# Children's Comprehension and Memory for Stories

DOROTHY POULSEN, EILEEN KINTSCH, WALTER KINTSCH

*University of Colorado*

AND

DAVID PREMACK

*University of Pennsylvania*

Sixteen 4-year-olds and sixteen 6-year-olds were shown four picture stories consisting of 15 to 18 pictures without text. The stories were well structured, consisting of two or more causally and temporally related episodes. The children were asked to describe each picture, and, after seeing all the pictures of a story, to recall the story without pictures. The pictures were either presented in their normal order or in scrambled order. The data analysis concentrated upon the comparison between the responses in the normal condition when the children were telling a story and in the scrambled condition, when they were merely responding to the pictures as such without the story context. The results showed that even the 4-year-olds, but especially the older children, were interpreting the pictures as stories in the normal condition and that their knowledge about stories, i.e., the story schema, determined the nature of their responses. Even in the scrambled condition the 6-year-olds tried to make sense of the pictures in terms of a story by making inferences, attributing thoughts and emotions to the characters, and using narrative conventions, while the 4-year-olds often reverted to a simple labeling strategy. In recall all of these trends were emphasized. Those parts of the descriptions that were best integrated into a story were recalled best, while nonintegrated descriptions tended to be forgotten.

The way people recall stories depends to a large extent on the nature of the story: well-structured, schema-conforming stories are easy to remember, but disorganized stories that deviate from our culturally shared story schema are likely to be distorted in recall. The notion of "schema" used here is based upon an extensive body of work in anthropology, linguistics, and philosophy, concerned with the theory of story structure (for useful discussions of this literature, see van Dijk, 1972, or Grosse,

379

1977). In this work, a distinction is made between the surface expression of the text of a story and an underlying, conceptual base, the "immanent level" in the terminology of French structuralism. Immanent to all narratives are, on the one hand, the culture-specific structures of narratives, and on the other, the universal principles of human thinking in general. The first comprise the story schema proper, while the second consist of general knowledge schemata organized in terms of hierarchical frames, together with the rules for the operation of this system. In most of the work to be discussed below these two components are inextricably confounded, but we shall make some effort to distinguish between them in the experiment we report.

What, then, are the features of stories that are subsumed under the concept of "story schema"? They key ideas were proposed some time ago by the Russian linguist Propp in 1928, and were further elaborated by many others (e.g., Lévy-Strauss, 1970; van Dijk, 1972; Brémond, 1973).[1] We can only outline them here. A story is built around *actors* (or rather actor-types) and *functions,* which are major story relevant actions that change the story from one state into another (such as *departure, marriage, betrayal*). While there is a requirement for continuity in the topic actor, the actions change throughout the story. Originally, Propp, who was working with folk tales of the simplest kind, conceived of functions in a rather fixed, static manner. Later investigators stressed that the order of actions in a story could not be fixed. Instead, only the category to which an action belongs forms a fixed sequence in a story (e.g., Brémond, 1973): some action that functions as an *exposition* is followed by one in the *complication,* and eventually in the *resolution* categories. Expositions must introduce the setting and the main character; complications require a remarkable or interesting event (some unexpected twist in the course of events), and resolutions must return the story to a stable state (there must be a proper ending with no dangling or unresolved events). Furthermore, Brémond described some explicit combination rules that permit one to construct complex stories from these simple building blocks, e.g., rules of concatenation whereby stories of more than one episode may be formed, even permitting an overlap of function (so that, for instance, the resolution of one episode may at the same time function as the exposition of the next one). Another important rule that we shall use below is insertion, which allows one episode to be embedded within another, frame-like episode. For example, a complication may be expanded to form an entire episode, or even a series of episodes.

[1] The "story-grammars" of Mandler and Johnson (1977) and Stein and Glenn (1979) present an alternative to this approach in that they do not distinguish story-specific rules from general knowledge about human action, and involve a rather more detailed breakdown of stories than is given here (though its general features can be mapped into the present system).

This knowledge that people have about how stories may be constructed is called the "story schema." In comprehending an actual story, this schema is used to organize the input material, generating what we call the macrostructure of the story (Kintsch and van Dijk, 1975). Thus, the schema is general and formal, an expectation in a psychological sense; a story supplies specific content to the schema, thereby producing a macrostructure. A schema may be compared to a structure with many open slots that stand in specified relations to each other; the macrostructure is obtained by filling in these slots with labels that subsume the material from a particular story.

There is a certain amount of evidence that schemata function in this way in story comprehension. Bartlett (1932) has shown that disorganized stories that deviate from our culturally shared schema are poorly recalled. More recent work has both replicated and qualified these results by demonstrating that the distortions observed by Bartlett fail to occur when schema-conforming texts are used (Thorndyke, 1977; Kintsch & Greene, 1978). Indeed, readers have very little trouble reconstituting the proper sequence of events in a story when the paragraphs of the story are presented out of order, as long as the story fits their story schema well (Kintsch, Mandel, & Kozminsky, 1977). If it does not fit the schema, the reader will try to make it fit, thereby distorting it (Bartlett, 1932).

We take the general importance of schemata in adult story comprehension as established. Our interest is in how and when the schema enters into the psychological processes involved in story comprehension. For this reason, we decided to study children who are still in the process of acquiring a story schema. Not only were we interested in the role that the story schema plays in the child's comprehension of stories, but we wished to know something of the development of the comprehension process itself.

From his observation of 6- to 8-year-old children Piaget (1928) concluded that they were not yet capable of recalling fairly complex stories. The children made many cause and effect confusions in the texts Piaget gave them. The nature of his experimental materials, however, may have led Piaget to greatly underestimate the abilities of his subjects. One of his stories was fully as bizarre as some of Bartlett's stories, and others, though structurally well formed, were written in a cryptic style which, as Mandler and Johnson (1977) point out, tends to hide and even mislead the reader about the nature of the connections between episodes. Simply by rewriting one of these stories, Mandler and Johnson obtained a better record of recall even from first graders.

Piaget (1928) also reported that preoperational children were unable to reproduce the correct order of events in a story. Fraisse (1963) confirmed Piaget, but added that children ordered stories better than other kinds of texts. Recent findings about order show that first-graders *can* recall the

basic order of events in a story with accuracy if the story can be organized around a familiar schema (Mandler & Johnson, 1977; Stein & Glenn, 1979). The importance of schemata in the ordering task should come as no surprise after Piaget's (1968) dramatic demonstration of the change in recall of a series of sticks that occurs when a child acquires the anticipatory schema underlying the operation of seriation. Similarly, the child who has acquired a story schema can structure a story according to that schema and thereby produce an orderly recall protocol—unless the experimenter interferes by telling an inadequate story which does not fit the child's schema.

If one wants to study the importance of a schema in the comprehension of stories one needs to work with fairly complex stories, where the organization provided by the schema can be expected to make a real difference. The problem then becomes to select a suitable measure of comprehension. Since Piaget (1968) it has been known that children who do very badly in recall will perform much better when given a recognition test. Hall, Cole, Reder, and Dowley (1977) also found that $4\frac{1}{2}$-year-olds often refused to recall a story spontaneously, but that performance increased from 25% recall to 63% when probe questions were given. Recall seems to be difficult for children not because of poor comprehension, but in part because of a lack of expository skills (Brown, 1975), and in part because of inappropriate control processes. Children have trouble with the notion of intentional recall, especially in a laboratory situation, and have to be taught the operations of voluntary remembering, say in a game situation, before they can use them in the laboratory (Istomina, 1948). Thus, free recall seems to be a poor index of comprehension, particularly with younger children. Probe questions, recognition, and reconstruction tests have their own problems when they are used to assess story comprehension.

We decided that if we told our stories by means of pictures rather than verbally, we could use a task simple enough that even our youngest subjects (4-year-olds) would give us a data set rich enough to be analyzed for various signs of story comprehension or incomprehension. We asked our subjects to describe each picture to the experimenter. In the normal condition, a series of 15 to 18 pictures telling a well-structured story of at least two episodes was shown in normal order to each child. So that the child could form an impression of the story the child previewed the series of pictures. Then, each picture was shown again and the child was asked to tell the experimenter about them. In a scrambled condition, the same procedure was followed, except that the pictures were shown out of order. By comparing the childrens' responses to the pictures in normal and scrambled order the role of the story schema could be evaluated. Consistent differences in the way the children describe the pictures when

they tell a story or when they do not tell a story would indicate the extent to which the children understood the story.

After describing each picture set, the children were asked to recall the story without the pictures. In this way free-recall data were also obtained and could be compared with the picture descriptions.

## METHOD

### Subjects

Informal observations led us to believe that while 6-year-olds are already quite familiar with stories, 4-year-olds seem to be still in the process of acquiring a story schema; hence our decision to use these two age groups as subjects for this study.

Sixteen 4-year-olds (mean age, 4 years 6 months) and sixteen 6-year-olds (mean age, 6 years 4 months) participated in the study. The children attended a private school which provided preschool and daycare services. Boys and girls were about equally represented, and the children generally came from middle-class backgrounds. The school they were attending emphasized the free development of cognitive and social skills in an informal atmosphere.

All children were individually tested in two sessions in a room provided by the school in which other activities also took place. Thus, the creation of a testing atmosphere was avoided, since the experimental session was not very different from many other routine interactions between teachers and helpers and individual children. Reading and telling stories, or looking at pictures, were not unusual activities at the school, and the two female experimenters were known to the children.

### Stimulus Materials

Four picture sequences were used to elicit stories from the children. The sequences were constructed from commercial children's stories selected because of their tightly structured order. Two of the sequences, *What Whiskers Did* (Carroll, 1972) and *A Boy, A Dog, and A Frog* (Mayer, 1967), were picture stories without narrative. These stories were shortened by deleting redundant or irrelevant pictures. The final sequences each contained 16 pictures. The second two sequences, *Corduroy* (Freeman, 1968) and *A Doll for Marie* (Fatio, 1957), were originally narratives with pictures. For the purposes of this study, the narratives were excluded. The final picture sequences contained 18 pictures for *Corduroy* and 15 for *A Doll for Marie*. Each picture was placed in an individual plastic document protector and bound in a loose leaf binder. A scrambled order was constructed for each story such that no two pictures were in their correct successive order. Figure 1 shows the 16 pictures of *Whiskers* in their correct order. In the scrambled order the sequence of

pictures was 12, 2, 5, 14, 16, 8, 3, 11, 13, 7, 1, 15, 9, 6, 4, 10. In the actual experiment the pictures were, of course, all presented one by one and in their original size (about 25 × 35 cm).

*Procedure*

The children were presented two picture sequences in each of two sessions. The sessions were on successive days. In Session I all children were shown a story sequence in normal order first. Instructions were always as follows:

> "I'm going to show you some pictures which tell a story. I want
> you to look at them the first time through and then I will ask you
> to tell me about them the second time through. (Leaf through
> pictures.) Now will you tell me about them?"

The child's statements about each picture were tape-recorded. The picture book was closed and the child was asked for recall. The instructions for recall stressed the idea of telling a story:

> "Now can you tell me the story without looking at the pictures?
> What happened in the story? What can you remember?"

The child was then presented a story in scrambled order and the procedure was repeated.

In Session II the children again received a normal story first. Then they were presented a scrambled story and were informed that the story was "all mixed up." They were asked to put the pictures in the right order before telling the experimenter about them. The child placed the pictures on the floor and was allowed to change the order until he was satisfied. The experimenter then picked up the pictures in the child's order and the procedure continued as before. The four picture sequences were counterbalanced across the four tasks, so that over the 16 subjects each story was used equally often in each condition. Two experimenters each tested half of each age group.

*Adult sample.* In order to establish an adult standard for the picture descriptions, 24 college students were paid to write one or two brief statements for each picture.[2] Their instructions were to tell how each picture related to the story, and they saw all stories in their normal orders. The four stories were presented in counterbalanced order, except that because of an experimental error two pictures of *Corduroy* were shown out of order, and a new sample of 24 students had to be asked to write brief responses for this one story. No adult sample was obtained for the scrambled order, because informal observation showed that adults can reconstruct the original story perfectly, or nearly so, from the scrambled

---

[2] We found that without this restriction, adult descriptions tended to be extremely long and wordy, and hence, useless for our purposes.

input, and hence, one would not expect systematic differences in their descriptions.

*Scoring.* The tape-recorded protocols were transcribed, and the transcripts were analyzed propositionally according to the procedures developed in Kintsch (1974), except that modifiers and negations were not counted as separate propositions. All scoring was done independently by the first two authors. A reliability of .94 was obtained. Discrepancies between the scores, most of which were simple oversights, were resolved in conference. Each proposition was classified in five ways.

(1) Each proposition was assigned to one of three response classes: *core propositions, extra propositions,* and *spurious propositions.* The core propositions were determined from the results of the adult sample: each proposition given in response to a picture by more than 50% of the adult subjects was designated a core proposition. The average number of core propositions per picture for the four stories was 1.6, 1.8, 2.1, and 1.5 for *Whiskers, Corduroy, Frog,* and *Doll,* respectively. Statements that were correct descriptions of a picture but that were not generated by more than half of the adult subjects were classified as extra propositions. While core propositions can be regarded as the essence of the story, as determined by adult consensus, extra propositions constitute the inessential detail. Wrong responses, on the other hand, were designated as spurious. Spurious statements did not describe the content of a picture correctly in the judgment of the experimenters. They were labeled spurious rather than false because in many cases they were possible, though far-fetched, interpretations of a picture from the standpoint of the adult, as when a 4-year-old said in response to picture 15 of *Whiskers* (Fig. 1) "His knees are broken"—they certainly are not, but they are drawn strangely, and the child was clearly interpreting the picture, not fabricating a response.

(2) The second classification distinguished between *picture propositions, story propositions,* and *narrative conventions.* The large majority of responses were picture propositions, that is, they were responses to some feature of the stimulus picture. Story propositions, on the other hand, cannot be derived from any one picture, but depend upon an understanding of the story. For instance, Picture 15 of *Whiskers* could be described in many ways ("The boy is sad," "the dog is behind the boy," etc.), but the response "The dog comes back to the boy" is possible only from an understanding of the whole story. Only a few pictures in each story were likely to elicit story propositions as responses. Narrative conventions similarly are not responses to the picture itself, but unlike story propositions they need not indicate an understanding of the story, merely a familiarity with the conventions that are observed in story telling, e.g., concerning settings ("once there was," "one day"), resolutions ("and he lived a long time there"), and mostly temporal connectives ("and when,"

"again," "finally"). The temporal connective "then" was not scored in this or any other category because it frequently occurred in the protocols of some subjects in a ritualistic way, introducing almost every statement they made.

(3) The third classification distinguished between *descriptive propositions* (the vast majority of the responses given) and *cognitive and affective statements*. Any statement attributing "saying," "thinking," "planning," "wanting," or an emotion to a story character was included in the cognitive/affective category.

(4) Each proposition was assigned to the picture which had elicited it. Each picture was then assigned to one of the narrative categories— *exposition, complication,* or *resolution*—according to the principles referred to above. Thus, for the recall responses a fourth classification for each proposition was obtained in terms of its role in the macrostructure of the story. Deleting a few ambiguous responses (less than 2%), this could be done reliably because many of the statements made in recall are repetitions of statements made in describing the pictures.

The assignment of pictures to narrative categories is illustrated in Table 1 for the *Whiskers* story which had the most complex macrostructure of the stories used here. *Whiskers* consists of three episodes, distinguished by a change in the actors: the actors of the first episode are a boy and his dog; the dog, a rabbit, and a wolf appear in the second episode, and the dog and the rabbit family in the third. More specifically, Picture 1 (see Fig. 1) functions as the exposition of Episode 1, introducing a boy leading his dog on a leash. The complication of that episode starts with the next picture: the leash breaks and the dog runs away. Episodes 2 and 3 are the adventures of the runaway dog, and thus, are embedded into the first episode as part of its complication. Picture 3 opens the second episode with a new character, the rabbit. The scene changes again with Picture 4, when a wolf appears who chases the dog and the rabbit (Pictures 5 and 6).

TABLE 1
THE MACROSTRUCTURE OF THE *Whiskers* STORY

*Episode* 1:
  *Exposition* 1: Boy with dog on leash (Picture 1)
  *Complication* 1: Dog runs away (Picture 2)
    *Episode* 2:
      *Exposition* 2: Rabbit sees dog (Picture 3)
      *Complication* 2: Fox chases dog and rabbit (Pictures 4–7)
      *Resolution* 2: Dog and rabbit escape (Pictures 8 and 9)
    *Episode* 3:
      *Exposition* 3: Dog meets rabbit family (Pictures 10 and 11)
      *Complication* 3: Events in rabbit's home (Pictures 12 and 13)
      *Resolution* 3: Dog leaves (Picture 14)
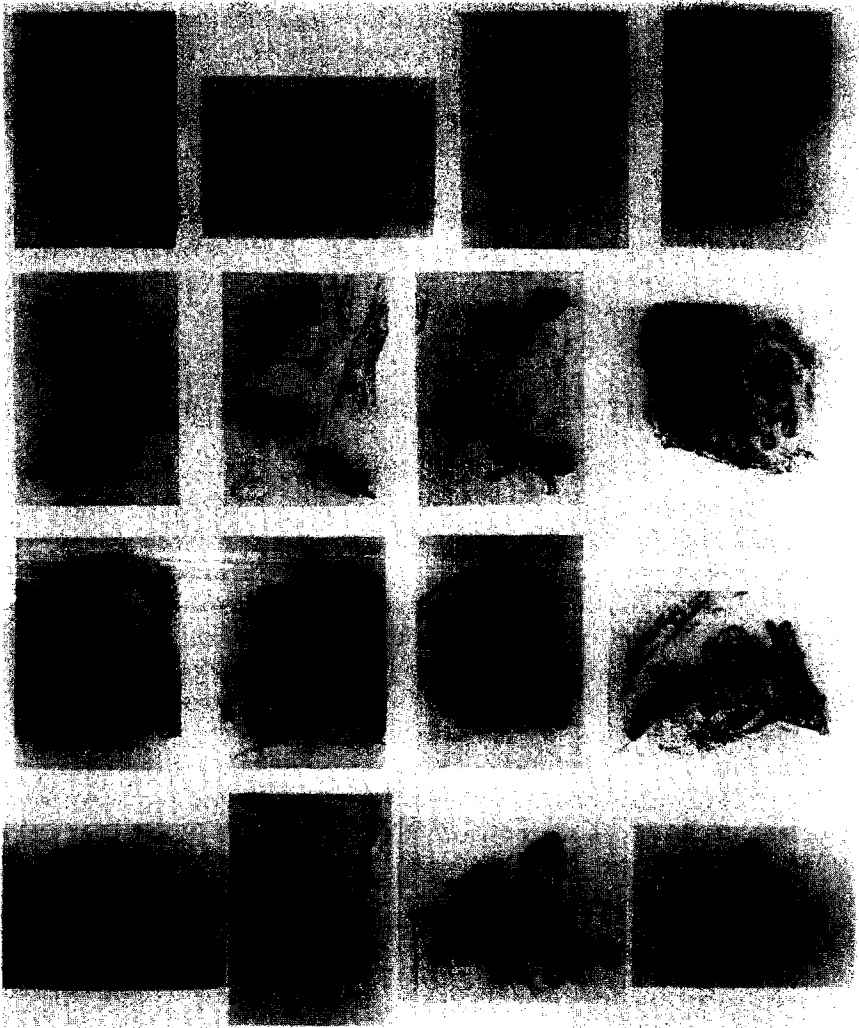  *Resolution* 1: Boy and dog together (Pictures 15 and 16)

FIG. 1. The *Whiskers* story.

These three pictures constitute the complication of Episode 2. The chase is resolved in Pictures 7 and 8 where the dog and rabbit escape into a hole (Resolution 2). A new episode is introduced in Picture 9—the dog is alone in the cave. The episode comes to its climax in Pictures 10 through 13, which tell of the adventures of the dog with the rabbits. A leave-taking scene, Picture 14, is the resolution of Episode 3, and the story returns to the boy: dog and boy are re-united in Picture 15 and 16, which form the resolution of the first, embedding episode.

The other three stories consisted of only two episodes each. For both

*Corduroy* and *Doll* the second episode forms the complication of the first episode. *Frog* simply has two coordinated episodes.

(5) The final classification depended upon the form of the subject's response. One-word responses, that is, labels, were distinguished from two- and three-or-more-word phrases. Only content words were considered in this analysis.

Thus, each propositional statement was classified in five ways. For instance, the response "The boy is sad," which occurred frequently to Picture 15 of *Whiskers* (Fig. 1), would be scored as (1) core proposition (because more than 50% of the adults responded in the same way); (2) picture proposition (since it depends upon the facial expression and posture of the boy in the picture); (3) affective proposition; (4) resolution category (because Picture 15 is part of the resolution of the story); (5) two-concept phrase (relating "boy" and "sad").

In some cases the classifications reported here were quite objective (e.g., the form classes, the determination of core propositions on the basis of an adult sample, the list-based narrative conventions, the cognitive/affective classification on the basis of key words). But even in those cases where subjective judgment was called for, this judgment was fairly reliable: the agreement between the two raters, as determined from a random sample of about one-fourth of the data from the 6-year-olds, was .89 for the core/extra/spurious classification; for the picture versus story proposition classification it was .97; for the descriptive versus cognitive/affective classification it was .99; and for the formal classification the reliability reached .996. Disagreements, except for obvious scoring errors, were resolved in favor of the scorer who had worked with that particular child during the experimental session, who presumably had a better idea of what the child was referring to than the scorer who merely read the protocol.

## RESULTS

The results for the five ways of scoring the data will be presented separately. In each case we shall first discuss the descriptions and then the recall protocols.

### Total Number of Responses

The total number of propositions in the description and recall protocols for normal and scrambled presentation order for both age groups is shown as a function of response class in Table 2. The standard errors shown in Table 2 were determined from the results of an ANOVA which included Age as a between-subjects variable and Experimental Condition (Normal vs Scrambled) and Response Class (Core, Extra, Spurious) as within-subjects variables. A .01 significance level is used throughout, unless noted otherwise. In the analysis of the responses given as descriptions of

TABLE 2
AVERAGE NUMBER OF PROPOSITIONS PER STORY FOR DESCRIPTIONS AND RECALL (IN PARENTHESES)

|  | 4 Years | 6 Years |
|---|---|---|
| Normal |  |  |
| Core | 11.94 | 15.09 |
|  | (4.44) | (8.16) |
| Extra | 8.97 | 11.59 |
|  | (2.38) | (5.28) |
| Spurious | 3.25 | 1.97 |
|  | (0.69) | (0.72) |
| Scrambled |  |  |
| Core | 8.88 | 12.38 |
|  | (2.06) | (4.50) |
| Extra | 9.00 | 12.38 |
|  | (1.19) | (4.44) |
| Spurious | 6.50 | 5.72 |
|  | (1.06) | (2.38) |

Note. The standard error of the mean is .95 for descriptions and .53 for recall.

the pictures two interactions were significant. Figure 2 shows the Response Class × Condition interaction, $F(2, 150) = 11.17$. Although subjects gave as many responses in the scrambled condition as in the normal condition ($F < 1$), the nature of their responses was different: in the normal condition they produced more core propositions and very few spurious propositions, but when the pictures were scrambled the children
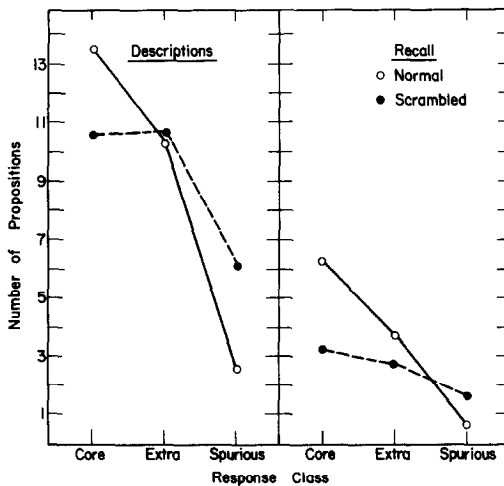


FIG. 2. Total number of propositions in descriptions and recall as a function of experimental condition.
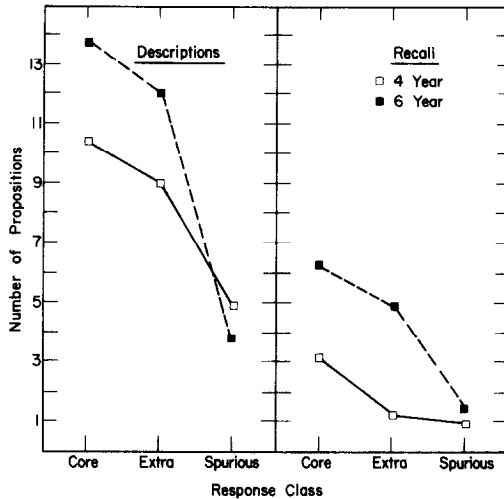
FIG. 3.  Total number of propositions in descriptions and recall as a function of age.

agreed less well with the adults and produced many spurious proposi-
tions.

Age similarly interacted with response class (Fig. 3). Although 4-year-
olds produced almost as many descriptions of the pictures as 6-year-olds,
$F(1,30) = 2.68, p = .11$, fewer of their statements were correct (core +
extra) and more of them were spurious, $F(2, 150) = 6.45$.

In recall, these effects are accentuated: recall was significantly better in
the normal condition, $F(1,150) = 10.77$, but again more core propositions
and fewer spurious propositions were observed in the normal condition,
$F(2,150) = 14.44$. Thus, the normal order apparently helped the subjects
to understand the story, so that they produced better, more adult-like
responses. The main effect of age was also significant, $F(1,30) = 7.10$, but
so was the Response Class × Age interaction with $F(2,150) = 6.86$.

*Story Propositions*

The frequency of story propositions in the various experimental condi-
tions is shown in Table 3. Story propositions were analyzed in the same
way as the total number of propositions above, that is, by means of an
analysis of variance with the factors Age, Response Class, and Condition.
This analysis yielded two significant interaction effects. The Response
Class × Condition interaction was significant ($F(2,150) = 4.70$). While the
story propositions in the scrambled condition were about equally likely to
be core, extra, or spurious, almost all of the responses in the normal
condition were correct inferences, and most of them were in fact core
propositions also given by the adult respondents. Note that equally many
story propositions were given in the two experimental conditions ($F < 1$).

TABLE 3
AVERAGE NUMBER OF STORY PROPOSITIONS PER STORY FOR DESCRIPTIONS AND RECALL (IN PARENTHESES)

|  | 4 Years | 6 Years |
|---|---|---|
| Normal | | |
| Core | 0.63 | 1.06 |
| | (0.37) | (0.78) |
| Extra | 0.35 | 0.81 |
| | (0.41) | (0.47) |
| Spurious | 0.93 | 0.28 |
| | (0.00) | (0.16) |
| Scrambled | | |
| Core | 0.38 | 0.81 |
| | (0.00) | (0.31) |
| Extra | 0.13 | 1.22 |
| | (0.12) | (0.37) |
| Spurious | 0.13 | 1.00 |
| | (0.00) | (0.47) |

Note. The standard error for the description means is .29 and for recall .18.

However, when the children described a picture sequence that told a story, they made correct inferences about it, whereas with the scrambled pictures, many of their inferences were necessarily spurious.

The 4-year-olds tended not to say anything when they had nothing to say. The 6-year-olds, on the other hand, made all sorts of inferences in the scrambled condition (often wrong ones, as we have just seen). This age change is reflected in the significant interaction between Condition and Age, $F(1,150) = 5.67$.

In recall, we observe a condition effect, $F(1,150) = 7.55$, modified by an interaction with Response Class, $F(2,150) = 9.34$: correct story propositions were produced when the pictures were shown in sequence, but fewer and more incorrect ones occurred when their normal order was disturbed. Although the Age × Conditions interaction was not significant in the recall data, $F(1,150) = 1.52$, the pattern of results was once more similar to that obtained with the descriptions.

The analysis of story propositions indicates that subjects, especially the 6-year-olds, were making an effort to interpret the pictures as a coherent story: when the pictures were in order, they were quite successful; when the pictures were out of order their attempts to tell a story frequently failed, yet they kept trying, so that the 6-year-olds produced in fact more story propositions in the scrambled condition than in the normal condition.

No separate analysis of picture propositions is presented here because that analysis duplicates in all respects the analysis for the total number of

propositions: by far the largest portion of all responses were, in fact, picture propositions.

*Narrative Conventions*

The same desire to make some sense out of the scrambled pictures also led the subjects to use considerably more narrative conventions in the scrambled than in the normal condition, .67 vs .23 per story. This difference was significant, $F(1,150) = 28.51$, in an analysis of the frequencies of narrative conventions employing the same factors as in the previous analyses. The only other significant effect in this analysis was age, with the older children employing narrative conventions much more frequently than the younger ones (.65 vs .25 per story, $F(1,30) = 7.65$). The analysis of the recall data yielded corresponding results, except that the overall incidence of narrative conventions was reduced to .20 per story.

*Cognitive and Affective Responses*

Table 4 shows the average frequencies of cognitive and affective responses that were produced when the children were describing the pictures and when they recalled them. The same analysis of variance design as above was used to analyze these data. Two interaction effects were found to be significant. The first is the Response Class × Condition interaction. While about the same number of core and extra responses were observed in the normal and in the scrambled conditions, there were four times more spurious cognitive/affective responses in the scrambled condition, $F(2,150) = 3.58$, $p = .030$. When the pictures did not make

TABLE 4
AVERAGE NUMBER OF COGNITIVE OR AFFECTIVE PROPOSITIONS PER STORY FOR DESCRIP-
TIONS AND RECALL (IN PARENTHESES)

|  | 4 Years | 6 Years |
|---|---|---|
| Normal |  |  |
| Core | 0.34 | 0.84 |
|  | (0.13) | (0.53) |
| Extra | 1.31 | 2.59 |
|  | (0.44) | (1.50) |
| Spurious | 0.38 | 0.31 |
|  | (0.06) | (0.19) |
| Scrambled |  |  |
| Core | 0.28 | 0.50 |
|  | (0.03) | (0.25) |
| Extra | 1.53 | 2.75 |
|  | (0.25) | (1.06) |
| Spurious | 1.47 | 1.31 |
|  | (0.13) | (0.72) |

*Note.* The standard error of the means is .68 for descriptions and .33 for recall.

sense, the children nevertheless tried to infer plans and motives from them, resulting in many spurious cognitive/affective responses.

Six-year-olds tended to give more cognitive and affective descriptions than 4-year-olds, though this effect was not statistically significant. However, a significant Response Class × Age interaction shows that the number of correct responses almost doubled with age, especially the extra propositions, while spurious cognitive/affective descriptions actually declined, $F(2,150) = 4.17, p = .017$.

In recall, the same pattern of results was obtained. Cognitive and affective responses that could be classified as core or extra propositions were more frequent in the normal condition than in the scrambled condition, but the reverse was true for spurious responses, $F(1,150) = 3.84, p = .024$. The overall frequency of cognitive and affective responses increased with age, $F(1,30) = 5.72, p = .023$, but the locus of this increase was primarily in the "extra" category, as shown by a significant Response Class × Age interaction, $F(2,150) = 4.49$.

Seventy-one percent of the responses shown in Table 4 were cognitive responses and the rest affective responses. These proportions were about the same for both age groups.

## Retention

If one studies the correspondence between descriptive responses and recall in the preceding analysis, it is clear that the two are highly correlated. Nevertheless, the ratio between the number of responses in the descriptions and in recall was by no means constant: it was highest for core propositions, especially in the normal condition. This suggests a sharpening effect in recall: correct responses that are well integrated in a story tend to be reproduced in recall, while either spurious or extra responses in the scrambled condition (both of which are not as well integrated into a story) tend to be forgotten. The clearest statistical justification for this claim is obtained by an analysis of those responses that were produced both as a picture description and in recall. Figure 4 shows the percentage of descriptions that were repeated in recall as a function of the experimental condition. An ANOVA of these data (after an arcsin transformation) showed that all three main effects (Age, Response Class, and Condition), were statistically significant, but none of the interactions were. Core propositions were more likely to be reproduced than either extra or spurious propositions, $F(2,150) = 31.00$. Indeed, in the normal condition over half of the core propositions generated in response to the pictures were retained in recall! Retention was generally higher in the normal than in the scrambled condition, $F(1,150) = 11.86$, and retention was higher for the 6-year-olds than for the 4-year-olds, $F(1,30) = 8.62$. All three of these observations are in agreement with the hypothesis that retention depends primarily upon the degree of integration
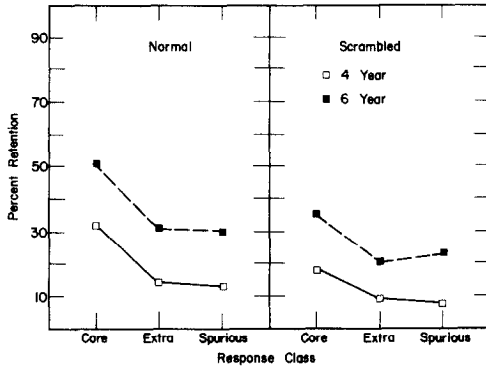
F

IG

. 4.   Percentage retention as a function of age and experimental condition.

achieved. Very few core propositions occurred in recall that were not given before in the descriptions: only 5% for the 6-year-olds and 17% for the 4-year-olds. This contrasts with 42% (44% for the 4-year-olds) new extra and spurious responses in recall.

The retention of story propositions shows the sharpening effect in recall even more clearly. Correct story propositions in the normal condition were retained very well: 57% for the 6-year-olds, 48% for the 4-year-olds. In contrast, story propositions in the scrambled condition were reproduced in recall only 25% of the time (10% for the 4-year-olds). However, new story propositions were quite frequently introduced in recall by both age groups and in all experimental conditions, attesting to the children's attempts to shape their recalls into coherent stories. For 6-year-olds, 26% of all story propositions in recall were not given as descriptions, and 41% of the 4-year-olds' story propositions were new.

*Macrostructure Analysis*

As we have shown above, each picture functions in the story either as part of its introduction, complication, or resolution. If one tabulates the total number of recall responses to the pictures in each story category, a very clear trend appears in the data: for all episodes in every story, by far the most responses occurred in the complication category, many fewer in the exposition category, and still fewer in the resolution category. Furthermore, the first exposition in each story as well as the final resolution are emphasized more heavily than the expositions or resolutions of other episodes. Thus, not only the structure of each episode, but that of the story as a whole is reflected in the recall protocols.

Unfortunately, these rather striking results are difficult to interpret. For the most part, they merely reflect the way the stories were constructed: most of the pictures (and hence, most of the responses) dealt with the interesting parts of the story i.e., the complication. Furthermore, the

complication pictures tended to be more salient than the more sedate pictures of the exposition or resolution; hence, again, they were favored in recall. That rather pronounced and stable differences existed between pictures in recall is shown by the fact that the number of responses per picture correlated $r = .69$ between the 4- and 6-year-olds.

The function of a picture in the story did, however, demonstrably influence recall over and above mere saliency differences. This is revealed by a comparison between the recall of each picture when the children were telling a story (the normal condition) and when they were not, or at least not as successfully (in the scrambled condition). The pattern of recall is quite different in the normal and scrambled conditions: the amount recalled per picture in the two conditions no longer correlates significantly, $r = .23$. What is a salient picture in a story is not necessarily salient out of the story context.

If one calculates, instead of the total number of responses in each narrative category, the average number of responses per picture in each category, an index of the importance of each category is obtained unconfounded by differences in the number of pictures per category. If the narrative structure does not influence how much a child says when describing a picture, then these average values should be approximately the same in the normal and scrambled condition. On the other hand, systematic differences between conditions imply that the mere fact that a picture belongs to a certain narrative category determines how the children describe it.

Figure 5 shows the difference in the number of responses per picture between the normal and the scrambled condition. The pictures that form the expositions of the episodes in our stories elicit about the same number of recall responses, whether they are interpreted as part of a story or not. Pictures that belong to the complication category show a small superiority in the normal condition. The main difference occurs, however, in the resolution category: the same pictures elicit substantially more responses when they form the resolution of an episode than when they are recalled out of the story context.

## Differences between Stories

The four stories used in this study were selected because they were well structured and not dependent upon a verbal text. There was no reason to suppose, however, that they would be equally difficult. Indeed, large differences between stories were observed. *Doll* and to some extent *Corduroy* were harder than the other two stories, at least for the 4-year-olds. In fact, it appears that *Doll* was too hard and that it was not understood well at all, even when presented in normal order: the finding of much better performance in the normal than in the scrambled condition, which is so clear in the data for the other three stories (as well as in
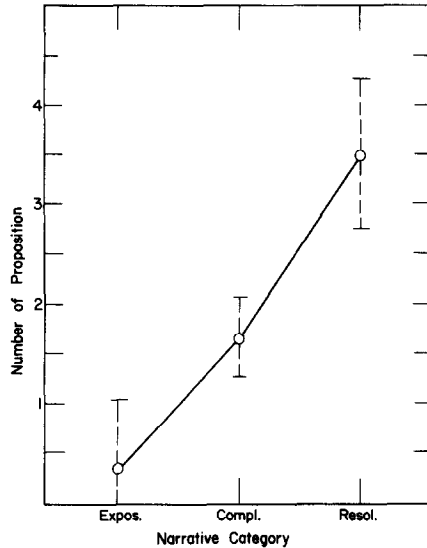
Fig. 5. Number of propositions recalled in the normal condition minus number of propositions recalled in the scrambled conditions for different narrative categories. The data are averaged over all episodes from all four stories and both age groups. The standard error of each mean is indicated.

the average data reported above) does not always hold for *Doll*. The children's reactions to the *Doll* story suggest that it was difficult because it contained a lot of pictorial detail which confused them. Furthermore, there were some gross changes in perspective: in some scenes the doll appeared life-size, and the younger children had difficulty identifying it with the little thing in the store or in the girl's hands.

Inter-story differences, as well as inter-subject differences which were equally prominent, cannot be investigated further within the design of the present experiment. But it should be noted that, because of the counter-balancing employed, they did not bias the analyses reported above, though both factors undoubtedly inflated the error variance.

*Formal Analysis*

Each statement was classified as a two-word phrase, a three-or-more-word phrase, or a one-word phrase, or label. Describing a picture simply by labeling some of the objects in it is a very primitive response form. As Table 5 shows, it occurred relatively frequently in the protocols of the 4-year-olds (15%), but only rarely in the 6-year-olds (2% of all responses). Two-word phrases were about equally frequent in 4- and 6-year-olds (26 and 20%, respectively), while longer constructions were observed more frequently in the older children (59 vs 78%). An analysis of variance of the labeling responses with Age as a between- and Conditions as a

TABLE 5
AVERAGE NUMBER OF LABELING RESPONSES PER STORY IN THE PICTURE DESCRIPTIONS

|  | Normal | Scrambled |
|---|---|---|
| 4 years | 3.22 | 5.00 |
| 6 years | 0.56 | 0.34 |

*Note.* The standard error of the means is .95.

within-subjects factor yielded a marginally significant main effect of age, $F(1,30) = 4.50, p = .042$, and an Age × Conditions interaction, $F(1,30) = 4.41, p = .044$. The latter reflects a floor effect: for the 6-year-olds labeling responses are very rare in either the normal or scrambled condition, while for the 4-year-olds labeling is more frequent in the scrambled condition than in the normal condition. Apparently, when the 4-year-olds cannot make sense of the pictures, they tend to regress to an earlier, more primitive response mode. Many of the labeling responses in the scrambled condition referred to details of the pictures that were entirely unrelated to the story, e.g., when a subject's only comment to Picture 2 of *Whiskers* was "That's a branch."

### Reconstruction

In Session 2 subjects were asked to put the scrambled pictures into their right order before describing them. Note that this is not a conventional reconstruction task, since the subjects had never seen the true order before. To order 15 to 18 pictures is a difficult task, and was quite beyond the capabilities of the 4-year-olds. In the sortings they produced, only 3% of adjacent pairs were in their correct order. The 6-year-olds did somewhat better: 27% of adjacent pairs were correct.

### Recall Order

In the normal condition, the order of recall corresponded perfectly to the order of events in the stories in 62% of the recall protocols. In 25% the order was reversed, the children having begun their recall with the end sequence of the stories. For the remaining 13% of the protocols the order was mixed. The results for the scrambled condition are entirely different; only 15% of the protocols duplicated the input order, while 35% started with the last-presented pictures, and half of the protocols were mixed up. There were no major age differences in recall order.

## DISCUSSION

When children describe a sequence of pictures it makes a profound difference whether or not the pictures tell a story. If they do, the children's descriptions tend to emphasize those elements of the pictures that

are important for the story. Their general knowledge about the nature of stories—the story schema—is used to organize their descriptions so that they come close to telling a story. The children not only describe what they see in the pictures, but tell what it means in terms of the story as well, thus providing clear evidence of their ability to comprehend stories, and indirectly of having acquired an adult-like story schema upon which this comprehension is based. Perhaps one could arrive at this conclusion simply by reading through the protocols we have collected. However, we have avoided relying upon such anecdotal evidence and have presented instead a number of different data analyses designed to capture the strong impression one obtains from the protocols in a more objective and hence more reliable way. Several different analyses were necessary in order to do justice to the richness of the protocols. These analyses all converge on the same conclusions.

The data indicate the guiding role of the schema in various ways. With the 4-year-olds, even the form of the response was affected: primitive labeling responses were much more frequent when the children responded to incoherent pictures than when they could tell a story. In general, it was not the total number of descriptive statements that was affected by the presentation order, but their nature: when the order was normal, the descriptions were better, more adult-like, more true to the pictures. When the pictures were out of order, more spurious responses were given. The same was true for the comparison between the 4- and 6-year-olds: the older children possessed a better schema and therefore were better able to treat the pictures as a story.

While most of the responses to the pictures were descriptive, there were certain types of responses that connected different pictures, our "story propositions." The fact that the 6-year-olds gave substantially more such responses than the younger children confirms our conclusion about the general importance of the story schema, but the observation that the number of story propositions was about the same when the pictures were presented in or out of order may seem at first like a contradiction. Linking pictures by inferences would be expected to occur when pictures are well organized on the basis of some schema (as in Brown, 1975), but not in the absence of it (as in the study of Piaget, 1968). Closer inspection of our results reveals the reason for this apparent contradiction: there is a second factor operating here which Bartlett had termed the effort after meaning. When the pictures are out of order and do not make sense, the children do not simply give up, but try their best to make some sense anyway. They keep producing story propositions linking the pictures, except that now many of these are necessarily spurious. The same interpretation is suggested by the high incidence of spurious cognitive and affective responses as well as the greater use of narrative conventions in the scrambled condition. When there was no story to be

told the children used the now empty trappings of a story with special frequency. Age differences were interesting in this respect. What we have just described characterizes primarily the behavior of the older children. It is the 6-year-olds who try to make a logical sequence out of the scrambled pictures, whereas the 4-year-olds are just beginning to show these response tendencies (e.g., the increased frequency of cognitive and affective responses in the scrambled condition). However, for the most part they react in a more primitive manner: they label, or, in recall, produce only very few responses.

The recall data add strong support to both of these conclusions. The pattern of recall was much like that of the descriptions, except for a pronounced sharpening effect. What was recalled best was the material integrated into a story. Hence core propositions, and especially core story propositions, were reproduced very well in recall in the normal condition. Since the core propositions essentially correspond to the gist or macro-propositions of the story, they were the best integrated material and most likely to be recalled. Nonintegrated material, especially spurious proposi-tions, but also extra propositions which correspond more to the details of the story rather than its gist, were much less likely to be produced in recall in both conditions. In the scrambled condition where story integration was not achieved recall was generally much lower. However, the some-what elevated retention probability of core propositions even in the scrambled condition may be taken as evidence that some progress toward story integration was made at least by some subjects.

Overall recall was quite good in our experiment. For instance, only one child, a 4-year-old, refused to recall anything (on two of the four stories). This contrasts sharply with the results of Hall et al. (1977) where spon-taneous recall was obtained from 4½-year-olds on only 52% of their stories, or Mandler and Johnson (1977) who replaced 38% of their first-grade subjects and even 19% of the fourth-grade pupils for failure to recall! Obviously, the relatively high level of recall in the present experi-ment depends upon the fact that the children were recalling not only the pictures themselves but also their own previous descriptions of these pictures.

The question of which parts of a story are recalled best is a complicated one. In the present experiment most of the recall concerned the complica-tions of the episodes. Stein and Glenn (1979) and Mandler and Johnson (1977) reported the best recall for major settings and beginnings (which are here subsumed under exposition) and outcomes–consequences (which here are part of the complication). However, such data are difficult to interpret because recall obviously depends also upon the input: in the present experiment most of the pictures are in fact about the complica-tions of the various episodes, that is, about the interesting, surprising events of the story, while expositions and resolutions are treated quite

briefly. Thus the higher recall in this category, in part, simply reflects the greater quantity of the input. In addition, not only the quantity of the input in the various story categories matters: complications are by their very nature exciting. Children will recall a dramatic picture of two dogs tearing up a doll not just because it constitutes the complication of an eposide, but simply because it is salient. The high correlations between the amount of recall for particular pictures across age groups shows that these saliency differences are both important and stable (for similar observations see Stein and Glenn, 1979). One way to control for the quantity and quality of input is by comparing recall for pictures that form a story and for the same pictures when they do not. If there are stable differences between the two as a function of the assignment of the pictures to narrative categories, these may be attributed to effects of the story schema. As we have seen in Fig. 5, the only consistent differences here were that pictures that form the resolutions of episodes were recalled better than the same pictures out of the story context. Such pictures are often not very salient by themselves (a girl putting a doll to bed, a dog standing behind a boy) and hence are poorly recalled; within the story, however, they play an important role, and are correspondingly recalled better.

This observation is important because it permits us to distinguish between the two aspects of the schema referred to in the introduction of this paper: those that relate specifically to the structure of narratives, and the general human knowledge that is needed to interpret the actions of the narrative. The latter have nothing to do with the stories per se, but belong to a general theory of knowledge. Since stories are mostly about human actions, the most relevant knowledge is that concerned with the principles and rules of human action—the theory of action upon which much of human behavior in and out of stories is based. For instance, the children in the present experiment were able to understand the causal relationships in the chase scene in Fig. 1 not because of any knowledge specific to stories, but because they could organize it on the basis of a "chase script" that is part of their general knowledge about actions. On the other hand, Fig. 5 suggests that superimposed on these effects is a differential response to the pictures, depending upon the role they play in the story. All pictures, whatever their category membership, are interpreted in terms of the appropriate action scripts, but those belonging to the resolution category are additionally emphasized because of their function in the narrative schema. Thus, story comprehension involves not only knowledge about the causal and temporal regularities of action sequences, but also a truly story-specific component.

As Table 4 shows, the overall incidence of cognitive and affective responses was rather low. This was also the case in the data of Mandler and Johnson (1977) and Stein and Glenn (1979). We know that this is not due to an inability of the children to infer thoughts and feelings of

other persons. Under simple task conditions children are able to describe correctly another person's reactions regarding familiar events (e.g., King, 1971; Marvin, Greenberg, & Mossler, 1976). The fact that they do relatively little of it in story recall might mean that knowledge about the actors' goals and motivations is not necessarily made explicit in story comprehension.

As far as the order of recall is concerned, it was good when the input was well structured and poor (with a heavy recency bias) when it was random. Since this result was achieved with much more complicated materials than used in other studies, it provides a strong confirmation of similar results discussed in the introduction.

Note that our use of the reconstruction task probably served to diminish the differences that were obtained between the normal and scrambled presentation orders in the present experiment. On half of the stories that were presented in the scrambled order the children were allowed to reorder them as well as they could and were then shown the pictures in the order they had chosen. However, since the 4-year-olds could not reorder the pictures at all, and even the orderings generated by the 6-year-olds were only a little better than random, this feature of our experimental procedure had relatively little effect. In any case, whatever effects there were, they worked against the experimental hypothesis tested here: without the reconstruction task the differences between presentation orders could only have been more pronounced.

The main conclusion of this study is that the availability of a story schema affects the way children describe and recall picture stories. The effect is not revealed in the quantity of responses, but in the quality. The schema guides the subjects' responses. In the absence of such guidance responses tend to be less appropriate and frequently spurious. In addition we have observed a decided effort after meaning in the behavior of our older subjects: they tried to make sense of an input, even when the game was rigged and no sense could be made. Presumably, this is simply another manifestation of certain general conventions that regulate human communication. Messages are supposed to make sense, and the stories that the child hears and the pictures that he looks at, at home or in school, are meaningful. Therefore they tried to treat our random picture sequences in the same way. The 4-year-olds, on the other hand, reacted to the lack of meaning rather differently: they labeled or failed to respond. Apparently, a very noticeable developmental change occurs in this respect between the ages of 4 and 6.

The developmental approach used here permitted us to arrive at these conclusions in two ways. Once by looking at the differences caused by presenting the pictures in normal or scrambled order, arguing that in the first case the schema influence would be apparent, while in the second it would be absent or at any rate greatly reduced. Then, using the same

argument, we claimed that 6-year-olds have acquired a better, more adult-like schema than 4-year-olds. The similarity in the effects of presentation order and age in our data indicates that this approach succeeded.

Given that the availability of a schema has such important consequences for comprehension as observed here, further study of the educational implications of this finding seem indicated. To what extent are comprehension problems caused by the lack of an appropriate schema? How can schema use be trained? What is the effect of the many children's books, educational materials, and television programs composed of brief sequences not organized or organizable into a larger framework? What other schemata, in addition to the story schema, do we need to study? The methods exist now to approach such questions.

## REFERENCES

Bartlett, F. C. *Remembering*. Cambridge, England: Cambridge Univ. Press, 1932.

Brémond, C. *Logique du récit*. Paris: Seuil, 1973.

Brown, A. L. Recognition, reconstruction, and recall of narrative sequences by preoperational children. *Child Development*, 1975, **46**, 156–166.

Carroll, R. *What Whiskers did*. New York: Scholastic Book Service, 1972. (Originally published, 1932.)

van Dijk, T. A. *Some aspects of text grammars*. The Hague: Mouton, 1972.

Fatio, L. *A doll for Marie*. New York: McGraw–Hill, 1957.

Fraisse, P. *The psychology of time*. New York: Harper & Row, 1963.

Freeman, D. *Corduroy*. New York: Viking Seafarer, 1970. (Originally published, 1968.)

Grosse, E. U. French structuralist views on narrative grammar. In W. U. Dressler (Ed.), *Current trends in text linguistics*. Berlin: de Gruyter, 1977. Pp. 155–173.

Hall, W. S., Cole, M., Reder, S., & Dowley, G. Variations in young children's use of language: Some effects of setting and dialect. In R. O. Freedle (Ed.), *Discourse production and comprehension*. Hillsdale, NJ: Erlbaum, 1977. Pp. 161–174.

Istomina, Z. M. The development of voluntary memory in preschool age children. *Soviet Psychology*, 1975, **13**, 5–64. (Originally published, 1948.)

King, M. The development of some intention concepts in children. *Child development*, 1971, **42**, 1145–1152.

Kintsch, W. *The representation of meaning in memory*. Hillsdale, NJ: Erlbaum, 1974.

Kintsch, W., & van Dijk, T. A. Comment on se rapelle et on résume des histoires. *Langages*, 1975, **40**, 98–116.

Kintsch, W., & Greene, E. L. The role of culture-specific schemata in the comprehension and recall of stories. *Discourse Processes*, 1978, **1**, 1–13.

Kintsch, W., Mandel, T. S., & Kozminsky, E. Summarizing scrambled stories. *Memory and Cognition*, 1977, **5**, 547–552.

Lévi-Strauss, C. L'analyse morphologique des contes populaires russes. *International Journal of Slavic Linguistics and Poetics*, 1960, **3**, 122–249.

Mandler, J. M., & Johnson, N. S. Remembrance of things parsed: Story structure and recall. *Cognitive Psychology*, 1977, **9**, 111–151.

Marvin, R. S., Greenberg, M. T., & Mossler, D. G. The early development of perceptual perspective taking: Distinguishing among multiple perspectives. *Child Development*, 1976, **47**, 511–514.

Mayer, M. *A boy, a dog, a frog*. New York: Scholastic Book Services, 1968. (Originally published, 1967.)

Piaget, J. *The language and thought of the child*. New York: Meridian Books, 1955. (Originally published, 1928.)

Piaget, J. *Memory and intelligence*. New York: Basic Books, 1973. (Originally published, 1968.)

Propp, V. *Morphology of the folktale*. Austin, TX: Texas Univ. Press, 1968. (Originally published, 1928.)

Stein, N. L., & Glenn, C. G. An analysis of story comprehension in elementary school children. In R. Freedle (Ed.), *Multidisciplinary approaches to discourse processing*. Hillsdale, NJ: Ablex, 1979. Pp. 53–120.

Thorndyke, P. W. Cognitive structures in comprehension and memory of narrative discourse. *Cognitive Psychology*, 1977, **9**, 77–110.