Psychol Res (1995) 57: 220–228 © Springer-Verlag 1995

ORIGINAL ARTICLE

Pierre Perruchet · Nicole Frazier · Jacques Lautrey

Conceptual implicit memory: A developmental study

Abstract The widely accepted standpoint that implicit memory emerges earlier in development than explicit memory, and is more stable from childhood to adult age, is based on experimental data essentially collected in perceptual tasks. The present study was aimed at investigating whether these findings still hold when a more conceptual task is used. We compared the performance of children at two age levels (2nd and 4th grades) on a category-exemplar generation task. Results showed that performances of the two groups were comparable when the target items were typical of their categories, as in Experiment 2, and for a subset of the items in Experiment 1. However, the older children outperformed the younger children in Experiment 1 when the items selected were atypical of their categories. Interpretations of these findings are discussed.

Introduction

The relative stability across the life span of performance in implicit memory tasks is a well-documented phenomenon. Although the majority of studies bear on elderly subjects (e.g. Russo & Parkin, 1993), experiments focusing on the other end of the life span, which constitutes our main concern here, also offer a consistent picture of this stability. To our knowledge, all published studies have reported an absence of reliable differences in the amount of priming from children of about 3 years to young adults (Blanc, 1988; Carroll, Byrne, & Kirsner, 1985; DiGiulio, Seidenberg, O'Leary, & Raz, 1994; Ellis, Ellis, & Hosie, 1993; Greenbaum & Graf, 1989; Lorsbach & Worman, 1989; Naito, 1990; Parkin & Streete, 1988; Perrig & Perrig, 1993; Wippich, Mecklenbräuker, & Brausch, 1989). This invariance strongly contrasts with the marked improvement of explicit memory performance with age. The improvement of explicit memory has been evidenced long ago (e.g. Dempster, 1981), and has been confirmed in the studies cited above that included recall (e.g. Naito, 1990) or recognition (e.g. Carroll et al., 1985) tests as controls. The resulting dissociation between implicit and explicit forms of memory has been viewed as powerful support for the notion of multiple, independent memory systems (e.g., Mitchell, 1993).

However, it is not clear whether developmental invariance may be generalized to any implicit measures of memory. As has often been noted, most implicit memory tasks involve mainly perceptual, or "data-driven" processing (Jacoby, 1983), in contrast with conceptual, or conceptually driven, processing which is primarily tapped in explicit tests of memory. Reflecting this general tendency, most of the tests used in the developmental studies of implicit memory cited above tap data-driven processing. That is the case for picture completion (Carroll et al., 1985; Lorsbach & Worman, 1989; Parkin & Streete, 1988), picture identification (Perrig & Perrig, 1993), face identification (Ellis et al., 1993), and word-completion tasks (DiGiulio et al., 1994; Naito, 1990).

It is quite possible to design implicit tests that rely more on the encoded meaning of concepts than on the perceptual record of the items. The prototype of these tests is the category-exemplar generation test, in which subjects are asked to name the first exemplars of a given semantic category that come to mind. This test is implicit, in so far as subjects are not instructed to

Preparation of this article was supported by the Centre National de la Recherche Scientifique (URA CNRS 316 and 1838), the Université de Bourgogne, and the Université René-Descartes

P. Perruchet (✉)
Université de Bourgogne, LEAD, Faculté des sciences, 6 Bd Gabriel, F-21000 Dijon, France; e-mail: perruche@satie.u-bourgogne.fr

N. Frazier · J. Lautrey
Laboratoire de Psychologie différentielle, Université René-Descartes, Paris, France

remember the items that were displayed in an earlier phase of the experiment, but nevertheless is conceptually driven, in so far as generating exemplars requires semantic processing of the material. Systematic investigations, essentially carried out by Roediger and his colleagues (e.g. Roediger, Weldon, & Challis, 1989) have shown that performance on a exemplar-production test may be sensitive to variables, such as the depth of processing, that were hitherto believed to affect only explicit tests of memory. Can the potential confounding between the nature of the instructions (implicit vs. explicit) and the nature of the operations predominantly involved at test (perceptual vs. conceptual) account for the dissociation observed with regard to age-related changes?

Empirical data are still inconclusive on this question. The only developmental study of implicit memory that has tapped conceptually driven processing is the Greenbaum and Graf (1989) experiment, in which children were asked to name category exemplars. As was mentioned above, the authors concluded that there is developmental invariance of performance. However, the reliability of this conclusion may be questioned on a twofold basis. First, descriptive data showed a moderate improvement in performance between 4-year- and 5-year-old children (proportion of target words produced: 23% and 30%). The lack of significant difference may be imputed to the small size of samples (12 subjects per group). Second, in a later paper, Graf (1990, p. 356) mentions briefly that he observed a striking age-related increase in priming when he used adult-normed instead of child-normed material, or when the study phase involved a non-semantic study task instead of a semantic study task.

There are also some theoretical arguments that lend favour to a positive response to the question raised above. Indeed, among the various factors supposed to play a role in the improvement of explicit memory with age, it seems likely that at least some of them also affect conceptual implicit memory. A first example is provided by the factors surrounding the notion of metamemory – that is, the knowledge that children have of the properties and limitations of their own memorization processes, and of the strategies that help to overcome these limitations. The stream of research initiated by Flavell, Beach, and Chinsky (1966) on this point has demonstrated that strategies such as verbal rehearsal or categorization of items to be recalled, are rarely spontaneously used by children before the age of seven or eight (Flavell & Wellman, 1977). Now, it has been shown that metamemory processes, such as those that elicit the organization of the study material, may improve conceptual priming (e.g. Rappold & Hashtroudi, 1991). It follows logically that when study conditions allow metamemory processes to come into play during the encoding phase, a developmental change should be observed in a subsequent conceptual implicit memory test.

Likewise, the developmental improvement of explicit memory is often related to the evolution of knowledge structures in semantic memory. The well-known experiment of Chi (1978) illustrates this point. She compared the memory span of adult novices at chess and relatively expert children. When these two groups were compared on digit span, the adults performed better, but when they were compared on recall of chess pieces, the children performed better. More generally, the knowledge-based explanation is that memory performance improves because, during the course of development, children pass from the state of universal novices to the state of more or less experts in most domains of knowledge (see Peverly, 1991, for a discussion of this position). Although they had no empirical supporting evidence, Naito and Komatsu (1993) have hypothesized that existing knowledge structures may also influence the priming system under some conditions.

The following experiments were designed to investigate the developmental changes in performance in a category-exemplar-generation task. Our primary interest was to investigate the possible influence of changes in knowledge structures. Because there is a large body of evidence that a considerable evolution in knowledge structures occurs between the ages of 7 and 10, we chose to explore the differences between children of the 2nd and 4th grades. One such evolution in this age range concerns the hierarchical organization of categories, such as is evidenced in Piaget's work. In order to explore the role of the hierarchical organization of categories in the exemplar-generation test, subjects were submitted to Piagetian inclusion tests, as described below.

Experiment 1

The development of the notion of inclusion was originally investigated by Piaget through what have become standard tasks in which children are requested to compare extensions of the class and the subclass. For example, the child is presented with 10 daisies and 2 roses and the experimenter asks, "Are there more flowers or more daisies?" and "Why?." According to Piaget and Inhelder (1959), the correct response, which appears at the age of 7–8, is the criterial measure of the existence of a cognitive organization that is isomorphic to the additive composition of classes. Later studies, of Voelin (1976), Markman (1978), Bideaud and Lautrey (1983), have demonstrated that the correct response given in this standard situation relies more on an empirical information processing than on logical reasoning. If the experimenter asks a child who has given the correct response and an adequate justification to the standard question: "Can we do something so that there are more daisies than flowers, or not?," up to the age of 10, many children reply "Yes we can do something... You have to add more daisies" or "You have to take some of the flowers away." This *modification task* can thus be considered to be a stronger criterion of the logical notion of class inclusion.

These considerations led us to use *standard* and *modification* inclusion tasks to evaluate the level of hierarchical structuring of conceptual knowledge of children. In order to disentangle the effect of the hierarchical structuring of knowledge from the other age-related factors, the inclusion level was included as a factor in the experimental design. Half of the subjects at each age level failed in both standard and modification inclusion tests, and the other half succeeded in both tests.

Method

Subjects. These were 56 children from Reims elementary schools. The initial sample was larger ($N = 71$), but some subjects were eliminated after running the inclusion task, in order to obtain the same number of subjects passing and failing the inclusion tasks within each age level. The final four groups of children were: 14 grade-2 failing (mean = 7:9 years, range = 7:4–8:8 years), 14 grade-2 passing (mean = 7:7 years, range = 7:5–8:4 years; 14 grade-4 failing (mean = 10.0 years, range = 9.6–10.9 years; 14 grade-4 passing (mean = 9.9 years, range 9.4–10.3 years).

Materials. For the inclusion tasks, the material consisted of a set of flowers (daisies and roses) and a set of beads (green and blue) drawn on a 6 × 7.5-cm index card. For the category-production task, the categories excluded flowers and colours, in order to prevent interference between the inclusion test and the implicit memory test. We used 28 other categories, and one exemplar was selected within each category to be displayed during the study phase. For illustration, Appendix 1 reports the part of this material concerning the categories that were also used in Experiment 2. Each exemplar was shown on a 7.5 × 12.5-cm index card.

Procedure. Each child was tested individually in a quiet room during two sessions, separated by a two- or three-week interval.

The first session was devoted to assessing the level of inclusion. Three items were proposed with each material (flowers and beads). The first two items consisted of the standard problem of inclusion with two different extensions for the subclasses (10/2 and 4/4) and the third item was the so-called modification problem. The order of presentation, flowers before beads and standard problems before modification problems, was the same for all children. The inclusion questions were worded as was indicated in the Introduction. An item was judged correct when both the response and the justification were correct.

The second session was devoted to the implicit memory task. During the study phase, the experimenter showed each word, and simultaneously read it aloud. In order to ensure that the word item was sematically processed, the children were asked a question involving this word. For instance, when the target item was "Friday," the question was: "Do you go to school on Friday?." The question never involved the name of the relevant category (for instance, questions such as: "Is Friday a day of the week?" were prohibited). Within each of the four experimental groups, half of the subjects were presented with 14 out of the 28 target items, and the other half were presented with the other part of the list.

Immediately after study, the category-production task was given. Subjects were asked to name the first exemplar of a given category that came to mind. For instance, they were asked: "What is the first day of the week that comes to your mind?." This general procedure was used for each of the 28 selected categories, with old and new categories randomly intermixed. The categories for which an exemplar was shown during the study phase (old) served to assess the rate of production of the primed items, whereas the other categories (new) provided a baseline measure. The rate of priming was assessed as the difference between the two scores.

After the completion of this task, the children were tested for explicit remembering of the exemplars displayed during the study phase. Each of the 14 category names presented earlier was used as a cue for explicit recall. For instance, subjects were asked: "A moment ago, I showed you a day of the week. Do you remember which day it was?."

Results

Category-production task

An analysis of variance (ANOVA) with three between-subjects factors (Lists, Age, and Level of Inclusion) and one within-subjects factor (Status of the test item: old vs. new) was performed on the proportion of target items produced in response to category names. The main effect of List was significant, $F(1,48) = 7.68$, $p = .008$, and likewise the interaction of List with Status of items, $F(1,48) = 4.77$, $p = .034$. These results provide evidence that the two lists were not homogeneous in their ability to elicit a priming effect. However, neither the interaction of Age with List and Status of items nor the interaction of Inclusion level with List and Status of items reached significance (respectively: $F < 1$, and $F(1,48) = 1.21$, $p > .10$). This indicates that the same pattern of results was obtained for the phenomena of interest (namely: the priming effect as a function of Age and Inclusion level) for the two lists. As a consequence, the data were pooled on the two lists in Table 1, which shows the proportion of target items named during the test as a function of the other factors.

The percentage of responses matching the target items was consistently higher for old than for new items (respectively: 30.1% and 14.4%), thus providing evidence for a reliable priming effect, $F(1,48) = 33.65$, $p < .0001$. The main effect of Age was also reliable, $F(1,48) = 4.93$, $p = .03$, and, more interestingly, the interaction of Age with Status of item reached significance (although only at the conventional significance threshold, $F(1,48) = 3.99$, $p = .05$). Inspection of Table 1 reveals that the priming effect is stronger for old than for young subjects (respectively: 21.1% and 10.3%). Both groups of subjects showed similar performances for the new items (baseline), whereas the older subjects produced more primed items than the younger subjects. Planned partial comparisons confirmed this conclusion (respectively: $F < 1$, and $F(1,48) = 5.37$, $p = .025$).

As Table 1 shows, the level of inclusion had no effect on word production. The ANOVA confirmed that the main effect of Inclusion Level, and likewise all the interactions involving the Inclusion Level as a factor, failed to reach significance, $F_{max} = 1.26$, $p > .10$.

Table 1 Performance in the exemplar-generation task and in the cued-recall test as a function of children's groups in Experiment 1 (%)

Subjects		Exemplar Generation					Cued Recall	
Grade	Inclusion test	Old		New		Priming	% Correct	
		M	*SD*	M	*SD*		M	*SD*
2nd	passing	25.5	16.3	17.9	13.1	7.6	58.7	20.2
	failing	23.7	17.2	10.7	8.7	13.0	57.7	16.0
4th	passing	35.7	16.6	13.3	8.8	22.4	72.4	11.8
	failing	35.7	19.8	15.8	11.6	19.9	74.0	12.7

Cued recall

An ANOVA with three between-subjects factors (Lists, Age, and Level of Inclusion) was performed on the proportion of target items correctly recalled when the category name was given as a prompt. In contrast to the performance on the category-production test, the main effect of List, as well as all interactions involving this factor, failed to reach significance, $F\text{max} = 2.76$, $p > .10$. The data were pooled on lists for the subsequent analyses.

The data reported in Table 1 show that the pattern of cued-recall performance is strikingly similar to the pattern observed on implicit scores. Older subjects recalled more items than younger subjects ($M = 73.2$ vs. 58.2, $F(1,52) \sim 13.13$, $p = .0007$), and this result was obtained whatever the Inclusion Level, as confirmed by the non-significant interaction between Age and Inclusion Level ($F < 1$). There was again no main effect of the Inclusion Level ($F < 1$).

Discussion

The effect of age on the cued-recall test could be anticipated on the basis of the available literature, the age-related improvement of explicit memory being a well-documented phenomenon.

The findings on the category-production test are more informative. We found no effect of the inclusion level, and a reliable effect of age, with older children exhibiting more priming than younger children. The lack of influence of the level of inclusion provides some guarantees that differences in the implicit category-production test cannot be considered to result from differences in the amount of knowledge about the logical principles that underlie conceptual organization. On the other hand, the reliable effect of age suggests that the developmental invariance of priming effects reported hitherto could be due to the predominant use of implicit-memory tests primarily relying upon perceptual processes. However, before expanding upon this interpretation, it seemed advisable to strengthen further the empirical basis of the phenomenon at hand. Experiment 2 was intended to replicate the results of Experiment 1 with another sample of subjects, and with better-controlled experimental material.

Experiment 2

In Experiment 1, the target exemplar of each category was selected somewhat arbitrarily, given that there is no children-normed category-association table available in French. The words were selected only on the basis of their probable inclusion within the children's lexicon, and in fact the responses to questions asked during the study phase confirmed that all the words were known by children. However, a common practice within the implicit-memory area consists in selecting the target items so that their baseline rate – that is, their rate of production under unprimed conditions, is neither too low nor too high, in order to prevent any floor or ceiling effects.

In an attempt to follow this practice, despite the lack of available norms, we performed a pilot study on a sample of 23 children, selected from grade-3 classes of a Reims elementary school. Children were tested individually. The name of a semantic category was given both orally and in the written form, as during the experiments, and the children were asked to name all the exemplars of this category that came to mind. They were given a maximum of 1 min, but the next trial was given before the deadline if children claimed to have no more responses. This procedure was repeated for a set of 22 categories. Four categories were then removed from the experimental set, either because they were represented by a single, or very few exemplars, or, on the contrary, because no consistency in the pattern of responses emerged. For the 18 selected categories, the target exemplars for Experiment 2 were chosen among the items that were produced by at least 5 out of the 23 children (but these were in no cases the most frequent responses).

There were two other differences between Experiments 1 and 2. First, given that the level of inclusion had no effect in Experiment 1, this factor was no longer included in the design of Experiment 2. Subjects were still submitted to inclusion tasks, but no attempt was

Table 2 Performance in the exemplar-generation and word-stem completion tasks as a function of age level in Experiment 2 (%)

		Exemplar Generation			Word-stem Completion		
		Old	New	Priming	Old	New	Priming
2nd grade	M	41.4	10.2	31.2	23.9	12.4	11.5
	SD	23.8	10.8		21.0	14.2	
4th grade	M	35.5	9.0	26.5	18.4	10.3	8.1
	SD	24.1	8.3		18.6	12.1	

Table 3 Performance in the exemplar-generation Tests as a function of the level of typicality of items for each age level (%)

	Experiment 1			Experiment 2		
	Old	New	Priming	Old	New	Priming
2nd grade						
Typical	46.7	27.1	18.6	46.9	9.0	37.9
Atypical	2.9	4.3	– 1.4	36.3	12.3	24.0
4th grade						
Typical	48.6	27.1	21.4	40.0	8.8	31.2
Atypical	17.1	2.9	14.2	29.6	9.4	20.2

Note: Item typicality was assessed in a separate pilot study. For Experiment 1, these results apply only to the 10 categories included in this pilot study. For Experiment 2, the attentive reader may note that when averaged across the levels of typicality, the data do not exactly fit those reported in Table 2. This is due to the fact that there were nine typical and nine atypical items, and that a given subject saw either five typical and four atypical or the reverse. The resulting discrepancies between the Tables (which are quite minor and do not affect the conclusions in any way) are due to differences of weighting

made to equalize the group sizes. This avoids a potential bias in the selection of subjects, which may undermine the conclusions pertaining to the age factor. Second, the explicit cued-recall test that followed the category-production task in Experiment 1 was removed, and replaced by a word-stem completion task. This was intended to provide evidence of a dissociation between conceptual and perceptual implicit memory tests within a single experimental arrangement.

Method

Subjects. Fifty-two new children from the same pool as for Experiment 1 served as subjects. They were 26 grade-2 (mean = 7:6 years, range = 7:2–8:1 years) and 26 grade-4 (mean = 9:7 years, range = 9:1–10:1 years).

Material and procedure. Children were submitted successively to the category-production task, the tests for level of inclusion, and the word-stem completion task. The three tasks were performed within a single session, in the same order for all subjects.

For the first two tasks, the only change from Experiment 1 (in addition to the running order) pertained to the choice of categories and exemplars in the production task. As was mentioned above, the 18 items were selected from a pilot study. A large sample of these items is provided in Appendix 1.

The word-stem completion test was performed on the same list of words used for the category-production task. (Care was taken that all the words should begin with a different syllable, and that several words could be generated from each auditive stem. As assessed from the phonetic description of words in a standard dictionary, the syllables could be completed to form at least eight words.) During the test, the first syllable of the 18 items was read to children through a tape-recorder. Subjects were asked to complete each syllable by the first word that came to mind.

Results

Category-production task

An ANOVA with two between-subjects factors (Lists and Age) and one within-subjects factor (Status of the test item: old vs. new) was performed on the proportion of target items produced in response to category names. There was no main effect of List, and no interaction between List and the other factors (Fs < 1). The data were therefore pooled on the two lists for subsequent analyses.

Table 2 shows the proportion of target items named during the test as a function of the other factors. The scores were consistently higher for old than for new items ($M = 38.5\%$ and 9.6%), thus providing evidence for a strong priming effect, $F(1, 50) = 56.49$ $p < .0001$. However, in contrast to the results from Experiment 1, neither the effect of Age, nor the interaction between Age and the Status of items reached significance, $F\text{max} = 1.20$, $p > .10$. The priming effect was even slightly higher for younger than for older children (respectively: 31.2% vs 26.5%).

Experiments 1 and 2 essentially differed by the choice of items. A possible explanation for the discrepancy in the pattern of results was that the items selected differed with regard to their a-priori probability of production as an exemplar of the categories. A way of exploring this hypothesis consisted in investigating whether the priming effect differed as a function of the typicality of items, as assessed during the pilot study. An index of typicality was computed for each word as the ratio between the number of occurrences of the target word and the number of occurrences of the prototype (i.e., the word the most often produced in response to the same category name). For instance, the prototype of the category "sport" was "Tennis" (8 occurrences), and the selected exemplar was "Basket" (5 occurrences). This item was assigned a typicality score of 5/8, i.e.: .625. The scores go from .28 to .82 as a function of items.

The next step in the analysis consisted in ordering the items according to these scores, then dividing the list into two parts. A new analysis was run, with Age as a between-subjects factor, and Status of items and Typicality as within-subjects factors. Detailed results are displayed in Table 3. The priming effect was reliably stronger for the more typical (34.5%) than for the less typical items (22.1%), as is shown by the significant interaction of Status of item with Typicality, $F(1, 50) = 6.05$, $p = .017$, an effect that cannot be imputed to a metrical artefact stemming from differences in baseline. Indeed, the difference in the baseline level did not differ significantly for typical and atypical items, and it was even slightly higher for atypical than for typical items (respectively: 10.9% and 8.9%). However, this effect is the same for young and old children, as is attested by the non-significant three-way interaction between Age, Status of item, and Typicality ($F < 1$). At first glance, this result invalidates the hypothesis that differences in typicality of target items may account for the differences between the results of Experiments 1 and 2. We postpone to the Discussion section an evaluation of how well this hypothesis may work, despite this empirical counterargument.

A final analysis dealt with the effect of the inclusion level. As was mentioned above, no attempt was made to orthogonalize Age and Inclusion level. It turned out that most of the grade-2 subjects failed the inclusion tasks (19 out of 26), hence making unfeasible any analysis of the inclusion level for this group. However, grade-4 subjects were more variable: 10 children failed the two tasks, 5 children succeeded in the empirical inclusion test, but failed the logical inclusion task, and 11 subjects were successful in the two tasks. A 3×2 ANOVA was performed, with Level of Inclusion as a between-subjects factor and Status of item as a within-subjects factor. Neither the main effect of the Level of Inclusion, nor the interaction between the two factors reached significance (Fs < 1).

Word-stem completion

An ANOVA with two between-subjects factors (List and Age) and one within-subjects factor (Status of the items) was performed on the proportion of word stems completed as study words. As is shown in Table 2, more stems were completed as the target items when these items were studied than when they were not studied ($M = 19.8\%$ vs 13.1%, $F(1, 48) = 15.50$, $p = .0003$). However, there was a significant effect of List, $F(1, 48) = 4.47$, $p = .039$, and a significant interaction of List with Status of items, $F(1, 48) = 59.3$, $p < .0001$, demonstrating that the two lists did not elicit the same amount of priming.

Neither the main effect of Age, nor the interaction of Age with the other factors reached significance ($F\text{max} = 1.85$). This replicated the well-documented developmental invariance of implicit memory as assessed through a perceptual task. As for the category-production task, a 3×2 ANOVA was also performed on the grade-4 group, with Level of Inclusion as a between-subjects factor and Status of item as a within-subjects factor. Neither the main effect of the Level of Inclusion nor the interaction between the two factors reached significance, $F\text{max} = 2.21$, $p > .10$.

It is worth recalling that the word-stem completion test was performed on the same items as the category-production task. If, as we expected, an effect of age on the category-production task were obtained, the fact of not obtaining an effect of age in the word-stem completion task would have strengthened a position stressing the independence between the processes at hand. Admittedly, in the present context, this particular design somewhat undermines our conclusions on the word-stem completion test. However, this design offers an opportunity for testing another form of independence, namely stochastic independence (Tulving, Schacter, & Stark, 1982). Two indexes were computed for each word, revealing the amount of priming each word elicits in the conceptual test on the one hand, and on the perceptual test on the other hand. The Pearson correlation between the two sets of measures across the 18 words turned out to be near zero ($r = -.05$). Therefore, previous exposure to a word may increase the likelihood of its evocation in a subsequent category-production test and not in a subsequent word-completion test, and conversely. This independence appears despite the fact that the production of a study item in the category-production task may have strengthened further the probability of completing the relevant word stem by the same item.

General discussion

This study replicated two well-documented findings, namely, an improvement in explicit memory (Experiment 1), associated with stability in perceptual priming

(Experiment 2), with increased age. However, the results of the two experiments notably differed on our main point of interest: while Experiment 1 showed evidence of a clear increase in conceptual priming in a category-exemplar generation task with age, Experiment 2 failed to replicate this effect. What causes this apparent discrepancy?

The two experiments differed mainly in the choice of items. The target exemplars were chosen arbitrarily in Experiment 1, whereas in Experiment 2 they were chosen as a function of their typicality, as assessed by a pilot study. In an attempt to explore the potential influence of this factor, we obtained evidence, in Experiment 2, that the amount of priming was far higher for items that had a high relative probability to be produced at a category prompt in unprimed conditions. But this effect was the same for young and old children, hence apparently making the differences in typicality an ill-suited candidate for an explanation of the discrepancy under consideration.

However, it is worth stressing that the variation in typicality of items selected for Experiment 2 was quite moderate. In fact, the rate of spontaneous production of typical and atypical items was nearly equivalent, in so far as, in keeping with the current practice in the field, the items were selected to keep this rate within a restricted range for all items (see Method section). The variance in the scores of typicality used above stemmed primarily from the fact that the rate of production of the target items was divided by the rate of production of the prototype of the relevant category. It is possible that the material used in Experiment 1 included items that were more extreme on the typicality dimension. A test of this hypothesis is made possible by the fact that there was some overlap in the categories used in the two experiments. Concerning these categories, our pilot study provides an index of typicality of the target exemplars selected for Experiment 1. The 10 overlapping categories were dichotomized according to the typicality of the exemplars selected for Experiment 1, with the typicality being assessed as above (see Experiment 2, Method section). The mean score of typicality for the five more atypical items was .19, and for the five more typical items .90. It is noteworthy that these are mean values, but they fall out of the range of individual typicality scores of the items selected for Experiment 2 (.28 to .82). This indicates that the variation in typicality was far greater for Experiment 1 than for Experiment 2. In addition, the rate of generation of the target exemplars differed largely between categories. It turned out that for the previous division into two sets of five categories, the classification was exactly the same, whether it was based on the relative typicality as above (i.e., the ratio between the rate of generation of the target items and of the prototype of the category), or on the rate of generation of the target items itself.

The results of this reanalysis are shown in Table 3 (left panel). Overall, the amount of priming was far higher for typical than for atypical items (20% vs. 6.4%), thereby confirming the findings from Experiment 2. However, it turned out that, in contrast to Experiment 2, the effect of typicality was stronger in the younger than in the older subjects. The amount of priming for atypical items was notably higher for older children (M = 14.2%) than for young children, who exhibited in fact no priming at all ($M = -1.4\%$). However, when the more typical items were considered, the difference in performance was considerably attenuated ($M = 21.4\%$ and 18.6%). Thus the age-related difference obtained in Experiment 1 tended to vanish when the analysis was circumscribed to a subset of exemplars for which the relative frequency of production at a category prompt in unprimed conditions was high enough.

These a-posteriori analyses are only indicative, and warrant being replicated in further experiments especially designed to manipulate the level of typicality of items, and to tease apart the effects linked to the rate of generation of the target item itself, and the effects linked to the rate of generation of the prototype of the categories. However, pending complementary data, a provisional conclusion is that whether or not developmental differences in an exemplar-generation task are observed depends on the selection of the material. If items are selected to be typical of their categories for the children, no – or minimal – age-related differences in priming emerge. However, if the material includes items that are sufficiently atypical of the categories, a reliable improvement with age occurs, a conclusion that is consonant with the earlier claim of Graf (1990). Graf (1990) briefly reported an experiment demonstrating no difference in priming between 3-, 4-, and 5-year-old children when child-normed categories were used. However, a strong improvement in performance appeared when adult-normed categories were involved ($M = 0\%$, 17%, 28%, estimated from Graf, 1990, Fig. 8).

These empirical data may be interpreted in several ways. An uninteresting account consists in attributing these phenomena to floor or ceiling effects. In particular, it could be argued that the lack of age-related differences with typical items is due to the fact that the high probability of producing the target exemplars in young children leaves no place for any further improvement in performance in older children. Although impossible to rule out definitely, this interpretation is somewhat undermined by the fact that this probability is still far from its maximum theoretical value (i.e, 1; the maximum value reached here was .41).

If we take for granted that the effect discussed here is not, or not only, a metrical artefact, we have to examine the explanatory power of the two factors suggested in the Introduction as potential inductors of a developmental trend in conceptual implicit memory, namely metamemory and knowledge organization. Although metamemory can play a role in implicit tests in some conditions (e.g. Rappold & Hashtroudi, 1991), its explanatory power appears strikingly limited in the present context. Indeed, metamemory processes are primarily engaged during encoding when people are intentionally attempting to memorize the displayed material. Because we used incidental conditions of learning in our experiments, even the older children had no reason to use their knowledge of memorization strategies during the study phase. In addition, the development of mnemonic strategies based on the structuring of the material was not easy, given that items belonged to independent categories, in order to fit the requirements of the exemplar-generation task.

A second possibility lies in the development of knowledge organization. This development may be thought of in at least two different ways, according to whether the main concern bears upon the logical principles underlying this organization or upon more content-related features. In order to disentangle these two possible sources of influence, children were submitted to several Piagetian inclusion tasks. The performance of the subjects in these tasks was unpredictive of the amount of priming in the exemplar-generation tasks in our two experiments, an argument against the involvement of the logical principles underlying the categorial organization in this case.

An alternative hypothesis is that the content of the categories changes from younger to older children. Because our assessment of typicality was derived from a pilot study carried out on children from a single sample of intermediate age, it is possible that the resulting data did not provide a reliable picture of the category content for each of the two experimental populations. This interpretation may find support in studies that show that young children tend to demonstrate underextension in category membership judgments in comparison with older children, and that they tend not to include in their categories items that older children consider to be atypical category members (Anglin, 1977, Caplan & Barr, 1989, Rosch, 1973). If at least some of the target items selected for Experiment 1 were not valid exemplars of their categories for young children, it follows that these items should fail to be activated by the category prompt, even though they had been studied previously.

In order to test this hypothesis, complementary analyses were performed on the 10 categories that were used both in the pilot study devised to assess the typicality of exemplars and in the two experiments. The main objective was to use the data collected in unprimed conditions in the two experiments to provide a separate measure of typicality for each age group. In passing, it may be noted that the results nicely replicated the results from the pilot study for comparable aspects. For instance, when assessed on the whole population, the prototypes were found to be the same in the pilot study and in the unprimed conditions of the two experiments for 8 out of the 10 categories. Furthermore, in the two residual cases, the departure was small. But more importantly for the present concern, various attempts failed to reveal any reliable difference in performance between older and younger children. So it appears that the proportion of occurrences in unprimed conditions of the atypical exemplars used in Experiment 1 was exactly the same for the two age groups (2.2%). As for the typical exemplars, the corresponding values were closely similar (31.9% and 28.9% for older and younger children respectively).

This post-hoc analysis runs counter to the view that the age-related differences observed in this study are related to a difference in the category content for the two ages. The final picture emerging from these data is that the effect on a subsequent exemplar-production task of an earlier exposure to items that are equally atypical for younger and older children is only (or at least mainly) manifest for older children. However, this conclusion has only tentative value, because the database used in the post-hoc analysis was collected in conditions in which children had to generate only one word to the category prompt. As a consequence, both atypical exemplars and items not included in the target category tend never to occur in children's productions. This line of reasoning led us to suspect that the fact that young and older children generated a very low proportion of atypical items (2.2%) is due to some kind of floor effect inherent to the method of measure, rather than to the genuine identity of the category contents, for the two groups of children. Future investigations will have to study the actual source of the age-related differences put forward in this paper, and to elucidate the kind of cognitive mechanisms underlying this developmental trend.

References

Anglin, J. M. (1977). *Word, object, and conceptual development.* New York: Norton.

Bideaud, J., & Lautrey, J. (1983). De l'inclusion empirique à l'inclusion logique. *Cahiers de Psychologie Cognitive, 3*, 295–324.

Blanc, C. (1988). Etude génétique de la mémoire implicite chez les enfants de 5 ans à 11 ans. Unpublished thesis, University of Grenoble.

Gaplan, L. J. & Barr, R. A. (1989). On the relationship between category intensions and extensions in children. *Journal of Experimental Child Psychology, 47*, 413–429.

Carroll, M., Byrne, B., & Kirsner, K. (1985). Autobiographical memory and perceptual learning: A developmental study using picture recognition, naming latency, and perceptual identification. *Memory & Cognition, 13*, 273–279.

Chi, M. T. H. (1978). Knowledge structures and memory development. In R. S. Siegler (Ed.), *Children's thinking: What develops?* Hillsdale, NJ: Erlbaum.

Dempster, F. N. (1981). Memory span: sources of individual and developmental differences. *Psychological Bulletin, 89*, 63–100.

DiGiulio, D. V., Seidenberg, M., O'Leary, D. S., & Raz, N. (1994). Procedural and declarative memory: A developmental study. *Brain & Cognition, 25*, 79–91.

Ellis, H. D., Ellis, D. M., & Hosie, J. A. (1993). Priming effects in children's face recognition. *British Journal of Psychology, 84*, 101–110.
Flavell, J. H., Beach, D. R., & Chinsky, J. M. (1966). Spontaneous verbal rehearsal in a memory task as a function of age. *Child Development, 37*, 283–299.
Flavell, J. H., & Wellman, H. M. (1977). Metamemory. In R. V. Kail & J. V. Hagen (Eds.), *Perspectives on the development of memory and cognition* (pp. 3–33). Hillsdale, NJ: Erlbaum.
Graf, P. (1990). Life-span changes in implicit and explicit memory. *Bulletin of the Psychonomic Society, 28*, 353–358.
Greenbaum, J. L., & Graf, P. (1989). Preschool period development of implicit and explicit remembering. *Bulletin of the Psychonomic Society, 27*, 417–420.
Henry, L. A., & Millar, S. (1991). Memory span increase with age: A test of two hypotheses. *Journal of Experimental Child Psychology, 51*, 459–484.
Jacoby, L. L. (1983). Perceptual enhancement: Persistent effects of an experience. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 9*, 21–38.
Lorsbach, T. C., & Worman, L. J. (1989). The development of explicit and implicit forms of memory in learning disabled and nondisabled children. *Contemporary Educational Psychology, 14*, 67–76.
Markman, E. (1978). Empirical versus logical solutions to part–whole comparisons problems concerning classes and collections. *Child Development, 49*, 168–177.
Mitchell, D. B. (1993). Implicit and explicit memory for pictures: Multiple views across the lifespan. In P. Graf & M. E. J. Masson (Eds.), *Implicit memory: New directions in cognition, development, and neuropsychology*. Hillsdale, NJ: Erlbaum.
Naito, M. (1990). Repetition priming in children and adults: Age-related dissociation between implicit and explicit memory. *Journal of Experimental Child Psychology, 50*, 462–484.
Naito, M., & Komatsu, S. (1993). Processes involved in childhood development of implicit memory. In P. Graf & M. E. J. Masson (Eds.), *Implicit memory: New directions in cognition, development, and neuropsychology*. Hillsdale, NJ: Erlbaum.
Parkin, A. J., & Streete, S. (1988). Implicit and explicit memory in young children and adults. *British Journal of Psychology, 79*, 361–369.
Perrig, W. J., & Perrig, P. (1993). Implizites Gedächtnis: Unwillkürlich, entwicklungsresistent und altersunabhangig? *Zeitschrift für Entwicklungs und Pädagogische Psychologie, 25*, 29–47.
Peverly, S. (1991). Problems with the knowledge-based explanation of memory and development. *Review of Educational Research, 61*, 71–93.
Piaget, J., & Inhelder, B. (1959). *Genèse des structures logiques élémentaires*. Neuchâtel: Delachaux et Niestlé.
Rappold, V. A., & Hashtroudi, S. (1991). Does organization improve priming? *Journal of Experimental Psychology: Learning, Memory, and Cognition, 17*, 103–114.
Roediger, H. L., Weldon, M. S., & Challis, B. H. (1989). Explaining dissociations between implicit and explicit measures of retention: A processing account. in H. L. Roediger & F. I. M. Craik (Eds.), *Varieties of memory and consciousness: Essays in honour of Endel Tulving* (pp. 3–14). Hillsdale, NJ: Erlbaum.
Rosch, E. (1973). On the internal structure of perceptual and semantic categories. In T. E. Moore (Ed.), *Cognitive development and the acquisition of language*. New York: Academic Press.
Russo, R., & Parkin, A. J. (1993). Age differences in implicit memory: More apparent than real. *Memory & Cognition, 21*, 73–80.
Tulving, E., Schacter, D., & Stark, H. A. (1982). Priming effects in word-fragment completion are independent of recognition memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 8*, 352–373.
Voelin, C. (1976). Deux expériences à propos de l'extension dans l'épreuve de quantification de l'inclusion. *Revue Suisse de Psychologie, 35*, 269–284.
Wippich, W., Mecklenbräuker, S., & Brausch, A. (1989). Implizites a und explizites Gedächtnis bei Kindern: Bleiben bei indirekten Behaltensprüfungen Altersunterschiede aus? *Zeitschrift für Entwicklungs- und Pädagogische Psychologie, 21*, 294–306.

Appendix

The exemplars used in Experiment 1 and 2 for the 10 categories common to both experiments

Category name	Exemplars	
	Experiment 1	Experiment 2
Meuble (furniture)	Lit (bed)	Table (table)
Maladie (illness)	Varicelle (chickenpox)	Varicelle (chickenpox)
Fruit (fruit)	Golden (a kind of apple)	Banane (banana)
Jouet (toy)	Poupée (doll)	Poupée (doll)
Légume (vegetable)	Carotte (carrot)	Poireau (leek)
Outil (tool)	Marteau (hammer)	Pince (pliers)
Moyen de transport (means of transport)	Clio (a popular french car)	Camion (truck)
Vêtement (clothes)	Blue Jean (jeans)	Chemise (shirt)
Instrument de musique (music instrument)	Piano (piano)	Saxophone (saxophone)
Sport (sports)	Gymnastique (gymnastics)	Basquet (basketball)