Developmental Psychology
1994, Vol. 30, No. 2, 178-191

0012-1649/94/$3.00

Development of Gist Versus Verbatim Memory in Sentence Recognition: Effects of Lexical Familiarity, Semantic Content, Encoding Instructions, and Retention Interval

Valerie F. Reyna and Barbara Kiernan

Fuzzy-trace theory is used to explore children's memory and comprehension of sentences describing spatial or linear relationships. Recognition tests were given immediately and after a week's delay, and test sentences' truth, wording (original or novel), and premise–inference status were varied. When children were instructed to recognize only verbatim sentences (Experiment 1), premise recognition (memory) was independent of systematic misrecognition of true inferences (reasoning), and experimental manipulations (delay; spatial vs. linear stimuli) drove memory and reasoning in opposite directions. Therefore, verbatim memories were not semantically integrated with gist, such as inferences. When children were specifically instructed to process gist (Experiment 2), however, memory and reasoning were positively dependent. Results are discussed from the perspectives of constructivism, theories of suggestibility, and fuzzy-trace theory.

What is the developmental relationship between memory and reasoning? It has long been thought that the answer is obvious: The relationship must be one of mutual interdependence. After all, accurate reasoning demands accurate memory for problem information, and reasoning also shapes what is remembered. However, it has been found in recent studies that age improvements in reasoning are often independent of age improvements in memory for problem information, and sometimes the two are even antagonistic (for reviews, see Reyna, 1992; Reyna & Brainerd, 1991a). Fuzzy-trace theory was developed to account for such counterintuitive results.

In this article, we use fuzzy-trace theory to explore the relationship between children's memory for sentences and their sentential inferences. Previous research on this topic has been dominated by constructivism. According to constructivism (which we discuss later), children remember the gist of sentences, including inferences. Therefore, performance on memory tests and reasoning tests must be interdependent because both involve processing the same semantic gist. On the other hand, fuzzy-trace theory holds that children retain independent gist and verbatim representations of sentences. In the present experiments, we investigated alternative predictions that these two theories make about (a) whether children, once they are properly instructed, can distinguish gist and verbatim memories and (b) whether gist and verbatim memories are functionally interdependent.

Valerie F. Reyna, Department of Educational Psychology, College of Education, University of Arizona; Barbara Kiernan, Department of Speech and Hearing Sciences, University of Arizona.

The research reported in this article was partially supported by Grant HO29D90108 from the United States Department of Education to Barbara Kiernan. We gratefully acknowledge the assistance of the children, parents, teachers, staff, and principal, JoAnn Sayre, of St. Cyril's School, Tucson, Arizona.

Correspondence concerning this article should be addressed to Valerie F. Reyna, Department of Educational Psychology, College of Education, University of Arizona, Tucson, Arizona 85721.

Background: Previous Research

There are numerous demonstrations that subjects misreport gist as having been directly experienced, including the classic sentence memory experiments of Bransford and Franks (1971). The standard interpretation of such experiments (e.g., Bjorklund, 1989) is that subjects' memories have become modified so that the original verbatim memories are lost (e.g., Jarvella, 1971) or cannot be retrieved apart from a semantically integrated representation of presented information (e.g., Alba & Hasher, 1983). Memory, therefore, is said to be constructive in the sense that experienced events are integrated with inferences and other elaborations that go beyond direct experience (McKoon & Ratcliff, 1992). This view of memory implies that it is difficult, perhaps impossible, to know what we have actually experienced (Schwartz & Reisberg, 1991; Turvey, 1974).

The roots of constructivism in cognitive developmental theory can be traced to Piaget. In his work on constructive memory (e.g., Piaget, 1968), he demonstrated improvements in memory (e.g., for correctly seriated arrays) with advances in reasoning. Although Piaget and Inhelder (1973) distinguished between memory in the strict sense (memory for presented stimuli) and memory in the broad sense (the product of understanding), constructivism was still assumed: "They argued that memory in the broad sense inevitably influences memory in the strict sense" (Siegler, 1991, p. 199). Subsequent research extended these ideas to distortion in memory for logically impossible stimuli (Furth, Ross, & Youniss, 1974; Liben, 1974; Samuels, 1974). Memory distortion of such stimuli in the direction of logically possible patterns increased with age, a trend that was again attributed to advances in reasoning.

Paris and Carter (1973) studied constructive memory for sentences by adapting Bransford and Frank's (1971) *false recognition* paradigm for use with children. Paris and Carter presented children sentences containing spatial relationships much like those that had been given to adults (see the Appendix for examples) and obtained comparable results. Children, too, falsely recognized true inferences, a result routinely cited in textbooks

as evidence that children's memory is constructive (e.g., Bjorklund, 1989; Siegler, 1991). (These "recognitions" are false because children are explicitly instructed to accept only the exact sentences that were presented.) As in the Piagetian research, the tendency to falsely recognize unpresented inferences has generally been attributed to the development of reasoning (e.g., Bjorklund, 1989; Johnson & Scholnick, 1979; Prawatt & Cancelli, 1976; Weismer, 1985). The prediction that such false recognitions should increase with development follows from the constructivist assumption that "the schemata used by the memory are borrowed from the intelligence" (Piaget & Inhelder, 1973, p. 382). As development progresses, therefore, certain kinds of memory errors—those that are consistent with correct reasoning—should become more common (see also Liben, 1977).

However, evidence for this prediction is mixed. For example, although two experiments identified differences in false recognitions according to cognitive level (Johnson & Scholnick, 1979; Prawatt & Cancelli, 1976), Liben and Posnansky (1977) concluded that "the degree of constructive memory was not related to children's performance on Piagetian transitivity tasks" (p. 1490). Age trends are similarly inconsistent. For example, Paris and Carter (1973) failed to find developmental differences in types of false recognitions: "The pattern of errors is identical for both grade levels" (p. 111). Indeed, younger children were more likely to misrecognize true inferences (as well as other sentence types). Finally, in some articles, different age trends have been reported for different conditions (e.g., Brown, Smiley, Day, Townsend, & Lawton, 1977; Paris & Mahoney, 1974).

Despite the absence of clear developmental trends, the fact that unpresented sentences that preserve the meanings of presented sentences (true inferences) are misrecognized more often than unpresented sentences that do not preserve meaning (false sentences) would seem to confirm that children's memories are constructive. However, several investigators (Liben & Posnansky, 1977; Paris & Mahoney, 1974; Small & Butterworth, 1981) have pointed out that even this finding is subject to a critical methodological ambiguity. In the standard design, true test items (presented sentences and true inferences) contain words that were presented, and all false items contain new words. Therefore, it is possible that children erroneously recognize true inferences not because memory is constructive but because of lexical familiarity—true items (both old and new sentences) are more superficially similar to presented sentences. In summary, close scrutiny of the developmental literature on memory for sentences suggests that evidence for constructive memory (memory that reflects understanding) is inconclusive.

Possible Relationships Between Memory and Reasoning

Developmental studies of the false recognition paradigm have focused on constructivism, ostensibly demonstrating the effects of reasoning on memory. Although variations in both memory and reasoning have been measured, their covariation has not been assessed. Three predictions about their covariation (i.e., stochastic dependency) can be derived from differing theoretical perspectives (including constructivism). In this section, we discuss these predictions, along with manipulations (such as delay) that bear on them.

First, constructivism predicts positive dependencies between measures of reasoning (e.g., systematic misrecognition of true inferences) and measures of memory for the information on which reasoning is based (e.g., recognition of presented sentences). This is because judgments about inferences and about presented sentences are both affected either by common schemata (e.g., Piaget & Inhelder, 1973) or by their consistency with a semantic representation of presented information (e.g., Paris & Carter, 1973). Therefore, the two judgments share variability. Positive dependency is not predicted only by constructivism, however. Many information-processing theories incorporate the assumption of necessity, that is, that memory for inputs (e.g., for presented sentences) must be preserved until reasoning can operate on them (e.g., Bryant & Trabasso, 1971; Smedslund, 1969). Necessity also predicts positive dependency, because reasoning accuracy is limited by memory accuracy. Paris and Carter (1973), for example, made this claim, arguing that, in the false recognition paradigm, subjects should be more likely to commit the inferential error when they correctly remember the premises.

But, in an important variant of the false recognition paradigm, Loftus's (e.g., 1979) misinformation procedure, theorists have predicted the opposite relationship. In discussing misinformation effects, a number of authors have suggested that better memory decreases false recognitions (Brainerd & Reyna, 1988; Ceci, Ross, & Toglia, 1987; Ceci, Toglia, & Ross, 1988; Howe, 1991; Loftus & Hoffman, 1989; Loftus, Levidow, & Duensing, 1992; Tousignant, Hall, & Loftus, 1986). Loftus et al. (1992) for example, called this the *principle of discrepancy detection:* The better the memory for original events, the easier it is to reject mismatches. This leads to the prediction of negative dependencies between memory and reasoning, because better memory for what was presented should allow subjects to better detect what was not presented. Not surprisingly, misinformation theorists have also argued that older children should be less prone to misrecognition because of developmental improvements in memory (e.g., Ceci & Bruck, 1993; Ceci et al., 1988; Howe, 1991; Loftus et al., 1992).

Fuzzy-trace theory (e.g., Reyna & Brainerd, 1990, 1992) leads to a third prediction, namely, reasoning–remembering independence. This prediction turns on the fact that misrecognition of true inferences cannot be based on verbatim sentence memory because such sentences were never presented. If true inferences are systematically misrecognized, it is because of their consistency with the *gist* of presented sentences. (By "systematically," we mean misrecognized more often than false sentences with familiar words.) According to fuzzy-trace theory, both gist and verbatim memories are based on inputs, but they are stored separately. It follows, then, that memory for gist can be accessed without accessing verbatim memory, and vice versa. Therefore, if children access sentence gist when they misrecognize true inferences but they access verbatim sentence memory when they recognize presented sentences, those judgments will be independent of one another because verbatim and gist traces are functionally dissociated (e.g., Brainerd & Reyna, 1992; Reyna & Brainerd, 1991a, 1991b). Precisely because gist and verbatim memories are not integrated, reasoning–remembering independence would be expected in the false recognition paradigm.

Prior research suggests that there are two manipulations that might bear on such predictions: sentence content and timing of

tests. Virtually all research, with children and adults, has involved sentences that express spatial relationships. As early as 1903, however, Alfred Binet noted that spatial problems have special properties that distinguish them from other kinds of reasoning problems, such as linear inferences. As a matter of generality, therefore, it is critical to determine whether results in the false recognition paradigm are limited to this special class of materials. Similarly, although researchers have generally been careful to test subjects outside the range of short-term memory, delays have not exceeded the memory consolidation interval, usually estimated as a few hours to a few days after encoding (Brainerd, Reyna, Howe, & Kingma, 1990). Therefore, it is also important to determine the extent to which results are limited to immediate as opposed to long-term tests, reflecting acquisition versus retention processes, respectively.

Experiment 1: Verbatim Recognition Memory

The purpose of the first experiment was to investigate reasoning–remembering relationships by varying memory tests so as to disentangle effects of verbatim versus gist representations. In this connection, we adopted the modifications used by Paris and Mahoney (1974) and Liben and Posnansky (1977) to unconfound meaning and lexical familiarity. This allowed us to compare misrecognition of true (but unpresented) sentences with false sentences that had familiar wording. A preference for the former, then, can be attributed to memory for gist rather than memory for surface form.

As in psycholinguistic research generally, gist is defined as an abstract representation of semantic content that does not incorporate details of surface form (e.g., Clark & Clark, 1977; Glucksberg & Danks, 1975; Jarvella, 1971). Consistent with this definition, we tested recognition of true inferences as well as differently worded versions of presented premises (i.e., *paraphrases* of premises). By including true paraphrases, we could separate constructive memory that was the result of reasoning—misrecognition of true inferences—from constructive memory for the meaning of individual sentences.

To assess memory for *surface form* (so-called *verbatim memory*; Reyna, 1992), we compared acceptance rates for presented sentences to equally true sentences that were not presented. Thus, *yes* responses to true premises with original wording (presented sentences) would be hits, and *yes* responses to other true sentences would be false alarms. The degree to which hits exceed such false alarms indicates discrimination of surface form. For example, if children discriminate presented premises from true paraphrases, they must remember exact wording because these sentences are synonymous.

The present experiment differed from previous work because it compared linear and spatial sentences, tested children at long as well as at short retention intervals, and, most important, directly assessed dependency between memory and reasoning. Further, we changed the instructions by providing explicit examples of true sentences that children were supposed to reject (as well as examples of other sentence types). Such instructions are pivotal, because subjects ordinarily assume that they should respond on the basis of meaning (Alba & Hasher, 1983). Therefore, we avoided the possibility that children's responses would be consistent with gist (and so appear to be constructive) only because they misunderstood the task. Using this modified false recognition paradigm, we examined children's memory for gist versus surface form, developmental differences in gist-based responses, and dependencies between responses (e.g., to presented premises and true inferences).

Method

Subjects. Fifty children, twenty-five 6-year-olds (M = 6.54 years) and twenty-five 9-year-olds (M = 9.64 years), participated in the experiment. The children were drawn from an elementary school serving a middle-class residential area, none was identified as learning or language disabled, and there was an equal number of boys and girls.

Design and materials. The design was a $2 \times 2 \times 2 \times 2 \times 2 \times 2$ factorial. The first factor, age (6 or 9 years old), was varied between subjects, and the other five factors were varied within subjects: (a) Each subject was tested twice, immediately after "story" presentation and again after a week's delay. Stories consisted of three sentences: two true premises plus a filler sentence. (Consistent with prior usage, we call these connected sets of sentences *stories;* however, they should be distinguished from full-blown narratives.) (b) Subjects received two types of stories: four expressed spatial relationships and four expressed linear relationships. The remaining three factors characterized recognition sentences: recognition sentences were either (c) true or false, (d) premises or inferences, and (e) contained either novel or original wording (see the Appendix for examples). In addition to these eight recognition sentences, subjects also verified the other two presented sentences: the filler sentence (a presented sentence that was not involved in constructing the true inference) and the other true original premise.

Procedure. Each child was tested orally and individually. Six different experimenters were used, and they were randomly assigned to children. In addition to securing consent from parents, the child signed an assent form before the session began. After an initial period of familiarization, children were instructed that they would hear some stories and should try to remember them. They would then hear some sentences, and their task was to say whether these sentences had been in the story. (Sentences were presented and tested orally to guard against such artifacts as differences in reading level.) Questions relevant to each story were presented immediately after that story, so the subject received eight story-test cycles. Children were carefully instructed to say yes only if "it is exactly the same as in the story" and to say no otherwise. They were also told: "Say yes if I say the same words in the same way. Say no if anything is different from the story."

The experimenter then presented a series of examples. Before the examples, children were reminded to "try to remember exactly what the story says." Children were given three types of practice recognition sentences: a false sentence that was identical to a presented sentence, except that the order of subject and object nouns was reversed (a false premise with original wording; FPO), a presented premise (true premise with original wording; TPO), and a true inference with original wording (TIO). The child provided answers for the practice examples, and, for each sentence, the experimenter explained the basis for the correct response. Experimenters were trained to repeat the practice examples if children were hesitant or gave incorrect responses, but most children responded quickly and accurately and appeared to have no difficulty understanding the instructions. For the true inference, the experimenter acknowledged that the sentence "is true," that this is what happened in the story but that "the story didn't say that. You should only say yes if I say the same words in the same way." Stories and recognition sentences were presented in a different random order for each subject. Before each story, children were reminded to try to remember the exact words.

At the long-term session, children were told that they would be asked about the stories that had been read to them the previous week and that they should answer the questions just as they did then. (Stories were not presented again.) They were also reminded that they "should say yes if

Table 1
Results of Analyses of Variance (ANOVAs) and Analyses of Covariance (ANCOVAs)

Variable	Results
	Experiment 1
Gist vs. surface form	There was a four-way interaction of delay with the sentence factors (true–false, premise–inference, original–novel wording), $F(1, 48) = 23.06$, $p < .0001$, $MS_e = 1.80$, as well as component interactions: premise–inference with true–false, $F(1, 48) = 28.92$, $p < .0001$, $MS_e = 0.06$; premise–inference with wording, $F(1, 48) = 14.82$, $p < .0004$, $MS_e = 0.04$; and true–false with wording, $F(1, 48) = 78.75$, $p < .0001$, $MS_e = 0.05$; and main effects for premise–inference, $F(1, 48) = 80.64$, $p < .0001$, $MS_e = 0.04$; for true–false, $F(1, 48) = 129.53$, $p < .0001$, $MS_e = 0.21$; and for wording, $F(1, 48) = 41.51$, $p < .0001$, $MS_e = 0.05$.
Developmental interactions	Sentence factors interacted with age: true–false, $F(1, 48) = 4.50$, $p < .04$, $MS_e = 0.21$; premise–inference, $F(1, 48) = 21.83$, $p < .0001$, $MS_e = 0.04$; and original–novel, $F(1, 48) = 5.72$, $p < .03$, $MS_e = 0.05$. The four-way interaction with age was also significant, $F(1, 48) = 13.03$, $p < .0008$, $MS_e = 0.07$. Both true–false and
Spatial vs. linear	True–false interacted with spatial–linear, $F(1, 48) = 9.79$, $p < .004$, $MS_e = 0.04$. Age interacted with spatial–linear and true–false, $F(1, 48) = 4.87$, $p < .04$, $MS_e = 0.07$. Premise–inference interacted with true–false and spatial–linear, $F(1, 48) = 14.66$, $p < .0005$, $MS_e = 0.05$.
Long-term retention	Premise–inference, $F(1, 48) = 10.41$, $p < .003$, $MS_e = 0.05$, and wording, $F(1, 48) = 7.82$, $p < .008$, $MS_e = 0.04$, interacted with delay. True–false, delay, and premise–inference interacted, $F(1, 48) = 10.11$, $p < .003$, $MS_e = 0.03$, as did true–false, delay, and wording, $F(1, 48) = 14.71$, $p < .0005$, $MS_e = 0.04$. When acquisition differences were partialed out using ANCOVA, a three-way interaction remained for the sentence factors, $F(1, 47) = 4.19$, $p < .05$, $MS_e = 0.05$. Also, age interacted with true–false, $F(1, 47) = 6.02$, $p < .02$, $MS_e = 0.08$. Spatial–linear interacted with true–false, $F(1, 47) = 7.52$, $p < .009$, $MS_e = 0.08$. Spatial–linear and true–false also interacted with premise–inference, $F(1, 47) = 15.76$, $p < .0003$, $MS_e = 0.05$. Spatial–linear also interacted with premise–inference, true–false, and age, $F(1, 47) = 4.50$, $p < .04$, $MS_e = 0.05$.
	Experiment 2
Gist vs. surface form	True differed from false, $F(1, 48) = 172.91$, $p < .0001$, $MS_e = 0.44$; and true–false interacted with age, $F(1, 48) = 8.68$, $p < .005$, $MS_e = 0.44$. In an ANOVA from which false sentences were omitted, differences between true sentences with original versus novel wording were significant, $F(1, 48) = 18.01$, $p < .0002$, $MS_e = 0.03$. True–false interacted with delay, $F(1, 48) = 21.22$, $p < .0001$, $MS_e = 0.07$. ANCOVA results indicated that original wording differed from novel wording, $F(1, 47) = 4.19$, $p < .05$, $MS_e = 0.04$.
Spatial vs. linear	True–false interacted with spatial–linear, $F(1, 48) = 5.60$, $p < .03$, $MS_e = 0.26$. Premise–inference interacted with spatial–linear, $F(1, 48) = 6.32$, $p < .02$, $MS_e = 0.04$, and there was a three-way interaction with true–false, $F(1, 48) = 11.21$, $p < .002$, $MS_e = 0.07$. Delay interacted with spatial–linear, $F(1, 48) = 4.90$, $p < .04$, $MS_e = 0.03$. ANCOVA results indicated that spatial differed from linear relationships, $F(1, 47) = 4.40$, $p < .05$, $MS_e = 0.04$. Also, true–false interacted with spatial–linear, $F(1, 47) = 13.32$, $p < .0008$, $MS_e = 0.13$. Spatial–linear affected true premises and inferences differently, $F(1, 47) = 7.86$, $p < .008$, $MS_e = 0.07$.

I say the same words in the same way and to say no if I say anything different." The order of testing of stories was randomized again, as was the order of recognition sentences, separately for each subject. Before the recognition sentences for each story, children were again told to "try to remember the exact words."

Results and Discussion

The frequency of affirmative recognitions for each sentence type, except fillers, was summed across stories. These frequencies were then input to analyses of variance (ANOVAs) and analyses of covariance (ANCOVAs; for results, see Table 1). All differences were confirmed by Tukey honestly significant difference (HSD) tests.

Memory for gist versus surface form. Figure 1A displays acceptance rates for each sentence type on the immediate test (the test most comparable with previous studies). The likelihood of recognizing a presented premise (TPO) was 36% greater than for its nearest true competitors (true premise with novel wording [TPN] = TIO), indicating excellent memory for surface form. Such a pattern does not support constructivism in that the largest differences were between presented sentences and everything else, regardless of truth. This differentiation according to surface form is exactly the opposite of the findings that have been used to support constructivism, although it is in keeping with more recent results with adults (Alba & Hasher, 1983; McKoon & Ratcliff, 1992). If children's memory is constructive (i.e., based on gist), then it is difficult to explain how children can easily discriminate among true sentences, all of which are consistent with gist. Notably, children were even able to discriminate between presented premises and premise paraphrases (TPN sentences) with the same meaning, a datum that does not square with the assumption that recognition is based on the sentence as understood.

However, gist contributed significantly to recognition judgments, as the comparison between true unpresented sentences and lexically familiar false sentences makes clear: There was a difference of 18% (compared with FPO) favoring both TPN and TIO sentences (although this premium shrank to 7% for true inferences with novel wording [TIN]). Superficial familiarity cannot explain such premiums because TPN sentences, for ex-

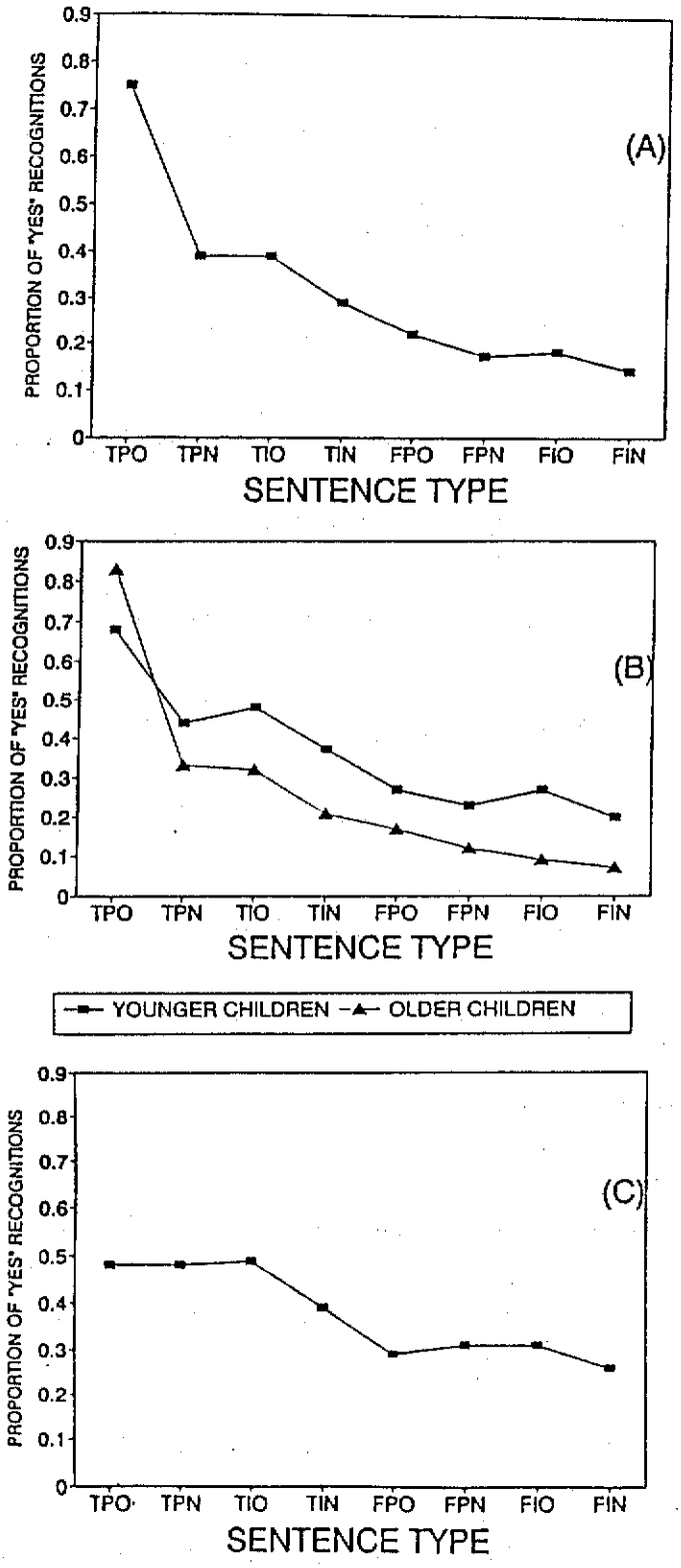

Figure 1. Mean proportion of affirmative recognitions for different sentence types in Experiment 1: (A) at acquisition; (B) separately for 6- and 9-year-olds at acquisition; and (C) at long-term retention. TPO = true premise with original wording; TPN = true premise with novel wording; TIO = true inference with original wording; TIN = true inference with novel wording; FPO = false premise with original wording; FPN = false premise with novel wording; FIO = false inference with original wording; FIN = false inference with novel wording. Retention means are adjusted for acquisition.

ample, are less similar to presented sentences than are FPO sentences (see the Appendix). We can ascribe such differences to gist because children favored less familiar true sentences over more familiar false ones. In summary, there was evidence of both memory for surface form as well as systematic misrecognition of gist on immediate tests.

1. Developmental interactions. As Figure 1B shows, younger and older children differed in their discrimination between presented sentences (TPOs) and other true sentences (see Table 1). The difference for older children between presented sentences and the closest unpresented true sentences (TPN) was 50%, almost twice the difference for younger children (24%). Thus, older children had better verbatim recognition (indicated by the crossover in Figure 1B). For the remainder of the graph, however, the lines for older and younger children are parallel. That is, misrecognition based on gist (the difference favoring true unpresented sentences over FPOs) was about the same for older and younger children. In terms of development, memory for surface form and misrecognition of gist appear to be independent: Gist-based errors were not lower in older subjects despite their better recognition of surface form.

Earlier studies stressed advances in logical reasoning across this age range, predicting a changing pattern of memory errors for true inferences versus false sentences. If we assume that reasoning does improve from ages 6 to 9, then, according to constructivism, although overall errors might decline, systematic errors (that reflect correct reasoning) should increase. This did not occur. Older children were less likely than younger ones to affirm true inferences, and the differences between true inferences and false sentences were about the same across age. This result disconfirms the prediction that memory distortions consistent with logical reasoning (e.g., misrecognition of true inferences) increase with age (e.g., Bjorklund, 1989; Johnson & Scholnick, 1979; Prawatt & Cancelli, 1976).

2. Spatial versus linear relationships. As noted earlier, most studies of false recognition in children have used spatial relationships. Type of relationship, spatial or linear, however, made a difference. Children were more likely to respond to truth when sentences described spatial relationships; surface familiarity mattered more for linear than for spatial sentences (Figure 2). These spatial–linear differences were confined to memory items (premises) as opposed to reasoning items (inferences). This indicates that, contrary to the necessity hypothesis, better memory for premises was not associated with more misrecognitions of true inferences.

Long-term retention. The impact of surface form declined across the retention interval: Differences between premises and inferences and between novel and original wording diminished with delay. Also, although there was no overall interaction between truth and delay, these factors did interact with surface cues; different forms having the same truth value became assimilated over time. Children's retention scores, however, generally depend on both initial acquisition level and subsequent

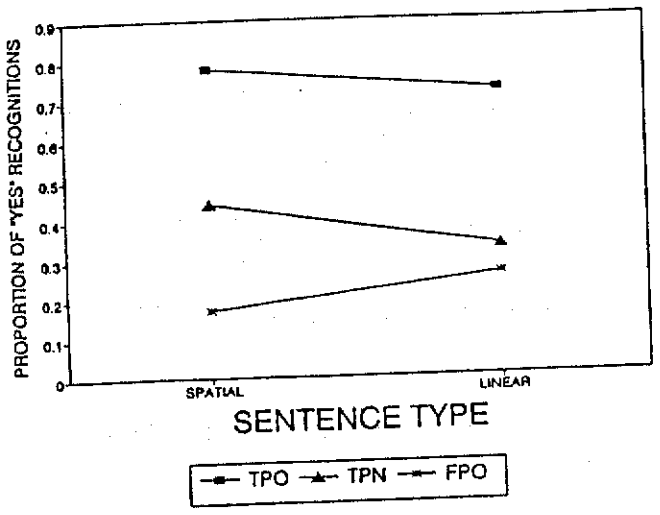

Figure 2. Mean proportion of affirmative recognitions for spatial and linear premises at acquisition in Experiment 1. TPO = true premise with original wording; TPN = true premise with novel wording; FPO = false premise with original wording.

forgetting (e.g., Brainerd & Reyna, 1990). To separate these effects, we performed an ANCOVA in which acquisition responses for each category of recognition sentence were covariates for the corresponding retention responses (a technique used in a similar context by Giambri & Arenberg, 1990).

At acquisition, presented premises (TPO) had differed from the rest of the true sentences (that had not been presented). As Figure 1C shows, however, this difference was entirely confined to acquisition—no difference remained between TPO and either TPN or TIO when acquisition differences were partialed out. (See adjusted cell means in Figure 1C). After the delay, the pattern among true sentences changed; now true inferences with novel wording differed from the other true sentences (see Figure 3). Novel wording and a novel combination of nouns (as in inferences) were no longer decisive cues by themselves (as they had been at acquisition). Both novelty cues were necessary to reject a true sentence. Recognition rates for all true sentences, however, continued to be higher than those for false sentences. Therefore, after a delay, consistency with the gist of information that had been presented was the major determinant of recognition, although surface features were detectable in combination.

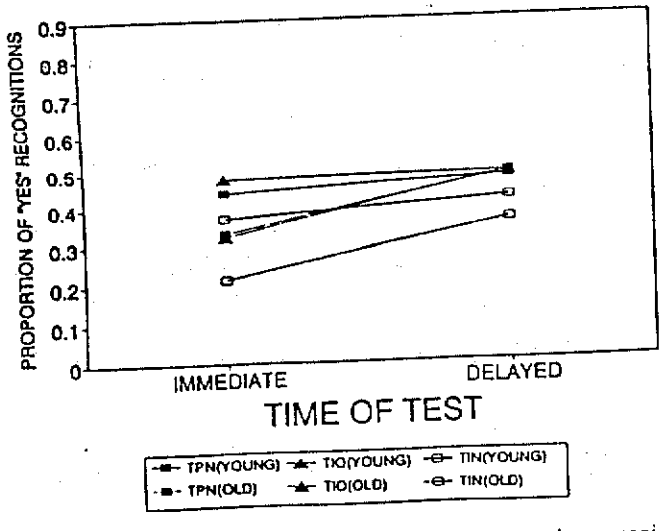

Figure 3. Age differences in mean proportion of affirmative recognitions for unpresented true sentences at acquisition versus retention in Experiment 1. TPN(Young) = true premise with novel wording for 6-year-olds; TPN(Old) = true premise with novel wording for 9-year-olds; TIO(Young) = true inference with original wording for 6-year-olds; TIO(Old) = true inference with original wording for 9-year-olds; TIN(Young) = true inference with novel wording for 6-year-olds; TIN(Old) = true inference with novel wording for 9-year-olds.

1. Developmental interactions. Unlike acquisition, the overall pattern of recognition rates across sentence types was similar for younger and older children (Figure 1C). Thus, the age groups converged at long-term retention, primarily because of forgetting of surface features in older children (cross-session differences were significant for older, but not younger, children; see Figure 3). Specifically, for older children, recognition of presented premises fell at about the same rate that misrecognition for true unpresented sentences increased. The ability to distinguish presented from unpresented true sentences deteriorated for older children (though they were still better able than younger children to reject TIN sentences because of acquisition differences). Recognition of false sentences changed little over the delay for either age group. Thus, it is not the case that one group forgot the gist of the sentences more rapidly than the other. Instead, developmental differences in forgetting had to do with memory for surface form. (This convergence for different age groups is a consequence of negatively accelerated forgetting curves [see Reyna, in press].) Older children were not responding randomly after a delay (i.e., 50 in Figure 3 does not represent chance level responding) because recognition rates differed systematically for true and false sentences; rather, they were responding more to gist.

2. Spatial versus linear relationships. Spatial–linear differences, including the dissociation between premises and inferences, became more exaggerated over time. Although children retained the gist of spatial premises better than that of linear premises, they were less likely to affirm true spatial inferences compared with linear inferences. Delay increased both variations in premise memory and misrecognitions of true inferences, either of which, according to necessity, ought to produce stronger associations between premises and inferences. As forgetting occurs, constructivists also predict increased integration between reasoning and memory (e.g., Bartlett, 1932; Liben, 1977; Piaget & Inhelder, 1973). However, dissociation between memory for premises versus misrecognition of inferences remained at retention and became more pronounced.

Reasoning–remembering relationships. The crux of fuzzy-trace theory's explanation of memory independence is that children's reasoning typically does not draw on verbatim traces of problem facts, although such representations are often used to answer memory questions. Instead, reasoning questions are usually answered by processing gist. Therefore, reasoning–remembering independence is really a question of gist–verbatim independence. For this task, then, verbatim judgments of pre-

sented premises, and gist-based judgments of the inferences that follow from those premises, should be independent.

Such independence requires that both gist and verbatim representations remain accessible, however (Reyna, 1992). It is a well-known finding in psycholinguistic research that verbatim representations become inaccessible at a faster rate over a delay than gist representations (Clark & Clark, 1977; Kintsch, Welsch, Schmalhofer, & Zimny, 1990). Similarly, verbatim representations were more accessible to older than younger children (at both immediate and delayed tests). Finally, verbatim representations were apparently accessed more readily in judging linear as opposed to spatial premises (e.g., Figure 2). Therefore, independence should be more likely at immediate tests, for older children, and for linear sentences (see, also, Reyna, 1992).

Alternatively, when children systematically misrecognize true inferences at above-chance levels (which they did), necessity means that those inferences must have been based on mental representations of presented premises. Hence, the conditional probability of saying yes to a true inference, given accurate memory representations for premises, should be higher than the unconditional probability (positive dependency). Conversely, if children say yes to true inferences because they forget exactly what was presented, the conditional probability should be lower than the unconditional (negative dependency). Finally, constructivism holds that memory is based on meaning, so recognition of all true sentences should be related (i.e., positive dependency).

Likelihood ratio tests of the dependencies between presented premises and true inferences were computed separately for spatial and linear sentences. Although the data were collapsed across stories (separately for spatial and linear), the contingencies were computed within each story. That is, presented premises were related only to the true inferences that followed from them. Such tests indicate whether variation in memory and reasoning measures are in any way related. Unless either memory or reasoning exhibits no variation at all, even small amounts of shared variation can be detected. For example, if children derived inferences from verbatim representations of premises immediately and (contrary to constructivism) stored them separately from premises, unless verbatim memory were virtually perfect, an association between inferences and premises would still be detected.

We evaluated the relationship between each premise (the first and the second premise) and each true inference (the one with original wording and the one with novel wording), which yielded eight comparisons (four spatial and four linear) per age and time of test (for a total of 32 tests). Overall, reasoning–remembering independence predominated. Although children's misrecognitions of true inferences were clearly not random but were based on the information presented in sentences, recognizing that information was generally independent of recognizing the true inferences.

Specifically, of the four spatial premise–inference comparisons for older children at the immediate test, all yielded independence. In three of these four comparisons, the conditional and unconditional probabilities were identical. Similar results were obtained for older children after the delay. For younger children, also for spatial sentences, independence was obtained for three of the four immediate tests, but the four delayed tests indicated positive dependency. For linear sentences, independence was obtained across both ages and times of test (14 of 16 comparisons yielded independence, with the remaining 2 comparisons split between negative and positive dependency).

The result that older, more proficient, children were more likely than younger children to display independence argues against the notion that independence results from guessing or comprehension deficiencies. Instead, there was a consistent pattern of independence across materials and time of test for older children (and a consistent pattern of independence with linear sentences for both ages). Only younger children responding to spatial sentences after a delay exhibited reliable dependency. Because these same children did show independence for the same sentences on immediate tests, such shifts to dependency after a delay cannot be ascribed to age differences in motivation, understanding instructions, and the like. These children had already performed the task correctly on the immediate tests. On the other hand, if independence occurs because children access verbatim memories for presented premises, then independence should be unlikely under conditions in which those verbatim memories were inaccessible.

The account of these results given by fuzzy-trace theory turns on the nature of representations used to verify sentences. Therefore, it leads to an additional set of predictions about between-sentences dependencies, which can be tested. If the account is correct, paraphrases of premises (TPN sentences) should be related differently to true inferences than verbatim premises (TPO sentences). Again, we can surmise that, like true inferences, true paraphrases were affirmed because of their consistency with gist, because they were favored over false sentences with greater surface familiarity (FPOs). Hence, judgments of true paraphrases and true inferences should be related because responses in both cases were based on gist (producing positive dependency). Judgments of true inferences with original wording and of true inferences with novel wording should also be related because they, too, are both consistent with gist. Thus, judgments of all unpresented true sentences should be positively dependent.

We compared judgments of premise paraphrases (TPN) to judgments for each true inference (original and novel wording) and also compared true inferences to one another (i.e., three comparisons per age, time of test, and materials for a total of 24 tests). In contrast to the results for presented premises, responses to premise paraphrases (TPN sentences) and to true inferences of both types were generally positively dependent. This result was more typical of spatial sentences than linear ones (10 of 12 comparisons for spatial sentences yielded positive dependency, compared with 7 of 12 for linear sentences), again, consistent with the idea that spatial sentences were more likely to be verified on the basis of gist. These findings for unpresented true sentences make it extremely doubtful that the independence observed for presented premises was a consequence of measurement insensitivity. Thus, dependency was obtained when gist was related to gist—either paraphrases to inferences or inferences to rewordings of those inferences. This demonstrates that the same children in the same task can maintain verbatim representations that are independent of the true inferences that follow from those representations and, simultaneously, can display dependence between synonymous paraphrases and those inferences.

Summary and conclusion. Our results differ from those of

previous studies that used the false recognition paradigm. The reasons for these differences include clarifying instructions (so that children did not assume that they were being asked about meaning) and unconfounding gist and verbatim attributes of recognition sentences. Unlike previous studies, children did not readily misrecognize true inferences. On the contrary, at acquisition, the major determinant of performance was verbatim identity, and there was substantial discrimination between verbatim sentences and other true sentences. Older children were less likely than younger ones, by a wide margin, to confuse presented sentences with true sentences that had not been presented. This pattern of clear segregation among true sentences contrasts with the findings of semantic integration that have been used to demonstrate that children's memory is constructive. However, despite highly accurate verbatim judgments, there was evidence of systematic misrecognition based on gist.

After a long-term retention interval, reliance on gist increased. The most important differences at acquisition, those involving verbatim identity, vanished. Despite initial superiority for older children, verbatim recognition performance converged for older and younger children after the delay. Although consistency with gist was the predominant influence on recognitions, and most true sentences were indistinguishable, children retained the ability to reject true inferences with novel wording. Therefore, even after a 1-week delay, memory was not constructive in that children could reject true sentences that had only two subtle changes (TIN sentences). This result runs counter to most models of sentence representation, which assume that details of sentential form are not accessible after a short interval (e.g., Kintsch et al., 1990).

The types of relationships expressed in sentences, spatial or linear, had a significant impact on recognition. In recognizing spatial sentences, children were biased toward gist, whereas superficial similarity was more decisive for linear sentences. Consistent with the idea that children represented spatial gist, the ability to discriminate between true and false statements about spatial relationships changed little over the course of a week. However, this ability declined for linear relationships, once verbatim memory for presented premises became relatively inaccessible.

At both acquisition and retention, there were significant dissociations between responses to true premises and to true inferences that followed from those premises. After a delay, differences between spatial and linear premises grew; differences between spatial and linear inferences grew also, but in the opposite direction. The most extreme premise–inference dissociations were found for older children after a delay. This pattern was echoed in the results of dependency analyses showing that responses to presented premises and true inferences remained independent for older children after a delay but sometimes became dependent for younger children. Thus, the relative accessibility of gist and verbatim representations was affected by age and also by delay and materials.

Fuzzy-trace theory would ascribe the patterns of independence and dependency to variations in the representations that children used to verify sentences: Judgments were independent when different types of representations, verbatim and gist, were interrogated, but when gist supported recognition of both sentences, judgments were dependent. Thus, misrecognizing a true inference was independent of recognizing the true premise that had been presented but was dependent on recognizing its synonymous paraphrase (TPN). Surprisingly, the relationship between true premises and true inferences did not depend on meaning but, instead, turned on whether equally true premises had or had not been presented.

Our results are similar to those that have been found with adults. Although adults appear to be adroit at remembering superficial details, they simultaneously display a host of well documented memory distortions, including suggestibility effects. Our data indicate that highly veridical verbatim memory coexists with systematic misrecognition of gist in the same individuals. However, the two types of representations did not interact. Better memory for presented sentences did not protect against misrecognition of gist (negative dependency), and, conversely, improved memory for presented information (e.g., for spatial, as opposed to linear, sentences) did not increase misrecognition of true inferences. Such findings argue against both negative and positive dependency, accounts that have been offered, respectively, by theories of suggestibility and by necessity and constructivist theories (e.g., schema theories). Instead, we find both experimental and developmental independence between gist and verbatim memory in children, in concert with predictions of fuzzy-trace theory.

Experiment 2

The findings from the first experiment left three fundamental questions about gist processing in children unanswered that were addressed in Experiment 2. First, are there age changes in the ability to recognize gist, as opposed to recognizing surface cues (e.g., Paris & Mahoney, 1974)? The data from Experiment 1 did not bear on this question because memory judgments did not inevitably reflect gist; in fact, as it turned out, the two were independent. Therefore, the ability to process gist can only be assessed by explicitly instructing children to base their judgments on gist, which was done in Experiment 2.

The second question addressed in Experiment 2 is whether recognizing the gist of true inferences, which involves reasoning, is harder than recognizing the gist of individual sentences (true paraphrases). Most theories would predict that reasoning poses additional difficulties, beyond those involved in recognizing the meaning of sentences, especially for younger children (e.g., Bjorklund, 1989; Siegler, 1991). Typically, however, studies fail to test recognition of true paraphrases and, thus, confound gist with true inferences.

Third, and finally, how are memory and reasoning related when children are told to base all of their responses on gist? Instructing children to base all their responses on gist is a simple manipulation, but one that speaks directly to fuzzy-trace theory's explanation of reasoning–remembering independence. If independence is the result of tapping verbatim memory in one case and gist in the other, then basing both responses on gist should produce reasoning–remembering dependency (Reyna, 1992). By asking about gist, we can also resolve an ambiguity about dissociations between memory and reasoning for spatial and linear stimuli. It could be argued that the failure to observe beneficial effects of spatial content for inferences in Experiment 1 was a result of verbatim instructions that discouraged drawing inferences. Therefore, encouraging children to draw inferences, as in Experiment 2, provides a more direct test of the dissocia-

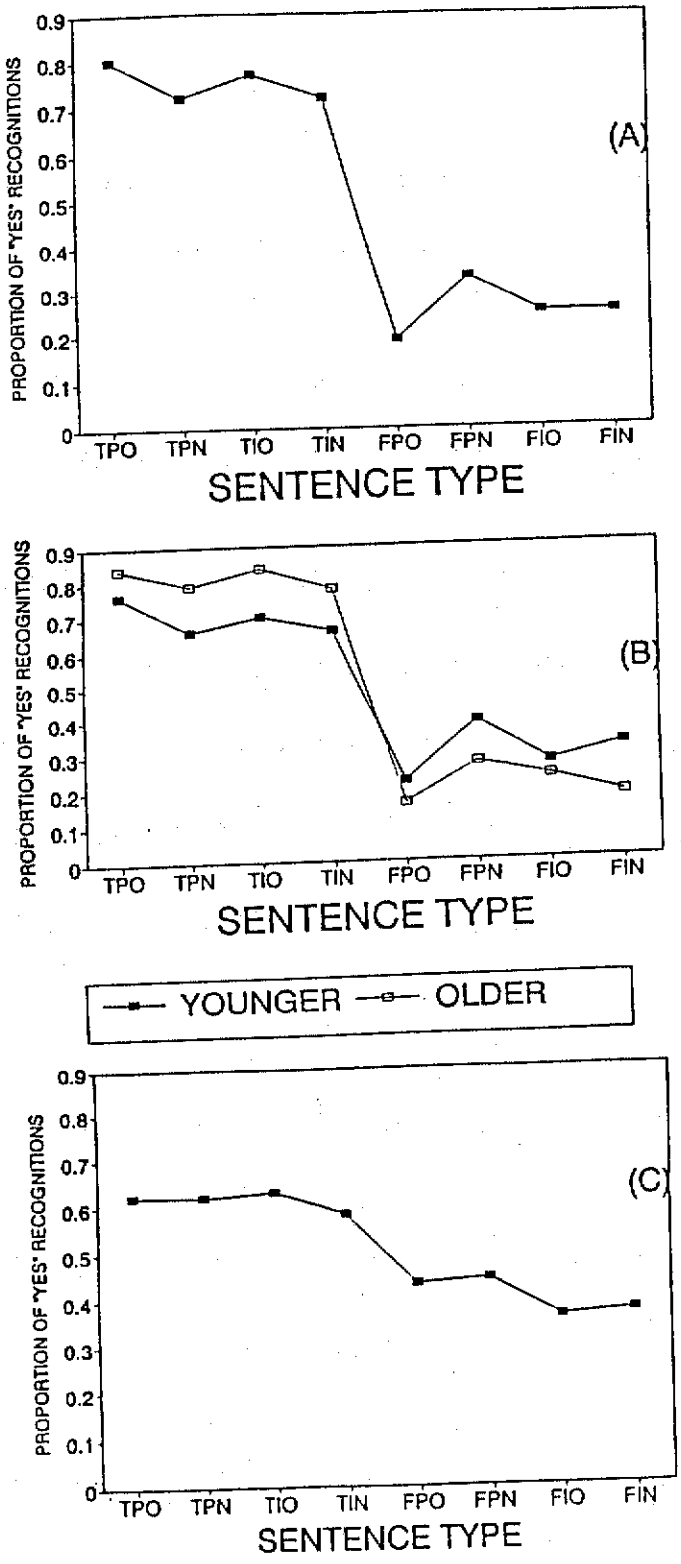

Figure 4. Mean proportion of affirmative recognitions for different sentence types in Experiment 2: (A) at acquisition; (B) separately for 6- and 9-year-olds at acquisition; and (C) at long-term retention. TPO = true premise with original wording; TPN = true premise with novel wording; TIO = true inference with original wording; TIN = true inference with novel wording; FPO = false premise with original wording; FPN = false premise with novel wording; FIO = false inference with original wording; FIN = false inference with novel wording. Retention means are adjusted for acquisition.

tion between responses to spatial and linear premises versus inferences.

Method

Subjects. Another 50 children, twenty-five 6-year-olds (M = 6.49 years) and twenty-five 9-year-olds (M = 9.58 years), participated in the second experiment. All of the children were drawn from the same school used in Experiment 1. Again, they were about evenly divided according to gender, and none was identified as learning or language disabled.

Design and materials. The design and materials were exactly like those in Experiment 1.

Procedure. The procedure, including oral and individual administration, was identical to that of Experiment 1 up to the point at which children were told that they would hear some stories. Instead of asking children to remember the stories, we instructed them to "try to understand the stories. Think about what the story means." Children were told that they would be asked about what happened in the story and that they should "say yes if that's what the story meant. Say no if the story didn't mean that."

The experimenter then presented a series of examples. Before the examples, children were reminded to "think about what the story means." Children were given three example recognition sentences: one that was identical to a presented sentence, except that the order of subject and object nouns was reversed (a false premise with original wording), a presented premise, and a true inference with original wording. Again, children provided responses to the practice examples, and most responded quickly and accurately. For each sentence, the experimenter also explained the basis for the correct response. For the true inference, the experimenter acknowledged that the story "didn't exactly say that." The experimenter then repeated the two premises and pointed out, "But, all of that means that . . .," and she repeated the true inference. False-premise original sentences were used to illustrate that surface familiarity was misleading and should be ignored. Children were told to say yes whenever sentences "mean the same thing" as what was said in the story. Stories and recognition sentences were presented in a different random order for each subject. Before each story, children were reminded to think about what the story means.

At retention, children were told that they would be asked about the stories that had been read to them the previous week and that they should answer the questions just as they did then. They were also reminded that they "should say yes if this is what happened in the story" and they should "say no if it didn't happen this way." The order of testing of stories was randomized again, as was the order of recognition sentences, separately for each subject. Children were reminded to answer according to meaning before the recognition sentences for each story.

Results and Discussion

Recognizing gist. Figure 4A displays children's accuracy in identifying sentences consistent with the gist of presented information. Sentences were sharply distinguished according to meaning. False sentences were rejected despite their superficial similarity to sentences children had heard. Among true sentences, differences were more subtle. To determine whether such differences were reliable, we conducted an ANOVA in which false sentences were eliminated. The analysis revealed

that there was a small, but consistent, difference between true sentences with original versus novel wording. However, there was no difference between true premises and inferences. Thus, although children were generally adept at identifying true premises and inferences, they were slightly less likely to do so when wording was changed. As Figure 4B shows, this zig-zag pattern among true sentences was obtained for both age groups.

Although younger and older children responded similarly to variations in surface form, overall discrimination between true and false improved with age. Given the traditional emphasis on logical limitations during this age range (as in Piagetian theory), it is important to stress that inferences did not pose any special difficulties for younger children. Errors were gist-based rather than logical in that younger children were less able to identify true statements, regardless of whether they were premises or inferences.

Long-term retention. Figure 4C shows the recognition pattern after the retention interval. Although there had been developmental differences in acquiring gist, there were no age differences in retaining it (once acquisition differences were partialed out). That is, true and false were discriminated at retention by both age groups (Figure 4C).

Spatial versus linear. The spatial–linear differences obtained in Experiment 1 were replicated. Truth affected judgments more for spatial premises than for linear ones, but lexical familiarity was more important for linear premises. Therefore, regardless of instructions, judgments of spatial premises were more likely to be based on underlying gist (further supported by greater declines in accuracy for linear sentences with delay, once verbatim representations have become relatively inaccessible). The dissociative pattern involving premises and inferences was also obtained again. Premise judgments were better for spatial sentences, but inference judgments were better for linear sentences. As in Experiment 1, delay intensified spatial–linear differences.

Reasoning–remembering relationships. In Experiment 2, children were instructed to recognize all sentences on the basis of gist. Therefore, according to fuzzy-trace theory, we should observe positive dependency between the same (presented) premises and inferences that were previously independent (Experiment 1). The results of analyses comparing linear to spatial sentences further suggest that because children are less likely to respond to the gist of linear premises, dependency should be less marked for linear sentences. Finally, if children base their judgments on gist at acquisition, as well as at retention, there should be little effect of delay.

We performed the same likelihood ratio analyses for the same comparisons made in Experiment 1. As expected, responses to presented premises and true inferences were generally found to be positively dependent. For spatial sentences, 14 of 16 tests were significant; 1 additional test was marginally significant ($p = .05$). For the linear sentences, 8 of 16 tests were significant (and, in contrast to Experiment 1, means generally differed in a positive direction for nonsignificant comparisons). Although the probability of affirming a true inference increased in Experiment 2, that alone would not have affected the contingency between recognizing true premises and true inferences. In short, we shifted the relationship between presented premises and true inferences from independence in Experiment 1 to positive dependency in Experiment 2 simply by instructing children to base all judgments on gist.

Recognitions of paraphrases (TPN sentences) and of true inferences were also generally positively dependent, consonant with our interpretation that paraphrases were verified on much the same basis as presented premises, that is, consistency with gist. Of the 12 spatial comparisons, all were positively dependent, as were 7 of the 12 linear comparisons. Despite gist instructions, linear sentences continued to display more independence between premises and inferences than spatial sentences. Such results can be accommodated by continuing to assume that children were less likely to base their responses to linear premises on gist. The absence of a systematic effect of delay on dependencies, unlike Experiment 1, suggests that children generally relied on gist at acquisition, as well as at retention, as instructed.

Just as recognition rates failed to support the idea that there were developmental differences in logical reasoning, neither did the outcomes of dependency analyses. Chapman and Lindenberger (1992), for example, argued that competent logical (i.e., operational) reasoning produces positive dependency between memory for presented premises and the drawing of true inferences, because such reasoning must process premises. (Incompetent reasoners, on the other hand, should show independence.) Of course, in this view, dependency is more likely for older children because they are more likely to reason logically. However, there were no developmental differences at acquisition (when reasoning, as opposed to forgetting, is more fairly assessed). In fact, the extent of dependency at acquisition was identical across age groups in both experiments. Second, when developmental differences between dependencies were observed (after a delay in Experiment 1), older children were less likely to show dependency than younger children.

Summary and conclusion. Overall, children were able to identify gist, older children were better than younger ones, and errors arose from an inability to see through differences in wording rather than from failures in reasoning. Such errors, however, were relatively rare. Despite the subtle differences in wording between true and false sentences, and despite the fact that children were capable of remembering verbatim features that favored acceptance of such false sentences as false-premise originals (Experiment 1), in general, children as young as 6 years were not misled by superficial similarity.

These results contrast with a number of studies indicating large developmental differences in logical and inferential ability across this age range (e.g., Brown et al., 1977; Paris & Lindauer, 1977). A key difference between the present study and others is the instruction to ignore differences in wording (with examples given of misleadingly familiar false sentences), and the pattern of errors confirms that sentences with novel wording were more difficult to recognize. Thus, prior findings of developmental differences in inferential ability may have been obtained because inferences differed superficially from presented sentences (Brainerd & Reyna, 1993; Reyna, in press).

Although gist instructions increased responses that were consistent with meaning, spatial–linear differences remained. Children were less likely to recognize the gist of linear premises, but this did not carry over to linear inferences. Although it seems puzzling from a constructivist perspective that judgments of premises and inferences would be divorced from one another,

this sort of verbatim–gist dissociation is assumed in fuzzy-trace theory. Over the delay, the ability to distinguish true from false diminished for linear relationships. In contrast, the meaning of spatial premises was retained, in line with the idea that children accessed spatial gist.

Dependency analyses confirmed that, overall, when both responses were based on gist, judgments of true premises and true inferences were positively dependent. Thus, the same true premises and true inferences used in Experiment 1 produced different results, depending on instructions. Poorer memory for linear presented premises did not imply poorer reasoning, a dissociation that is inexplicable from the assumption of necessity but that is consistent with the assumption that the representations underlying judgments of presented premises and true inferences do not necessarily interact.

General Discussion

The implications of the findings are discussed under three headings. First, we consider what the findings tell us about reasoning–remembering relationships, especially from the standpoint of constructivism and the necessity hypothesis. Next, we examine major developmental trends. There, we note some key differences between our results and those of earlier studies. Finally, we suggest a possible reconciliation of constructivist and realist perspectives on reasoning–remembering relationships. We emphasize that the degree of support that is obtained for either position is dependent on methodological choices such as encoding instructions (verbatim vs. gist in Experiments 1 and 2, respectively), the type of material that is presented (spatial vs. linear sentences), and the time at which performance is measured (immediately vs. after a retention interval).

Reasoning–Remembering Relationships

A view of memory presented in many textbooks is that it is constructive, and increasingly so with age. Constructivism, as instantiated in theories such as Piaget's, is typically contrasted with realism, the latter being the representation of events as they are, rather than as they are understood (e.g., Turvey, 1974). In this view, memory draws on reasoning, and in any representation of events, the two are linked. The weaker assumption of necessity also predicts that memory and reasoning are linked, but only in the sense that reasoning operates on memory for information (and, so, for example, variations in encoding problem information into memory ought to affect problem solutions).

Our data contradict both necessity and constructivism. When properly instructed, young children differentiated sentences that were actually presented from synonymous paraphrases and from true inferences, both of which differed only subtly in surface features from presented sentences. Such findings are not consistent with constructivism, especially as there was no significant dependency between memory and reasoning. However, there was evidence of significant misrecognition based on gist, despite highly accurate verbatim representations. Children misrecognized true inferences more often than chance, but those misrecognitions were not linked to memory for presented premises. Instead, children independently recognized presented premises and true inferences, and experimental manipulations affected recognition of premises (memory) and inferences (reasoning) differently. For example, regardless of instructions, children were more likely to respond to gist for spatial sentences, but they were more likely to respond to verbatim cues for linear sentences. However, better comprehension and memory for spatial premises did not foster spatial inferences, which were accepted at a lower rate than linear inferences. Thus, premises and inferences did not operate in tandem, and the developmental trend was not toward greater cohesion but toward greater dissociation.

These dissociations are explained in fuzzy-trace theory by assuming separate gist and verbatim representations. On immediate tests, both gist and verbatim representations are relatively accessible, especially for older children who more efficiently acquire verbatim representations. The quality of the verbatim representation at acquisition, however, did not predict the level of misrecognitions of gist, which was the same for younger and older children, nor did it predict increases in misrecognitions over the delay. Despite their initial verbatim superiority, older children's accuracy decreased rapidly and the loss of accuracy was limited to discrimination of the correct surface form among true sentences.

In summary, we observed experimental independence (differential effects of the spatial–linear factor on memory vs. reasoning) and stochastic independence between memory for presented premises and misrecognition of true inferences. Our conclusions about separate representational systems—gist and verbatim—rest not only on such findings but, more specifically, on the systematic effects of instructions to consult one type of representation or the other and on the effects of varying features of recognition sentences that are either consistent with gist or with verbatim representations. Assumptions about gist and verbatim representations allow us to account for both the appearance (with instructions to disregard verbatim memory in Experiment 2) and the disappearance (when verbatim recognition of presented sentences is related to misrecognition of gist in Experiment 1) of dependencies between reasoning and remembering.

Developmental Issues

Conclusions about ontogenetic changes in misrecognition of gist were qualified by acquisition–retention asymmetries. In situations in which misrecognition of gist has been previously reported in the false recognition paradigm (on immediate tests), our subjects generally did not mistake true inferences for sentences that had been presented, and such misrecognitions did not increase with age. In fact, discrimination among true sentences was greater in older children, and their judgments were more likely to exhibit verbatim–gist independence. Whereas constructivists have stressed the beneficial effects of integration, "the functional value of inferential, constructive processes, for access to memory" (Paris & Lindauer, 1977, p. 45), fuzzy-trace theory has emphasized the benefits of dissociation, consistent with the developmental trend observed in this study toward greater dissociation in older children (e.g., Brainerd & Reyna, 1993; Reyna, 1991, in press). Using the same tasks and materials, Lim (1993) found even greater verbatim–gist dissociations with adults, including a more pronounced pattern of independence and dependency.

Developmental differences of the sort expected by logicist theories (e.g., Chapman & Lindenberger, 1992) were not found, and predictions about dependencies between presented premises and true inferences based on logical development were not confirmed. Younger and older children did differ in their ability to recognize gist, but there were no developmental differences in recognizing (Experiment 2) or misrecognizing (Experiment 1) inferential gist (true inferences) as opposed to sentence meanings (paraphrases). Therefore, unconfounding gist and true inferences was revealing. Results from Experiment 2 suggest that previous findings of developmental differences in inferential reasoning are attributable to differences in processing gist generally (including sentence meanings).

Reconciling Constructivism and Realism

One way in which constructivism might be reconciled with accurate verbatim memory is by assuming that encoding is unbiased but that verbatim representations evolve into gist, so that memory becomes constructive over time (Alba & Hasher, 1983). This proposal is reminiscent of early gestalt analyses of forgetting (e.g., Wulf, 1922). Can we conclude that memory is veridical at acquisition but becomes constructive if we wait to measure it? Our data suggest that the answer to this question is no for three reasons.

First, the degree to which recognition appeared to be constructive, and when, differed depending on materials. At the extremes, recognition of spatial premises was biased by gist even on immediate verbatim tests (Experiment 1), but gist was unlikely to influence responses for linear relationships even at delayed tests, despite instructions to the contrary (Experiment 2). Second, even after a week, young children who received verbatim instructions were still able to detect some true sentences that were subtly different from presented ones. This occurred in the absence of special techniques that enhance verbatim memory performance after long intervals, such as repeated questioning (Brainerd et al., 1990).

Finally, the idea that verbatim representations evolve into gist leads to the same prediction that has been made by suggestibility theorists, namely that, as representations of original events fade, suggestions can supplant or overwrite those representations (Reyna, in press; Schwartz & Reisberg, 1991). This amounts to a prediction of negative dependency between the quality of original memories and the tendency to succumb to misrecognitions. However, misrecognizing a true inference as having been presented cannot be taken to imply that subjects fail to retain an accurate representation of original events, because the two were found to be stochastically independent. That is, a response indicating reasoning in the false recognition paradigm (e.g., saying yes to a true inference) does not imply a memory error because the representational systems used in the two judgments appear to be distinct.

A related pattern of findings for children's scientific problem solving led Reyna and Brainerd (1992) to argue that gist and verbatim representations coexist from the earliest stages of encoding until long after presentation and that their expression depends on the nature of testing (see also Carr, Brown, & Charalambous, 1989). Our independence results also suggest that we must relinquish a commonly held assumption in research on language, namely, that gist is derived from verbatim representations. In contrast, we accommodate the persistent findings of reasoning–remembering independence by assuming separate gist and verbatim representations. Such a "divided mind" explains how the truth of an inference can be recognized independently of the verbatim premises on which it was based.

Independence also explains how effects that seem to be evidence of constructivism can be present without implying contamination of verbatim memories. Traditionally, constructivism has been taken to entail memory integration. Demonstrating that memory was constructive, therefore, often implied that subjects could not recover "what really happened." The degree to which memory appears to be either veridical or constructive (integrative), however, may depend on methodological choices that affect which kind of representation, gist or verbatim, is accessed (Reyna, in press). These choices include encoding instructions, time of testing, and type of verbal materials (e.g., spatial or linear). As shown in this study, children can recognize verbatim sentences independently from sentences consistent with meaning or inferences, but this is more difficult for spatial relationships or after a delay. In the latter cases, although subjects misreport gist as presented, this is not because gist and verbatim memory have become integrated. Thus, memory appears to be both constructive and, to a remarkable degree, preserving of veridical details, and one type of representation neither enables nor distorts the other (i.e., is neither positively nor negatively dependent, respectively).

References

Alba, J., & Hasher, L. (1983). Is memory schematic? *Psychological Bulletin, 93,* 203–231.

Bartlett, F. C. (1932). *Remembering: A study in experimental and social psychology.* Cambridge, England: Cambridge University Press.

Binet, A. (1903). Étude expérimentale de l'intelligence [Experimental studies of intelligence]. Paris: Schleicher Freres.

Bjorklund, D. F. (1989). *Children's thinking: Developmental and individual differences.* Pacific Grove, CA: Brooks/Cole.

Brainerd, C. J., & Reyna, V. F. (1988). Memory loci of suggestibility development: Comment on Ceci, Ross, and Toglia. *Journal of Experimental Psychology: General, 117,* 197–200.

Brainerd, C. J., & Reyna, V. F. (1990). Can Age × Learnability interactions explain the development of forgetting? *Developmental Psychology, 26,* 194–204.

Brainerd, C. J., & Reyna, V. F. (1992). Explaining "memory-free" reasoning. *Psychological Science, 3,* 332–339.

Brainerd, C. J., & Reyna, V. F. (1993). Memory independence and memory interference in cognitive development. *Psychological Review, 100,* 42–67.

Brainerd, C. J., Reyna, V. F., Howe, M. L., & Kingma, J. (1990). The development of forgetting and reminiscence. *Monographs of the Society for Research in Child Development, 53*(Whole No. 222), 3–4.

Bransford, J. D., & Franks, J. J. (1971). The abstraction of linguistic ideas. *Cognitive Psychology, 2,* 331–380.

Brown, A. L., Smiley, S. S., Day, J. D., Townsend, M. A. R., & Lawton, S. C. (1977). Intrusion of a thematic idea in children's comprehension and retention of stories. *Child Development, 48,* 1454–1466.

Bryant, P. E., & Trabasso, T. (1971). Transitive inference and memory in young children. *Nature, 232,* 456–458.

Carr, T. H., Brown, J. S., & Charalambous, A. (1989). Repetition and reading: Perceptual encoding mechanisms are very abstract but not very interactive. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 15,* 763–778.

Ceci, S. J., & Bruck, M. (1993). The suggestibility of the child witness:

A historical review and synthesis. *Psychological Bulletin, 113,* 403–439.

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

Ceci, S. J., Toglia, M. P., & Ross, D. F. (1988). On remembering . . . more or less: A reply to Brainerd and Reyna. *Journal of Experimental Psychology: General, 117,* 201–202.

Chapman, M., & Lindenberger, U. (1992). Transitivity judgments, memory for premises, and models of children's reasoning. *Developmental Review, 12,* 124–163.

Clark, H. H., & Clark, E. V. (1977). *Psychology and language.* New York: Harcourt Brace Jovanovich.

Furth, H., Ross, B., & Youniss, J. (1974). Operative understanding in children's immediate and long-term reproductions of drawings. *Child Development, 45,* 63–70.

Giambri, L. M., & Arenberg, D. (1990, November). *Age differences in sentence memory and forgetting.* Paper presented at the annual meeting of the Psychonomic Society, New Orleans, LA.

Glucksberg, S., & Danks, J. (1975). *Experimental psycholinguistics: An introduction.* Hillsdale, NJ: Erlbaum.

Howe, M. L. (1991). Misleading children's story recall: Forgetting and reminiscence of the facts. *Developmental Psychology: 27,* 746–762.

Jarvella, R. J. (1971). Syntactic processing of connected speech. *Journal of Verbal Learning and Verbal Behavior, 10,* 409–416.

Johnson, J. W., & Scholnick, E. K. (1979). Does cognitive development predict semantic integration? *Child Development, 50,* 73–78.

Kintsch, W., Welsch, D., Schmalhofer, F., & Zimny, S. (1990). Sentence memory: A theoretical analysis. *Journal of Memory and Language, 29,* 133–159.

Liben, L. S. (1974). Operative understanding of horizontality and its relation to long-term memory. *Child Development, 45,* 416–424.

Liben, L. S. (1977). Memory in the context of cognitive development: The Piagetian approach. In R. V. Kail & J. W. Hagen (Eds.), *Perspectives on the development of memory and cognition* (pp. 297–332). New York: Plenum Press.

Liben, L. S., & Posnansky, C. J. (1977). Inferences on inference: The effects of age, transitive ability, memory load, and lexical factors. *Child Development, 48,* 1490–1497.

Lim, P. (1993). *Meaning versus verbatim memory in language processing: Deriving inferential, morphological, and metaphorical gist.* Unpublished doctoral dissertation, University of Arizona.

Loftus, E. F. (1979). *Eyewitness testimony.* Cambridge, MA: Harvard University Press.

Loftus, E. F., & Hoffman, H. G. (1989). Misinformation in memory: The creation of new memories. *Journal of Experimental Psychology: General, 118,* 100–104.

Loftus, E. F., Levidow, B., & Duensing, S. (1992). Who remembers best?: Individual differences in memory for events that occurred in a science museum. *Applied Cognitive Psychology, 6,* 93–107.

McKoon, G., & Ratcliff, R. (1992). Inference during reading. *Psychological Review, 99,* 440–466.

Paris, S. G., & Carter, A. Y. (1973). Semantic and constructive aspects of sentence memory in children. *Developmental Psychology, 9,* 109–113.

Paris, S. G., & Lindauer, B. K. (1977). Constructive aspects of children's comprehension and memory. In R. V. Kail & J. W. Hagen (Eds.), *Perspectives on the development of memory and cognition* (pp. 35–60). Hillsdale, NJ: Erlbaum.

Paris, S. G., & Mahoney, G. J. (1974). Cognitive integration in children's memory for sentences and pictures. *Child Development, 45,* 633–643.

Piaget, J. (1968). *On the development of memory and identity.* Worcester, MA: Clark University Press.

Piaget, J., & Inhelder, B. (1973). *Memory and intelligence.* New York: Basic Books.

Prawatt, R. S., & Cancelli, A. (1976). Constructive memory in conserving and nonconserving first graders. *Developmental Psychology, 12,* 47–50.

Reyna, V. F. (1991). Class inclusion, the conjunction fallacy, and other cognitive illusions. *Developmental Review, 11,* 317–336.

Reyna, V. F. (1992). Reasoning, remembering, and their relationship: Social, cognitive, and developmental issues. In M. L. Howe, C. J. Brainerd, & V. F. Reyna (Eds.), *Development of long-term retention* (pp. 103–127). New York: Springer-Verlag.

Reyna, V. F. (in press). Interference effects in memory and reasoning: A fuzzy-trace theory analysis. In F. N. Dempster & C. J. Brainerd (Eds.), *New perspectives on interference and inhibition processes in cognition.* San Diego, CA: Academic Press.

Reyna, V. F., & Brainerd, C. J. (1990). Fuzzy processing in transitivity development. *Annals of Operations Research, 23,* 37–63.

Reyna, V. F., & Brainerd, C. J. (1991a). Fuzzy-trace theory and children's acquisition of mathematical and scientific concepts. *Learning and Individual Differences, 3,* 27–59.

Reyna, V. F., & Brainerd, C. J. (1991b). Fuzzy-trace theory and framing effects in choice: Gist extraction, truncation, and conversion. *Journal of Behavioral Decision Making, 4,* 249–262.

Reyna, V. F., & Brainerd, C. J. (1992). A fuzzy-trace theory of reasoning and remembering: Paradoxes, patterns, and parallelism. In A. Healy, S. Kosslyn, & R. Shiffrin (Eds.), *From learning processes to cognitive processes: Essays in honor of William K. Estes* (Vol. 2, pp. 235–259). Hillsdale, NJ: Erlbaum.

Samuels, M. (1974). *Children's long-term memory for events.* Unpublished doctoral dissertation, Cornell University.

Schwartz, B., & Reisberg, D. (1991). *Learning and memory.* New York: Norton.

Siegler, R. S. (1991). *Children's thinking.* Englewood Cliffs, NJ: Prentice Hall.

Small, M. Y., & Butterworth, J. (1981). Semantic integration and the development of memory for logical inferences. *Child Development, 52,* 732–735.

Smedslund, J. (1969). Psychological diagnostics. *Psychological Bulletin, 71,* 237–248.

Tousignant, J. P., Hall, D., & Loftus, E. (1986). Discrepancy detection and vulnerability to misleading postevent information. *Memory and Cognition, 14,* 329–339.

Turvey, M. T. (1974). Constructive theory, perceptual systems, and tacit knowledge. In W. B. Weimer & D. S. Palermo (Eds.), *Cognition and the symbolic processes* (pp. 165–180). Hillsdale, NJ: Erlbaum.

Weismer, S. E. (1985). Constructive comprehension abilities exhibited by language-disordered children. *Journal of Speech and Hearing Research, 28,* 175–184.

Wulf, F. (1922). Über die Veranderung von Vorstellungen [Changes in performance]. *Psychologisch Forschung, 1,* 33–373.

Appendix

Examples of Stories and Corresponding Recognition Sentences

Spatial

The bird is inside the cage.
The cage is under the table.
The bird has yellow feathers.

Filler: The bird has yellow feathers.
True premise original (TPO): The bird is inside the cage.
True premise original (TPO): The cage is under the table.
True premise novel (TPN): The table is above the cage.
True inference original (TIO): The bird is under the table.
True inference novel (TIN): The table is above the bird.
False premise original (FPO): The table is under the cage.
False premise novel (FPN): The bird is above the cage.
False inference original (FIO): The table is under the bird.
False inference novel (FIN): The bird is above the table.
Additional examples of presented premises:
The sign is in front of the house.
The plane is higher than the trees.
The flowers are on the dresser.

Linear

The cocoa is hotter than the tea.
The tea is hotter than the coffee.
The cocoa is very sweet.

Filler: The cocoa is very sweet.
True premise original (TPO): The cocoa is hotter than the tea.
True premise original (TPO): The tea is hotter than the coffee.
True premise novel (TPN): The coffee is cooler than the tea.
True inference original (TIO): The cocoa is hotter than the coffee.
True inference novel (TIN): The coffee is cooler than the cocoa.
False premise original (FPO): The coffee is hotter than the tea.
False premise novel (FPN): The cocoa is cooler than the tea.
False inference original (FIO): The coffee is hotter than the cocoa.
False inference novel (FIN): The cocoa is cooler than the coffee.
Additional examples of presented premises:
The bear is older than the tiger.
The brick is heavier than the rock.
The orange is bigger than the apple.

Received February 1, 1992
Revision received August 24, 1993
Accepted August 24, 1993 ■