psychology, 38*, 454–477

Brainerd, C J , Kingma, J , & Howe, M L (1986a) Spread of encoding and the development of organization in memory *Canadian Journal of Psychology, 40* 203–233

Brainerd, C J , Kingma, J , & Howe, M L (1986b) Long-term memory development and learning disability Storage and retrieval loci of disabled/nondisabled differences In S J Ceci (Ed), *Handbook of cognitive, social and neuropsychological aspects of learning disabilities* (pp 161–184) Hillsdale, NJ Erlbaum

Brainerd, C J , & Reyna, V F (1990) Can age × learnability interactions explain the development of forgetting? *Developmental Psychology 26*, 194–203

Brainerd, C J , & Reyna, V F (in press) Acquisition and forgetting processes in normal and learning-disabled children A disintegration/reintegration theory In J Obrzut & G Hynd (Eds), *Advances in the neuropsychology of learning disabilities* New York Academic Press

Ceci, S J , Ross, D F , & Toglia, M P (1987) Suggestibility in children's memory Psycholegal implications *Journal of Experimental Psychology General 116*, 38–49

Ceci, S J , Toglia, M P , & Ross, D F (Eds) (1987) *Children s eyewitness testimony* New York Springer-Verlag

Halff, H M (1977) The role of opportunities for recall in learning to retrieve *American Journal of Psychology, 90* 383–406

Howe, M L (1988) Measuring memory development in adulthood A model-based approach to disentangling storage-retrieval contributions In M L Howe & C J Brainerd (Eds), *Cognitive development in adulthood* (pp 39–64) New York Springer-Verlag

Jahnke, J C , Davis, S T , & Bower, R E (1989) Position and order information in recognition memory *Journal of Experimental Psychology Learning Memory, and Cognition, 15* 859–867

Marbe, K (1901) *Experimentell-psychologische untersuchungen uber das urtiel* Leipzig Engelmann

Osgood, C E (1953) *Theory and method in experimental psychology* New York Oxford University Press

Paivio, A , Yuille, J C , & Madigan, S A (1968) Concreteness, imagery, and meaningfulness values for 925 nouns *Journal of Experimental Psychology, 76* No 1, Pt 2

Thumb, A , & Marbe, K (1901) *Experimentelle untersuchungen uber die psychologischen grundlagen der sprachlichen analogiebildung* Leipzig Engelmann

Toglia, M P , & Battig, W F (1978) *Handbook of semantic word norms* Hillsdale, NJ Erlbaum

(Received 11/13/89, Revision accepted 3/5/90)