Available online at www.sciencedirect.com

SCIENCE DIRECT®

Developmental Review 24 (2004) 396–439

DEVELOPMENTAL REVIEW

www.elsevier.com/locate/dr

Fuzzy-trace theory and memory development

C.J. Brainerd*, V.F. Reyna

Department of Psychology, University of Texas, Arlington, USA

Received 4 August 2004

Abstract

We review recent applications of fuzzy-trace theory to memory development, organizing the presentation around two themes: the theory's explanatory principles and experimental findings about memory development that follow as predictions from those principles. The featured explanatory principles are: parallel storage of verbatim and gist traces, dissociated retrieval of verbatim and gist traces, differential survival rates for verbatim and gist traces, retrieval phenomenology, and developmental trends in verbatim and gist memory. The experimental findings come from four different areas of research: the development of retrieval phenomenologies, "reversed" developmental trends in false memory, the development of mere testing effects, and the development of false persistence in memory.

The objective of this article is to sketch recent applications of fuzzy-trace theory (FTT) to memory development. FTT is not exclusively an account of memory development but, rather, is an interdisciplinary model of cognition. It is interdisciplinary in two senses. First, FTT's empirical focus is not confined to memory phenomena. It also deals with higher reasoning abilities and attempts to integrate theories and findings in the two domains by examining relations between memory and reasoning. In fact, FTT originated as model of reasoning that was motivated by two questions (Brainerd & Reyna, 1990; Reyna & Brainerd, 1991). One was the relation between representational processes in psycholinguistics (Reyna, 1981), and the other was

* Corresponding author.

E-mail address: brainerd@uta.edu (C.J. Brainerd).

0273-2297/$ - see front matter
doi:10.1016/j.dr.2004.08.005

the relation between children's ability to solve logical reasoning problems and their ability to remember the background information that authorizes those solutions (Brainerd & Kingma, 1985). In that era, the dominant answer to both questions was provided by theories such as constructivism (e.g., Bransford & Franks, 1971; Piaget & Inhelder, 1973) and working-memory capacity (e.g., Case, 1985). Constructivism posited that representations of linguistic inputs or of the background information in reasoning problems are not veridical in any bona fide way but, rather, are transformed (constructed) by the same reasoning operations that are used to make linguistic inferences or to solve problems. Working-memory capacity posited that the retention of problem information and the application of reasoning operations that process that information to solve problems both tax a single limited-capacity system, so that the accuracy of memory for problem information and the accuracy of reasoning are both controlled by the same capacity variable. Although constructivism and working-memory capacity have appealing features and although both remain influential models of cognition to this day, they make a characteristic prediction about the relation between linguistic inference and problem solving, on the one hand, and memory performance, on the other, that conflicts with large amounts of data. The prediction is that performance in the two domains ought to be very closely intertwined, both in the sense of displaying marked statistical covariations and in the sense of being influenced by the same experimental manipulations. However, several lines of experimentation that began during the 1980s and early 1990s revealed high levels of statistical independence and experimental dissociation (e.g., Hastie & Park, 1986; Katz, Lautenschlager, Blackburn, & Harris, 1990; Reyna & Brainerd, 1990). FTT originated as an attempt to account for such results.

The other sense in which FTT is interdisciplinary is that its empirical focus includes both developmental and nondevelopmental work. The theory aims to integrate evidence that is generated by three core approaches to cognition that are normally islands unto themselves: studies of normal adults (mainstream cognitive psychology), developmental studies (mainstream cognitive development), and studies of brain mechanisms (mainstream neuropsychology). Although there is much to be said for confining one's research to one or the other of these domains, it is obvious that theoretical models that rest upon findings from only one or another of them are at best limited and at worst fundamentally flawed. For eloquent discussions of this point, one need look no further than the writings of two of developmental psychology's giants, Werner and Piaget. As Werner (e.g., 1948) was fond of observing, developmental research that is not closely informed by what is known about the end results of ontogenesis, as revealed by nondevelopmental research with adults, misses the motivation for studying ontogenesis in the first place. Likewise, as Piaget (e.g., 1967) was fond of observing, nondevelopmental research with adults that fails to be closely informed by the ontogenetic processes that produce adult cognition runs the risk of becoming mired in hypotheses that, developmentally speaking, were never feasible to begin with. FTT's emphasis upon integrating experimental evidence that is generated by these different approaches has meant that research questions that produce instructive findings with one have been studied with the others. An illustration is provided by a line of research on the role of certain types of representations in true versus false-memory reports, which began as a series of developmental studies (e.g., Brainerd, Reyna, & Kneer, 1995b; Reyna & Kiernan, 1994), evolved into adult experimentation (e.g., Brainerd & Reyna, 1998b; Reyna & Kiernan, 1995), and is continuing in neuropsychological investigations (e.g., Budson, Sullivan, Daffner, & Schacter, 2003; Goldmann et al., 2003; Verfaellie, Schacter, & Cook, 2002). Another consequence of the emphasis upon integrating data generated by these different approaches is that FTT's explanatory principles are framed so as to span all of them. In practical terms, this means that processes that are used to explain memory and reasoning phenomena in adults are assumed to undergo specific types of developmental change.

As mentioned, the focus of this paper is upon memory development, not reasoning or adult cognitive psychology or neuropsychology. The discussion that follows proceeds in two steps. First, we consider some explanatory principles that are central to FTT's account of memory development. Those principles are empirical in origin, which is to say that each is an extrapolation from a foundation of experimental findings rather than a top-down hypothesis that reflects commitment to some a priori model of memory or ontogenesis. Second, we review a series of empirical findings about memory development that fall out of these principles as predictions. Along the way, it will be seen that the prediction of new findings is a prime criterion for evaluating the adequacy of explanatory principles in FTT.

Explanatory principles

FTT's approach to memory development has two overriding characteristics. First, the principles that are used to explain memory development center on how information is represented in memory and how those representations are subsequently retrieved and preserved. Here, FTT implements a core distinction between representations that capture the surface form of experience and representations that capture the gist of experience. Second, emphasis is placed upon securing strong tests of those principles by generating new predictions about memory development, particularly counterintuitive predictions. The first characteristic forms the substance of the present section. Discussion of the second is postponed until the next section.

FTT relies primarily upon five principles to explain memory development—where, for the most part, "memory development" boils down to ontogenetic variations in true- and false-memory responses to recognition and recall tasks. Those principles are: (a) parallel storage of verbatim and gist traces, (b) dissociated retrieval of verbatim and gist traces, (c) differential survival rates for verbatim and gist traces, (d) retrieval phenomenology, and (e) developmental variability in verbatim and gist memory. We explicate each of these principles in turn, along with the base of experimental findings from which it was originally extrapolated.

Parallel storage of verbatim and gist traces

For some years, a variety of evidence has pointed to the conclusion that the human brain deposits dissociated verbatim and gist traces of experience (for a review, see Reyna & Brainerd, 1995). Verbatim traces are integrated representations of a memory target's surface form, as well as associated item-specific information (e.g., contextual cues that are coincident with a target's occurrence). Gist traces, on the other hand, are representations of semantic, relational, and other elaborative information about a memory target. A key difference between verbatim and gist information is that the former is located in the memory targets that subjects experience (e.g., the fonts in which the words on a study list are printed, whether a word is presented as a printed letter string or as a picture), whereas the latter must be added to the targets by rememberers themselves. It would be more perspicuous to say, in the latter connection, that gist information must be accessed by rememberers themselves, using target items as retrieval cues, and then stored as gist traces of the index experience. Thus, verbatim traces may be thought of as representations of rememberers' "actual" experience, and gist traces may be thought of as representations of rememberers' "understanding" of their experience.

Because the information in verbatim and gist traces originates in the same target events and because gist traces are "interpretations" of those events, common-sense supposes that the brain ought to store the two types of representations in a serially dependent fashion; that is, verbatim traces ought to be stored first, perhaps in a short-term memory buffer, with their gist then being extracted and stored using verbatim traces to search memory for pertinent semantic and relational information. Just such a serial relation between verbatim and gist storage was posited in constructivism (but see Alba & Hasher, 1983) and was fashionable for some years in psycholinguistics. Despite its intuitive appeal, many experimental findings have rendered the serial model untenable, indicating instead that target encoding initiates rapid parallel access of meaning information (Wallace, Stewart, Shaffer, & Barry, 1998). The word-superiority effect (Ankrum & Palmer, 1989) and the missing letter-effect (Moravcski & Healy, 1995) are two illustrations. The former refers to the fact that when familiar words are presented at short exposure durations (e.g., less than 100 ms) and subjects are then confronted with either a probe word to identify (e.g.,"Was the word PAPER?") or a probe letter to identify (e.g., "Did the word contain the letter R?"), word probes produce above-chance performance at shorter exposure durations than their constituent letters. The missing-letter effect refers to the fact that when subjects are asked to read a passage of text and circle all instances of a particular letter as they read (e.g., *t*), they sometimes fail to circle the letter when it appears as part of a familiar repeated word (e.g., *the*).

Perhaps the most compelling evidence for parallel storage comes from experiments in which subjects have been found to deposit semantic information about targets less than 50 ms after target onset (e.g., Abrams & Greenwald, 2000; Draine & Greenwald, 1998; Seamon, Luo, & Gallo, 1998; Seamon et al., 2002), which is far too brief an interval to fully process targets' surface forms and store verbatim traces. A mock-up of such an experiment, which is based upon a procedure that was introduced by Draine and Greenwald (1998), is shown in Fig. 1. There are two tasks, one involving gist storage and the other involving verbatim storage. In the gist task, which is shown in Fig. 1, subjects sit in front of a computer screen. They are told that they will see a series of names (e.g., JANE, FRED, SALLY, and TOM) and that

GIST STORAGE TASK

	MATCHING TRIAL:	MISMATCHING TRIAL:
FORWARD MASK	KQHYTPEQFPBYL	KQHYTPEQFPBYL
PRIME	SALLY	FRED
BACKWARD MASK	ZBLYPZMVKTHLC	ZBLYPZMVKTHLC
TARGET	JANE	JANE
JUDGMENT	MALE OR FEMALE?	MALE OR FEMALE?

Fig. 1. An experimental procedure that demonstrates storage of meaning information about words in the absence of awareness of the words themselves (based upon Draine and Greenwald, 1998).

their task is merely to decide as quickly as possible (by pressing a key) whether the gender of each name is male or female. In fact, however, three other events occur on the computer screen, *outside the subjects' awareness*, just before each name is presented: (a) a string of consonants (called a forward mask) is flashed for 33 or 55 ms; (b) a name (called a semantic prime) is flashed for 30 or 55 ms; and (c) another string of consonants (called a backward mask) is flashed for 30 or 55 ms. It is the middle event that is crucial to diagnosing rapid parallel processing of meaning. Note in Fig. 1 that there are two types of trials—namely, matching, on which the gender of the briefly flashed name is the same as that of the test word (middle of the Fig. 1), and mismatching, on which the gender of the briefly flashed name is different than that of the test word (right side of the figure). Of course, the presentation of this name is far too brief for its surface form to be fully processed and stored, and indeed, various experimental checks are used to ensure that subjects are not even aware that anything was presented. The question is whether, nonetheless, meaning processing is so fast that subjects are sometimes able to extract the word's gender. If so, a standard semantic priming effect will be observed: the gender decision about the target (JANE) will be faster when the priming word matches the target's gender (SALLY) than when it does not (FRED). As the data in Fig. 2 illustrate, such a priming effect is observed. The size of the gender priming effect in Fig. 2 is plotted on the ordinate in d' units, for both 33 and 55 ms presentation rates, with a zero value indicating no priming. On the abscissa, the subjects' awareness that a word was actually presented, which is obtained from other experimental tests, is plotted in d' units, with a zero value indicating no awareness. Two findings demonstrate fast parallel processing of meaning that is independent of the processing of words' surface forms. First, note that there is a gender priming effect at both presentation rates *when there is no awareness that a word was presented* (i.e., the points on the far left are well above zero). Second, note that increasing levels of awareness of word presentation does not produce systematic increases in the size of the priming effect. Such data seem to force the conclusion that verbatim and gist traces are stored in parallel rather than serially because memory for meaning information can be detected before verbatim traces can be stored and because increasing the likelihood of verbatim storage does not increase the strength of memory for meaning information.

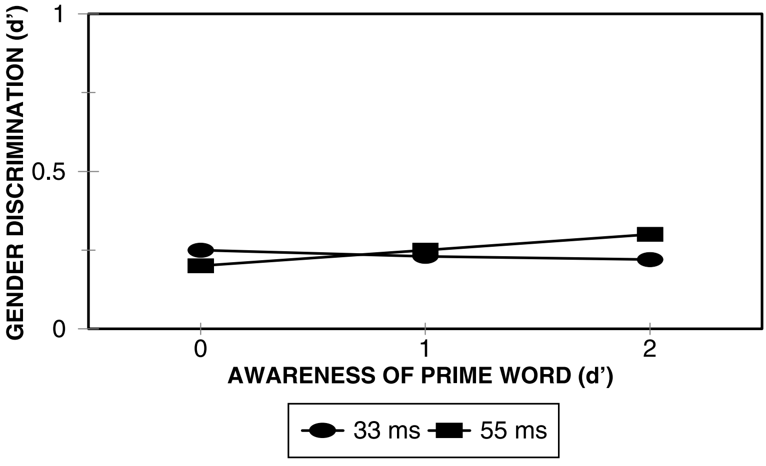

Fig. 2. Findings showing storage of meaning information about words in the absence of storage of targets' surface forms (based upon Draine and Greenwald, 1998).

With respect to verbatim storage, FTT stresses integration; that verbatim traces are cohesive representations of multiple surface features. Such integration means that when verbatim traces are retrieved on memory tests, they will induce realistic phenomenologies that reinstate the experiences that accompanied targets' earlier presentations. This leads to some surprising predictions about how subjects' tendencies to make false-memory responses are influenced by variations in the overlap between the surface properties of targets and memory tests (Reyna & Lloyd, 1997). With respect to gist storage, FTT stresses that gist traces are *episodic instantiations* of concepts (meanings, relations, and patterns) that subjects access in memory, or of elaborations that subjects generate as they encode targets' surface forms. To be more precise, as the encoding of a target's surface features begins, corresponding mechanisms of meaning access and elaboration are initiated, with verbatim and gist storage then running through to completion roughly in parallel. For that reason, subjects may be able to store considerable information about a target's meaning even though they fail to process its surface form more than partially. Moreover, subjects can store a variety of gist traces on the basis of a single target because, after all, the target items in memory experiments normally participate in multiple meanings (e.g., the word COLLIE participates in meanings such as "dog," "farm animal," and "pet").

Dissociated retrieval of verbatim and gist traces

If the brain stores separate verbatim and gist records of experience, it follows that there is potential variability in the types of representations that can be accessed on memory tests to support specific responses. This, in turn, means that the selection of retrieval cues will be of particular significance in controlling such variability. This possibility has been studied in our laboratory, and the results of that work, as well as related findings for other laboratories, have converged on the conclusion that in addition to dissociated storage, the retrieval of verbatim and gist traces are dissociated from each other. Much of our research on this point has been predicated upon a notion that grows out of the hoary rule of encoding variability (Tulving & Thomson, 1971): as long as verbatim traces are still highly accessible (see *Differential survival rates for verbatim and gist traces*), memory probes that are targets (e.g., COLLIE) ought to be better retrieval cues for their verbatim traces than memory probes that only preserve targets' meanings (e.g., POODLE), whereas memory probes that only preserve targets' meanings ought to be better retrieval cues for gist traces (''dog,'' ''pet,'' etc.) than for verbatim traces of the targets themselves. If so and if retrieval exhibits dissociation, performance on these two types of probes ought to display low levels of statistical covariation and ought to be influenced by different experimental manipulations.

Recognition designs have been used to gain experimental leverage on such possibilities because they allow the content of retrieval cues to be precisely controlled. The original studies were conducted by Reyna and Kiernan (1994, 1995), who used sentence recognition tasks. In their studies, children listened to three-sentence narratives (e.g., The bird is in the cage. The cage is on the table. The bird has green feathers.) and then responded to recognition tests that consisted of target probes (e.g., The cage is on the table.), probes that preserved the meaning of narratives (e.g., The table is under the bird.), and probes that violated the meaning of narratives (e.g., The bird is under the table.) Because recognition tests were administered immediately after each narrative, it was assumed, following the encoding variability rule, that correct recognition of targets would be based overwhelmingly upon retrieval of verbatim traces and *false* recognition of meaning-preserving distractors would be based overwhelmingly upon retrieval of gist traces (Reyna, Holliday, & Marche, 2002). When correlations between the two types of performance were computed, they were statistically independent, suggesting verbatim-gist retrieval dissociation. Consistent with the same conclusion, manipulations were identified to which the two types of performance reacted differently. A striking illustration is that narratives that revolved around spatial relations (e.g., the bird narrative) decreased correct recognition of targets and increased false recognition of meaning-preserving distractors, whereas narratives that revolved around magnitude relations (e.g., The coffee is hotter than the tea. The tea is hotter than the cocoa. The cocoa is sweet.) increased correct recognition of targets and decreased false recognition of meaning-preserving distractors. Another striking illustration is that delay also drove true and false recognition in opposite directions. When the recognition tests were repeated one week later, correct recognition was considerably lower but false recognition was considerably higher.

Both of Reyna and Kiernan's (1994, 1995) main findings, lack of statistical covariation when verbatim traces are still accessible and performance dissociation by experimental manipulations, have been replicated in children using tasks that involve verbatim and gist memory for numerical information (Brainerd & Gordon, 1994; Brainerd & Reyna, 1995) and using tasks that involve verbatim and gist memory for word lists (Brainerd et al., 1995b; Marx & Henderson, 1996). These same patterns have been obtained in normal adults (Lim, 1993; Reyna & Kiernan, 1995) and in patients with neurological impairments (Verfaellie, Rapcsak, Keane, & Alexander, 2004). Thus, the data base favoring the hypothesis of verbatim-gist retrieval dissociation is quite extensive, both with respect to subject populations and with respect to types of to-be-remembered material.

Beyond the fact that the principle of retrieval dissociation grows out of this data base, it fulfills an especially important explanatory function within FTT. It is the way that the theory explains the paradoxical duality of memory and the equally paradoxical duality of reasoning (Reyna & Brainerd, 1995). Concerning memory, contemporary research on false-memory phenomena has established the coexistence of mutually inconsistent memories of the same experience; that the same rememberer can supply true- and false-memory reports about the same event, often within a few moments each other. At first, the coexistence of conflicting memories—of subjects reporting with equal confidence that COLLIE was the single dog name on a word list and that POODLE was the single dog name on the same list—seems utterly incomprehensible. According to FTT, however, the explanation is quite elementary: memory is of two minds about experience, the verbatim and gist minds, and these minds are not well integrated with each other, owing to retrieval dissociation (Brainerd & Reyna, in press). Concerning reasoning, contemporary research on judgment and decision making tells a similar tale—specifically, a tale of the coexistence of mutually incompatible solutions to the same problems, even extremely simple problems (e.g., Tversky & Kahneman, 1981; Winer, Craig, & Weinbaum, 1992). Again, FTT's explanation is simple: different solutions result from the retrieval of different representations of problem information, and the fact that the retrieval of one type of representation tends to be dissociated from the retrieval of another allows a single reasoner to generate mutually incompatible solutions to the same problem (Reyna & Brainerd, 1995).

Differential survival rates for verbatim and gist traces

Under the principle of retrieval dissociation, whether verbatim or gist traces are accessed at any one time depends critically upon the retrieval cues that are supplied by memory tests, with verbatim access being favored by cues that reinstate the surface form of experience and gist access being favored by retrieval cues that reinstate the meaning but not the surface form of experience. As mentioned, however, there is another obvious factor that controls retrieval variability—namely, relative survival rates; the extent to which the two types of representations remain accessible in storage. This is a key consideration because many findings converge on the conclusion that verbatim traces are not preserved as well as gist traces. This differential seems to apply to both trace consolidation and to subsequent forgetting (Brainerd & Reyna, in press). In memory research, it is common to distinguish between a delimited post-event *consolidation interval*, lasting from a few minutes to a few days, during which stored representations are stabilized, and an indeterminate post-consolidation forgetting interval, when traces that survive consolidation become progressively less accessible (Eichenbaum, 2002). On the one hand, when memory tests are administered within the consolidation interval, findings from many studies suggest that verbatim traces are less likely to survive consolidation than gist traces (e.g., Bransford & Franks, 1971; Gernsbacher, 1985; Kintsch, Welsch, Schmalhofer, & Zimny, 1990; Murphy & Shapiro, 1994). A major reason is that verbatim traces seem to be far more sensitive than gist traces to sources of proactive and concurrent interference (Brainerd & Reyna, 1993). On the other hand, when memory tests are administered after the consolidation interval, findings from many other studies point to the conclusion that verbatim traces become inaccessible more rapidly than gist traces (e.g., Brainerd & Reyna, 1996; Koriat, Levy-Sadot, Edry, & de Marcas, 2003). Here, it appears that verbatim traces are more sensitive than gist traces to sources of retroactive interference (Reyna, 1995).

In sum, the tendency for recognition or recall to be based upon verbatim traces is time dependent as well as cue dependent: no matter how strongly the available retrieval cues favor verbatim access, there will be a progressive shift towards gist access as time passes. If so, as Reyna and Kiernan (1994) observed, this has an intriguing implication for the statistical independence and experimental dissociation patterns that were featured in the preceding section. If verbatim traces become rapidly inaccessible as time passes, memory tests that formerly provoked verbatim retrieval (i.e., tests on which the cues strongly reinstate the surface form of experience) will eventually come to provoke gist access, just like tests that provoke gist access in the first place (i.e., ones on which the cues preserve the meaning but not the surface form of experience). This, in turn, means that the patterns of statistical independence and experimental dissociation that predominate when the two types of tests are administered shortly after encoding will eventually be supplanted by patterns of dependency and association. These changes have been repeatedly detected in studies in which memory tests have been administered several days after encoding (Brainerd & Reyna, 1996; Brainerd et al., 1995b; Lim, 1993; Marx & Henderson, 1996; Reyna & Kiernan, 1994, 1995).

Retrieval phenomenology

In 1913, E.K. Strong reported the earliest systematic data on a topic that has become pivotal in contemporary memory research—the study of spontaneous mental experiences that accompany acts of remembering. Strong's experiments involved a familiar type of task on which subjects responded to an old-new recognition test after studying a word list. The interesting wrinkle was that Strong asked his subjects to introspect upon the mental experiences that accompanied the words that they recognized as old. He noticed that his subjects' introspections were dominated by two contrasting phenomenologies, one that was vague and generic and another that was realistic and specific: "Recognitions are made: (1) When the word was recognized by itself without any other associations of any sort coming to mind. For example, 'tarry [a list word]—I know it was there' (A) [a subject in the experiment], or 'period [another list word]—I know it was in the list' (D) [another subject]... Many words with such introspections were wrong, they had not been seen in the exposure lists... (2) Recognitions were made when at sight of the word in the second [test] list an association with it came to consciousness. The association was then recognized as having been met with before—very often directly connected with the exposure list—and then the word was identified as having been in the exposure list. Generally speaking, as soon as the association came to mind the subject was sure of the recognition" (p. 368).

Nowadays, Strong's (1913) first retrieval phenomenology is called *familiarity*, and his second is called *recollection*. The scientific study of retrieval phenomenologies languished for nearly three-quarters of a century until Tulving (1985) developed the remember/know procedure, in which subjects are given detailed instructions as to how to detect and report mental experiences as they make recognition judgments. Other procedures for measuring the incidence of recollection and familiarity have subsequently been devised, such as process dissociation (Jacoby, 1991), conjoint recognition (Brainerd, Reyna, & Mojardin, 1999), and conjoint recall (Brainerd, Payne, Wright, & Reyna, 2003). Evidence generated by all of them supports the conclusion that these are qualitatively distinct mental experiences, rather than opposite poles of a single phenomenological strength dimension, and that the phenomenologies are driven by different memory mechanisms. A particularly interesting feature of the two phenomenologies, it turns out, is their respective roles in true versus false memory.

FTT posits that the retrieval of verbatim traces induces realistic recollective phenomenology, while the retrieval of gist traces *usually* induces vague familiarity phenomenology. Because verbatim retrieval is assumed to predominate when true events are presented as retrieval cues, recollective phenomenology ought to be the rule with correct recognition and recall, which is a common finding in the literature (Donaldson, 1996). Because gist retrieval is assumed to predominate when false events that preserve the meaning of experience are presented as retrieval cues, familiarity phenomenology ought to be the rule with false recognition and recall, which is also a common finding (Conway, Collins, Gathercole, & Anderson, 1996; Heaps & Nash, 2001). Further, because verbatim retention is poorer than gist retention, over time there should be a shift away from recollective phenomenology towards familiarity phenomenology when true events are presented as retrieval cues, which is yet another standard finding (Gardiner & Java, 1991). Given that there is a reliable, albeit imperfect, correlation between retrieval phenomenology and the truth-falsity of remembered events, at least when events are recent, a potentially important, but largely unstudied, aspect of memory development is that infants and children may learn to rely on these phenomenologies as one method of distinguishing true events from false ones in everyday life (Brainerd, Reyna, Harnishfeger, & Howe, 1993; Reyna & Brainerd, 1998). In everyday life, infants and children have vast funds of cumulative experience with the fact that remembering true events often provokes vivid mental reinstatement of their occurrence, whereas remembering false events does not. A sensible generalization is that one ought to be hesitant about treating remembered events that are not accompanied by realistic phenomenology as true events.

This rule of thumb appears to be acquired during the course of memory development because research with adults has revealed that the rule is ubiquitous by young adulthood (e.g., Dodson & Schacter, 2002; Koriat, Goldsmith, & Pansky, 2000; Schacter, Israel, & Racine, 1999). Of course, the rule has its limitations, the most obvious one being temporal. Because the rule turns upon being able to access verbatim traces, it becomes progressively less practicable as time passes. A more subtle limitation is concerned with the phenomenologies that accompany the remembrance of false events—in particular, an illusory vivid phenomenology that is known as *phantom recollection* (Brainerd, Payne et al., 2003). Although, generally speaking, such remembrance provokes vague familiarity, FTT posits some special conditions in which it can provoke illusory vivid phenomenology that emulates the true recollective phenomenology that is induced by verbatim retrieval (as when the prior "occurrence" of a *false* event echoes in the mind's ear or flashes in the mind's eye). These phantom recollective experiences are predicted to occur at high levels when: (a) certain familiar meanings have been repeatedly instantiated by target experiences and (b) false events are especially good examples of those meanings (Reyna, 2000). Available data are consistent with this prediction (for a review, see Brainerd & Reyna, in press). The key feature of phantom recollection, for present purposes, is that it obviously confounds the strategy of treating remembered events as true if they provoke realistic mental phenomenology.

Developmental variability in verbatim and gist memory

It goes without saying that storage, consolidation, retrieval, and retention processes undergo massive change during infancy (e.g., Bauer, Burch, & Kleinknecht, 2002; Bauer, Wiebe, Carver, Waters, & Nelson, 2003; Howe, 2000), childhood (e.g., Howe, Courage, Vernescu, & Hunt, 2000; Ornstein, Ceci, & Loftus, 1998; Ornstein & Haden, 2001), and beyond. However, FTT's developmental variability principle has a narrower focus, being specifically concerned with age variability in what is known about memory for the surface form of experience versus memory for the meanings and patterns that are instantiated by experience. We consider these two topics separately.

Verbatim memory development

The developmental story for verbatim memory is straightforward. For the types of material that have usually provided the target events in memory development research (e.g., pictures, sentences, and word lists), memory for the exact surface form of such material exhibits marked improvements between infancy and young adulthood. Studies of infant memory require special procedures (e.g., Bauer et al., 2002), and those procedures may not produce findings that are comparable to findings that are generated by more traditional laboratory techniques for studying memory. However, the preschool-to-young-adult age range has been extensively studied using the same, traditional techniques, and the general pattern is that initial improvements in verbatim memory, during the preschool and early elementary-school years, are more pronounced than subsequent improvements (see Reyna, 1996). Evidence of age improvements in verbatim memory can be found in many studies in which children's task was to use memory for the surface form of experience to discriminate between actual targets and new items with very similar visual or auditory features, such as sentences that differ in a single word (e.g., Kiernan, 1993; Paris & Mahoney, 1974;

Weismer, 1985). As a general matter, the target materials in such studies have been highly meaningful, which means that the age trends in verbatim memory might not be pure developmental effects but, rather, might also depend upon the fact that children were processing meaning. However, pure developmental effects for verbatim memory have been obtained in studies in which children's task was to use memory for surface form to discriminate previously presented nonsense targets from new items with similar visual or auditory features, such as nonsense shapes (e.g., Posner & Keele, 1968) or nonsense words. Most evidence of pure developmental effects in verbatim memory comes from studies of nonsense words. To illustrate, we administered a long list containing nonsense words (e.g., CEXIB, ZUTEG) to 7- and 11-year-olds, with the children being asked to listen to each word and decide whether they had heard it earlier in the list. Some of the nonsense targets appeared twice on the list, early and late, but at other times, nonsense targets were replaced by *rhyming* distractors (e.g., LEXIB replaced CEXIB, KUTEG replaced ZUTEG). Between 7 and 11, there was a significant increase in children's ability to discriminate presented nonsense words from these very similar nonsense distractors (Brainerd, Stein, & Reyna, 1998).

Larger developmental trends in verbatim memory have been obtained in other studies that have used broader age ranges. In one such study (Brainerd & Reyna, 2001), 5- and 11-year-olds listened to the same list of nonsense words as in the just-mentioned study, with some words being read once and others being read thrice. Following a short buffer activity, the children responded to a test list that consisted of nonsense targets, distractors that rhymed with nonsense targets, and unrelated nonsense distractors. The key findings are displayed in Fig. 3, which shows the proportion of correct acceptances of targets corrected for the proportion of false

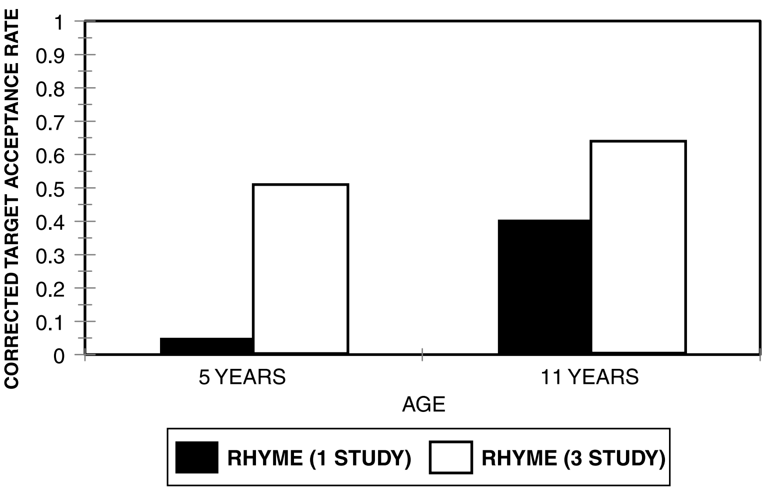

Fig. 3. Children's rate of recognition of nonsense targets (e.g., CEXIB, ZUTEG), corrected for the their rate of erroneous recognition of distractors that rhyme with nonsense targets (e.g., LEXIB, KUTEG).

acceptances of rhyming distractors. Two major results are apparent. First, the black bar on the left of Fig. 3 shows that the 5-year-olds were almost completely unable to discriminate nonsense rhymes from nonsense targets that had been presented only once (which should produce weaker verbatim traces), but this ability had improved dramatically by early adolescence. Second, 5-year-olds exhibited good discrimination of nonsense rhymes from nonsense targets that had been presented thrice (which should produce stronger verbatim traces), but this ability had also improved considerably by early adolescence.

Gist memory development

Not surprisingly, developmental improvements in gist memory are more complex than those for verbatim memory. On the one hand, it is clear that even infants or young preschoolers are quite adept at storing certain types of gist, while it is equally clear that the storage of other types requires lengthy periods of development. Early gist memory abilities are illustrated by the fact that infants are able to extract and preserve various simple meanings from their experience, with well-studied cases in point being perceptual concepts, numerical concepts, and cause-effect concepts (for reviews, see Haith & Benson, 1998; Mandler, 1998). A further example is the fact that preschoolers readily classify pictures of everyday objects according to their appropriate taxonomic categories (e.g., pictures of shirts, pants, and coats as clothing, pictures of hamburgers, hot dogs, and pizza as food; see Bjorklund, 2004). Nevertheless, it is well established that certain forms of gist memory lag far behind in young children. Important forms of such delayed gist memory are connecting the gist, unfamiliar exemplars of familiar meanings, and unfamiliar meanings.

Connecting the gist. Connecting the gist refers to children's ability to extract meaning *relations* that hold between objects and events whose *individual* meanings are well understood (Brainerd, Reyna, & Forrest, 2002). Historically, as it happens, this ability has been the focus of some very influential lines of research involving memory tasks known as discrimination shifts (e.g., Kuenne, 1946). In the basic paradigm, children memorize the classifications of objects that vary along some simple perceptual dimensions (e.g., color, shape, and number). These classifications are determined by a rule that involves one of the values of one of the dimensions (e.g., the red objects are all winners, the blue objects are all losers). The question is, As they memorize the classifications of individual objects, do young children connect this simple gist across the objects? They appear to have difficulty doing so.

Once memorization is complete, the procedure for diagnosing children's tendency to connect the gist consists of changing the object classifications in either of two basic ways (which is where the "shift" in discrimination shift comes from) and then determining the relative difficulty of the two changes through further memorization trials. One type of change preserves the gist of original learning, whereas the other does not. The gist-preserving change, which is called a reversal shift, is to introduce a new memorization rule that involves the same perceptual *concept* as before (e.g., red objects are now losers, blue objects are now winners). The gist-violating change, which is called a nonreversal shift, is to introduce a new memorization rule that involves some other perceptual concept (e.g., square objects are now winners, circular objects are now losers). If children connect the color gist across the target objects during initial memorization, the reversal shift will be easier because it involves this same concept. This pattern predominates in adolescents and adults (for a review, see Esposito, 1975). But if children fail to connect the color gist and simply memorize the winner–loser classifications of the individual objects, the nonreversal shift will be easier because it preserves the initial winner–loser classifications of some of the objects (while the reversal-shift changes all of them). In a classical literature review, Kendler and Kendler (1962) concluded that across many different types of object-memorization tasks, young children usually find nonreversal shifts to be easier to than reversal shifts. That conclusion was based on several studies in which the relative ease of reversal versus nonreversal shifts had been compared for age levels that ranged from preschoolers to college students. When Kendler and Kendler plotted the percentages of subjects at the various age levels who found nonreversal shifts to be easier than reversal shifts, the resulting values exceeded 50% in preschoolers and younger elementary schoolers. Later, Tighe and Tighe developed a method of analyzing young children's performance during shift tasks that provided direct evidence that they had not connected the gist during the initial memorization phase (Kulig & Tighe, 1976; Tighe, Tighe, & Schechter, 1975).

Other standard examples of children's connecting-the-gist limitations can be found in the extensive literature on the development of free recall of categorized word lists. In the standard design, children of various ages study and then recall a list of words or pictures, with the individual items on the list belonging to familiar taxonomic categories (e.g., some items are clothing, some items are foods, and some items are animals). When adolescents or adults study such lists, they store higher-order meaning relations between list items and display knowledge of the fact that the lists consist of a few semantic groups. Typical ways in which adults show that they have connected the gist across same-category exemplars are by exhibiting high levels of category clustering during free recall and high levels of proactive interference when they study new lists that involve the same categories (see Bjorklund & Muir, 1988). However, up until fairly advanced ages, children fail to exhibit these indicators of connecting the gist across same-category exemplars. For instance, there is an immense developmental literature on category clustering in free recall, the denouement of which is that most children do not spontaneously exhibit this type of clustering before adolescence (e.g., Bjorklund & Jacobs, 1985). With respect to the other diagnostic sign, proactive interference, the developmental literature is more sparse, but the few extant studies converge with clustering studies in showing that spontaneous category-based proactive interference is difficult to detect before early adolescence (e.g., Bjorklund, 1978).

Unfamiliar exemplars and unfamiliar meanings. The bottom line of the preceding section is that even when target materials are composed of familiar items whose individual meanings are well understood by young children, there are major improvements between early childhood and adolescence in the tendency to establish connected semantic relations across individual exemplars of a meaning. Naturally, there is an even simpler gist memory ability that undergoes substantial developmental improvement during the same age range. Here, we refer to the extraction of meanings from individual target events, rather than the extraction of meaning relations, when the targets are items that are unfamiliar to young children, and hence, children are unaware that the items are exemplars of certain meanings, even otherwise very familiar meanings. As an example, the average preschooler would know that RED is a color but not that CRIMSON is the same color, that COUCH is an article of furniture but not that CHESTERFIELD is the same article of furniture, and that PANTS is an article of clothing but not that BRITCHES is the same article of clothing.

In addition to unfamiliar events that are exemplars of otherwise familiar meanings, there are also meanings that are unfamiliar to children, meanings that take time to acquire. The literature on schematic memory, which has been extensively studied in adults (for a review, see Alba & Hasher, 1983), is a rich source of illustrations. The most commonly studied forms of schematic memory involve event schemas, spatial schemas, and social schemas (such as ethnic and gender concepts). The classical finding is that adults exhibit especially good recall of events that fit within some familiar schema, and when they encounter such events, they store and retain the fact that a particular schematic situation was encountered. Turning to children, there are some schematic situations that are quite familiar even to young children (e.g., "getting ready to go to school in the morning," "playing at the park."), and for these particular situations, they display some of the same schematic-memory effects that are found in adults (e.g., Liben & Posnansky, 1977). However, there are numerous schematic situations that are familiar to adults but that fall outside the scope of young children's everyday experience, such as the office schemas that have been frequently studied in college students (Brewer & Treyens, 1981; Lampinen, Copeland, & Neuschatz, 2001). Other examples are provided by schematic situations that involve mature social concepts, such as ethnicity and emotional relationships. In these cases, young children will not extract salient event meanings from target items because those meanings are not yet understood, and many years must pass before they are.

Predictive control of memory development

It is not widely appreciated that the litmus test for a good scientific theory is its ability to forecast new experimental results, rather than to account for existing ones. The reason for prediction's elevated status is that the history of science has taught us that it is always possible to devise multiple competing explanations of any set of findings. Psychology is rife will illustrations: How many explanations of the serial position curve in free recall or the 7 ± 2 capacity of short-term memory or the frequency effect in recall have you encountered? This circumstance is not unique to psychology or other relatively immature sciences. It is endemic to all sciences. In physics, for instance, examples are just as thick on the ground. At last count, quantum mechanics alone had spawned dozens of competing explanations of the behavior of atoms and subatomic particles (see Gribben, 1995).

In the face of such proliferation, other criteria are needed that allow working scientists to decide which theories are preferable, when all of them are able to account for target findings (e.g., the serial position curve) tolerably well. Additional criteria have sometimes been recommended that are in the nature of aesthetic preferences. Perhaps the most common example is plausibility, which is the notion that theories that explain target findings with intuitively reasonable (commonsensible, feasible, etc.) assumptions are preferable to those whose assumptions trouble our commonsense. Although plausibility has a good feel, it suffers from historical relativity and many instances of failure. Historical relativity refers to the fact that common-sense is educable; that ideas that vex the intuition of one generation (e.g., that the Earth is round, that heavier and lighter objects fall at the same speed) seem completely prosaic to another. With respect to plausibility's failures, the history of science presents many examples of theories that made highly counterintuitive assumptions that proved to be more successful than their more circumspect competitors. Again, physics is rife with illustrations. Returning to quantum mechanics, counterintuitive theories such as quantum chromodynamics and the so-called Copenhagen interpretation have completely supplanted theories that seem more plausible, a fact that prompted Richard Feynman's familiar quip, "No one understands quantum mechanics."

Another criterion for what constitutes a good theory—one that is more widely accepted because it is objective and able to produce public agreement among scientists—is *predictive power*: when different theories explain the extant facts equally well, the most successful one is the one that does the best job of predicting new and interesting findings. That predicted effects must be new is self-evident, but what does "interesting" mean? In forecasting results, implausibility is a plus. The most intriguing predictions are the ones that violate our preconceptions, the reason being that a theory that generates such predictions is pointing us in novel directions that would otherwise have seemed silly to pursue. We hasten to add that predictive iconoclasm does not automatically count in a theory's favor. The experimental confirmation hurdle remains. That is, although counterintuitive predictions show that a theory *may* be leading us along new paths, the paths are dead ends if data confute the predictions. So, the prediction criterion for a successful theory is that it forecasts effects that are verified, with verifications of counterintuitive predictions scoring several bonus points.

We have made these prefacing remarks about prediction by way of motivating a feature of applications of FTT to memory development that runs throughout the sections that follow—namely, that those applications are strongly hypothesis-driven. By hypothesis-driven, we mean research that generates and evaluates specific developmental predictions that follow transparently from one or the other of the theory's principles, with the predictions often being implemented in mathematical models, rather than developmental studies that involve bottom-up, exploratory data gathering. Further, special emphasis has been placed upon predictions that most investigators would regard as surprising. In the sections that follow, we review findings in four areas that were originally motivated by predictions of this ilk: (a) development of retrieval phenomenology; (b) "reversed" developmental trends in false memory;

(c) development of mere testing effects; and (d) development of false persistence in memory. It will been seen that several predictions are concerned with false rather than true memory. This is a consequence of the fact that, as a relatively new research topic, the study of children's false memories provides especially fertile ground for prediction.

Development of retrieval phenomenology

We mentioned Strong's (1913) early work on the mental experiences that accompany the retrieval of specific memories, along with the fact that this has become a productive area of research with adults. In sharp contrast, there is virtually no data base on how retrieval phenomenologies first emerge in children. This is a gap worth filling, given that adult research shows that these phenomenologies have some key functions in remembering, especially when it comes to subjective discrimination of the truth-falsity levels of candidate memories. The explanation for the lack of developmental data lies in the lack of child-appropriate methodologies. Naturally, Strong's method of free introspection, which was so popular in his era, is beyond young children's capabilities, even if one could ignore the well-known reliability criticisms of the procedure. In modern times, retrieval phenomenologies have been studied in adults with more reliable judgment-based methods, in which subjects receive instructions that: (a) spell out the specific phenomenologies that they can expect to experience on memory tests and (b) ask them to rate the intensity of individual phenomenological experiences (Johnson, Foley, Suengas, & Raye, 1988) or to select between different possible phenomenologies (Tulving, 1985). The complex instructions and judgments that are involved in these newer paradigms means that they too are beyond children's capabilities (Ghetti, Qin, & Goodman, 2002).

To solve this problem, we devised a simple behavioral methodology, conjoint-recognition, that does not require children to introspect upon mental experiences that accompany remembering and, instead, extracts estimates of those experiences *directly* from their responses to memory tests (Brainerd et al., 1999). The methodology measures the two phenomenologies mentioned earlier, recollection and familiarity, along with others that will not be discussed here. For the most part, the procedure is a simple recognition-memory task, a procedure at which even two- and three-year-olds are quite adept (Merriman, Azmita, & Perlmutter, 1988). The new wrinkle that delivers estimates of retrieval phenomenologies is that children respond to recognition tests under three sets of simple instructions, with performance differentials between the instructional conditions being used to diagnose levels of specific retrieval phenomenologies via a mathematical model that simultaneously corrects the data for response bias. Further description of this methodology would be tedious and space-consuming, so we proceed instead to the more interesting topics of predictions and findings.

There are two broad views of developmental trends in retrieval phenomenology (Brainerd, Holliday, & Reyna, 2004). One grows out the hoary notion that children's conscious mental experiences are extraordinarily rich in realistic, compelling imagery that leads to robust fantasy lives that adults do not share. This element of children's conscious mental experiences has traditionally been used to explain why children, particularly young children, have more difficulty discriminating fact from fiction than adolescents or adults. On this view, there are no grounds for predicting developmental increases in the realistic recollective phenomenology that is central to adult remembering, quite the contrary. Because realistic mental experiences are the rule in young children, the measured levels of this phenomenology could be higher at earlier as opposed to later age levels. Such a trend should be especially apparent for false memories, which ought be accompanied by high levels of phantom recollection in young children (Brainerd et al., 2004). The other view can be derived from FTT. Here, remember, first, that true recollective phenomenology is assumed to be a by-product of retrieving verbatim traces and, second, that phantom recollection is assume to a by-product of retrieving especially strong gist traces. Remember too that available data show that retrieval of both types of representations increases between early childhood and young adulthood. Therefore, FTT's predictions are the opposite of those of the first view: true recollection and phantom recollection should both display developmental increases. Note that the second prediction is highly counterintuitive from two perspectives—specifically, from the perspective that vivid fantasy imagery is at its peak during early childhood and from the perspective that the processes that support false memories decline with age because the overall accuracy of memory is improving.

We evaluated these predictions in two recent experiments, using different memory materials, and the results converged on the same conclusions. Both experiments were developmental conjoint-recognition designs. In Experiment 1, children of three age levels (7-, 11-, and 14-year-olds) studied the well-known Deese/Roediger/McDermott (DRM; Deese, 1959; Roediger & McDermott, 1995) word lists. A DRM list is generated by selecting a common word, such as DOCTOR, and then selecting the first 15 associates of the word (NURSE, SICK, HOSPITAL, ILL, PATIENT, CURE, STETHOSCOPE, SURGEON, etc.) from available norms of word association (e.g., Nelson, 1999). The associate words *but not the generating word* are then presented to participants as a study list, followed by a standard yes/no recognition test. On such a test, adults falsely recognize the generating word at very high levels—usually, 70–80% of the time. Experiment 1 was simply a conjoint-recognition variant of this tried-and-true task. The results are shown in Fig. 4, where it can been seen that the theoretical predictions, including the counterintuitive one, were confirmed: the tendency of targets to provoke true recollective phenomenology increased by .32 during this age range, and the tendency for semantically related distractors to provoke phantom recollective phenomenology increased by .25. For true recollection, the increase was more marked between 11 and 14 than between 7 and 11, whereas for phantom recollection, the increase was more marked between 7 and 11 than between 11 and 14. In FTT, this difference in the overall developmental patterns for true versus phantom recollection would be interpreted as showing that *in this particular paradigm*, the development of gist retrieval is leading the development of verbatim retrieval.

In Experiment 2, we assessed the replicability of these results using different materials and added a manipulation called gist cuing that FTT predicts should selectively

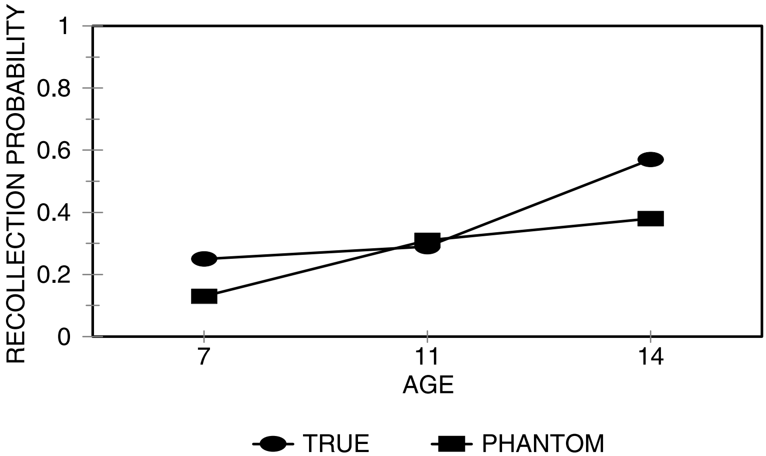

Fig. 4. Developmental trends in true recollective phenomenology and phantom recollective phenomenology (based upon Brainerd et al., 2004).

affect phantom recollection. The children were of two-age levels, 5- and 11-year-olds. First, they studied three categorized word lists, composed of 10 exemplars apiece (e.g., a list of 10 animals, 10 flowers, and 10 women's names). Second, the children responded to a recognition test for these lists under the three types of conjoint-recognition instructions. Third, the children received two more cycles of three categorized lists followed by conjoint recognition tests. Once again, there was a dramatic age improvement in true recollective phenomenology, from an average of .18 in 5-year-olds to an average of .42 in 11-year-olds. In contrast, there was no evidence of age improvements in phantom recollection. However, unlike Experiment 1, the procedures in this experiment produced very low levels of phantom recollection in all conditions, making it difficult to detect age variability.

The overriding message of these experiments is that there is a theoretically predicted vague →vivid shift in retrieval phenomenology, with older children's and adolescents' recognition responses being far more likely to provoke vivid realistic mental experiences than younger children's recognition responses. This pattern has been identified for both true and phantom recollection, though the level of empirical support is better for true recollection.

"Reversed" developmental trends in false memory

Age trends in false memory are a second area in which FTT has generated counterintuitive predictions about memory development. Because false memory will come up again in connection with later topics, it would be well to define our terms at this point. There are two basic errors that people make when they attempt to remember the events of their lives: errors of omission, which are collectively known as forgetting, and errors commission, which are collectively known as false memory. The number of real-life circumstances and laboratory paradigms that generate false memories is enormous. However, most of them have a key feature in common: the false memories that are measured are ones that are consistent with subjects' understanding of the meaning of their experience (i.e., with their gist memories). Here are a few illustrations of this gist-consistency thread of false memories: after studying a DRM list, a subject falsely recalls the unpresented generating word (e.g., DOCTOR); a suspected victim of child sexual abuse falsely reports having been touched on the left breast, rather than correctly reporting having been touched on the right breast; a witness to a convenience store robbery falsely reports that the robber was wearing a jacket, rather than correctly reporting that he was wearing a trench coat; a subject in a language experiment falsely recalls hearing the sentence "The plane is above ground," rather than correctly reporting that he heard the sentences "The plane is above the tree" and "The tree is on the ground." It is easy to see in each instance that although the reported information is false, it is congruent with the gist of experience.

Developmentally, the standard pattern across several paradigms is that younger children are especially vulnerable to false memories. When gist-consistent false information is suggested to them (Remember when we had potatoes at dinner last night?), younger children are more likely than older children or adults to remember it as a true experience on recognition or recall tests (for reviews, see Bruck & Ceci, 1997; Ceci & Bruck, 1995). Similarly, a repeated finding about spontaneous false memories (e.g., remembering drinking a coke rather than a 7-Up at a restaurant) is that they become less common between early childhood and young adulthood (for reviews, see Brainerd & Reyna, 1998a; Reyna, 1996). In the face of such extensive evidence, FTT makes the surprising prediction that there will be instances of false memory that will be more prominent in older children and adults than in younger children. According to FTT's analysis of false memory (Reyna & Brainerd, 1995), the processing of verbatim traces of targets and of gist traces work in opposition when it comes to false memory, with verbatim processing ("7-Up") supporting suppression of false memories ("No, I could not have drunk a Coke, because I clearly remember drinking a 7-Up") and gist processing ("soft drink") supporting acceptance of false memories ("7-Up"). Thus, increases in false memory can occur via two distinct routes, decreases in verbatim processing or increases in gist processing. As both types of processing improve with age, the first route obviously could not be a source of developmental increases in false memory. This leaves gist processing as a potential source of developmental increases, and more specifically, such increases should occur in situations that meet two criteria: the forms of gist processing that support false memories are more likely to occur at older age levels, and it is to hard suppress false memories using verbatim traces of targets (so that the memory-falsifying effects of developmental improvements in gist processing cannot be neutralized by parallel developmental improvements in verbatim processing).

In some recent papers (Brainerd et al., 2002, 2004; Forrest, 2002; Karibian, 2003), it has been noted that both criteria are satisfied by the DRM paradigm. With respect to the first one, earlier we discussed a gist ability that undergoes marked improvement throughout childhood and early adolescence—namely, spontaneous extraction of meaning *relations* that hold among different targets that overlap in meaning, such as exemplars from the same taxonomic category or words that are semantic associates of each other. DRM lists are materials of the latter sort, with the unpresented generating word being the meaning relation that leads to false-memory responses. Spontaneous extraction of this meaning relation will be less likely in younger children than in older children or adults. With respect to the second criterion, although verbatim traces of multiple DRM targets will normally be available for retrieval, this is not of much assistance when it comes to suppressing false memories of generating words. Because *all* DRM targets have related meanings, the fact that the presentation of NURSE, SICK, HOSPITAL, ILL, PATIENT, and CURE can be vividly recollected, whereas the presentation of DOCTOR cannot be, is not trustworthy evidence that DOCTOR is a distractor. Subjects who have extracted the gist of the list will be aware that there are many other medical words on the list that they are unable to recollect vividly, and DOCTOR might be one of them. Thus, developmental improvements in verbatim processing are unlikely to neutralize development improvements in gist processing because verbatim processing is not particularly helpful in suppressing false memories of generating words.

We have conducted several developmental studies of the DRM task in the past few years, using very similar procedures, with adults, adolescents, and children. Sometimes, we studied recall of DRM lists, and other times, we studied recognition of DRM lists. Although different theoretical questions and manipulations were explored in different studies, our concern here lies with the overall developmental picture. Therefore, we pooled data by age level across studies (to enhance stability), and the results are plotted in Fig. 5. The specific data that have been plotted are the mean probabilities of false recall and false recognition of unpresented generating words. The global pattern is straightforward: both measures of false memory increase steadily between early childhood and young adulthood. In addition, recognition tests are more sensitive measures of false memory than recall tests at all age levels. Although the recognition measure exhibits the same qualitative developmental trend as the re-

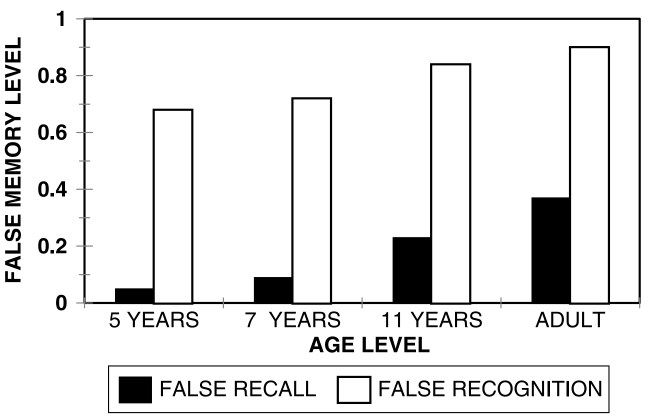

Fig. 5. Developmental increases in false recall and false recognition of generating words for Deese/Roediger/McDermott lists.

call measure, absolute levels of false recognition are high at all age levels—ranging from a low of .65 in 5-year-olds to a high of .91 in young adults.

DRM lists are not the only tasks that produce developmental increases in false memory. Brainerd and Mojardin (1998) used a sentence-recognition procedure that satisfied both of the aforementioned criteria. Children listened to lists of 48 sentences that were composed of trios of related sentences (e.g., The coffee is hotter than the tea. The tea is hotter than the cocoa. The cocoa is hotter than the soup.) Thus, the study materials provided an especially good opportunity for the memory-falsifying effects of developmental increases in the extraction of meaning relations to express themselves because meaning could be connected in two ways: across words within sentences and across sentences within each trio. Recognition tests were administered that contained target sentences (e.g., The coffee is hotter than the tea.), gist-consistent distractor sentences (e.g., The cocoa is cooler than the tea.), and gist-violating distractor sentences (e.g., The coffee is cooler than the cocoa.) These tests provided little opportunity for the suppressive effects of developmental improvements in verbatim retrieval to express themselves because they were delayed until after children listened to the entire study list and performed an attention-consuming buffer activity. These features of the tests should make it difficult for even the oldest subjects to retrieve verbatim traces of these complex targets, leaving gist processing as the default option. The measure of false memory in this study was the probability of falsely recognizing gist-consistent distractors, corrected for the probability of response bias (the probability of false recognition of gist-violating distractors). The pooled results for three age levels—6-, 9-, and 11-year-olds—are displayed in Fig. 6. It can be seen that this type of false memory exhibited the same pattern of major age increases as DRM lists do, with the level of false memory for these more semantically complex materials only being slightly above floor in the youngest children. Therefore, although the prediction of developmental increases in false memory

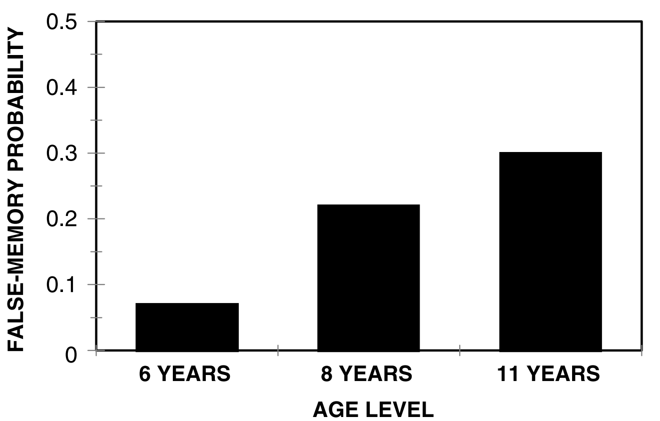

Fig. 6. Developmental increases in false recognition of gist-consistent sentences (based upon Brainerd and Mojardin, 1998).

is counterintuitive, it appears that the prediction can be verified with tasks that satisfy the preceding criteria.

Development of mere memory-testing effects

We now consider some predictions that FTT makes about how repeated testing affects the accuracy of children's memories. During the past decade, the question of how different methods of tapping the contents of children's memories can influence the subsequent accuracy of their memory reports became the focus of intense research. The impetus came from a high-stakes area of children interviewing; investigative interviews of children who are believed to be witnesses to crimes or victims of crimes, especially sexual abuse. During the course of a criminal investigation, child witnesses/victims are typically interviewed numerous times by many people (e.g., fathers, mothers, siblings, other relatives, teachers, classmates, child protective services investigators, police investigators, prosecutors, and defense attorneys). For instance, one researcher estimated that the average child witness is interviewed up to 12 times before trial (Whitcomb, 1992). The fact that this number is so large has fomented concerns, at least in the minds of memory researchers, about whether the interviews themselves may alter children's memories of original events and taint the reports that they eventually provide at trial. The particular issue of whether repeated interviewing elevates children's *false*-memory reports has been the linchpin of research.

Most of this research has focused upon whether the use of leading questions during the course of interviews implants false memories in children (e.g., He had a gun in his hand, didn't he? Remember when he took the money from the cash register? He touched you on the bottom, didn't he?). Two overriding conclusions that have emerged from numerous studies (e.g., Bjorklund et al., 2000) are that: (a) the memory suggestions that are supplied by leading questions do, indeed, falsify the contents of children's memories and (b) children, especially young ones, are more susceptible to false-memory implantation than adults. Consequently, it has become routine in guidelines for child investigative interviewing to include strong warnings against the use of leading questions, except in extreme circumstances (e.g., see Poole & Lamb, 1998). But, what about the effects of "neutral" questions that do not make deliberate suggestions to children (What did he have in his hand? Did he have a gun in his hand? Did he have a gun or something else in his hand?) Might repeated interviewing that is confined to such "mere memory" testing have benign or even beneficial effects?

Early research seemed to show that the answer was "yes" and, therefore, that repeated interviewing might not be problematical as long as child witnesses were not subjected to leading questions. This research consisted of studies in which children received immediate and delayed recall tests about events that they had experienced in quasi-forensic contexts (e.g., Baker-Ward, Hess, & Flannagan, 1990; Dent & Stephenson, 1979). Studies of this sort were reviewed by Poole and White (1995), and on balance, they found that the effects of repeated neutral testing were beneficial in that children's reports of *true* events increased as a result of repeated interviews, but their reports of *false* events did not: "laboratory simulations... offer strong corroborating evidence that multiple testing sessions preserve memories over time... We are aware of no studies in which multiple interviews with nonsuggestive questions were associated with increases in the amount of inaccurate information recalled by children" (p. 34).

However, as we have just seen with developmental trends in false memory, the fact that experimental findings run consistently in one direction does not preclude the possibility that those findings are not telling the whole story. After reviewing the same studies as Poole and White (1995), we concluded that, for two reasons, the findings failed to provide conclusive evidence that repeated neutral testing does not increase children's false-memory reports (Brainerd & Reyna, 1996). First, null results are always difficult to interpret because the logic of statistical inference demands that a null hypothesis (in this instance, "repeated testing does not elevate false memory") must be *rejected* before inferences can be made, the reason being that failures to reject a null hypothesis can be due to host of uncontrolled noise factors. Second, the early studies that were reviewed by Poole and White contained a number of design limitations—particularly, small sample sizes and small numbers of memory responses per subject—that may have reduced statistical power to levels that were inadequate to detect memory-falsifying effects of repeated neutral testing.

There is also a third reason for doubt: contrasting theoretical predictions. FTT's assumptions about memory retrieval lead one to expect that even neutral memory testing can elevate false-memory reports on later tests. Whether it does ought to depend on the type of retrieval, verbatim or gist, that serves as the basis for children's responses on the initial memory test. Remember, here, that verbatim processing supports the acceptance of true events and the rejection of false ones that preserve the meaning of experience, whereas gist processing supports the acceptance of both true and false events. An initial memory test can therefore be viewed as providing children with practice at retrieving the two types of representations, which should enhance their retrieval on later tests. Enhancements in verbatim retrieval will not elevate subsequent false-memory levels, of course, because verbatim processing is not responsible for false memories, and instead, it should elevate subsequent true-memory reports. However, enhancements in gist retrieval will elevate subsequent false-memory reports. Thus, because neutral memory tests provide practice with both verbatim and gist retrieval, the baseline expectation is that such testing should elevate both true- and false-memory rates on subsequent tests.

To test this prediction, we conducted some experiments whose designs had sufficient power to detect even fairly small influences of repeated testing on children's memories. An additional feature of these experiments is that they allowed the effects of repeated testing on true and false memory to be separated. The basic design of the experiments can explained using Fig. 7. There were three steps. First, a large sample of children of two-age levels (5- and 8-year-olds) studied a long list of familiar words, such as those shown on the far left in Fig. 7. Some of the words were designated as control targets and others were designated as experimental targets, with the former being presented as probes on immediate and delayed recognition tests and the latter being replaced by false-memory items (semantically related distractors). There was

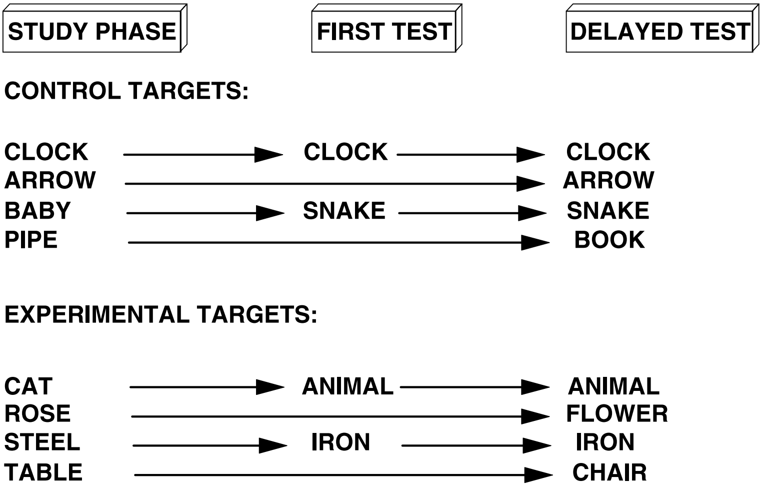

Fig. 7. Design of experiments that measured the tendency of repeated testing to inflate children's true and false memories (based upon Brainerd and Reyna, 1996).

also a word repetition manipulation in the study list, with half of the words being presented once and the other half being presented thrice. This manipulation was included because it is one that is known to enhance the strength of verbatim memory relative to gist memory, as illustrated by the fact that one of its effects is to increase the hit rate while simultaneously reducing the false-alarm rate to semantically related distractors (see Kensinger & Schacter, 1999). Second, after the complete list, including repetitions, had been presented, children responded to an immediate test on which some probes were just-studied targets (CLOCK in the center of Fig. 7), other probes were false-memory items that were semantically related to the targets that they replaced (ANIMAL and IRON in the center of Fig. 7), and the remaining probes were distractors that were unrelated to the targets that they replaced. One week later, the children responded to a delayed test, which was twice as long as the first one. It included all of the items from the first test, these being the repeatedly tested items, and it also included new probes that had not been presented a week earlier, these providing the baseline for measuring the effects of repeated testing. The new probes, like the old ones, were divided into targets (ARROW on the far right of Fig. 7), false-memory items (FLOWER and CHAIR on the far right of Fig. 7), and unrelated distractors. The effects of repeated testing were then determined by comparing performance on the delayed test for two pairs of items: (a) old versus new targets, which measured the effects of repeated testing on true memory, and (b) old versus new false-memory items, which measured the effects of repeated testing on false memory.

The key results are shown in Fig. 8. To begin, as the research reviewed by Poole and White (1995) had suggested, repeated testing improved true memory; the hit rate was higher for old targets than for new ones (cf. panel A). Crucially, however, this effect interacted with age and with repetition. Concerning age, repeated testing

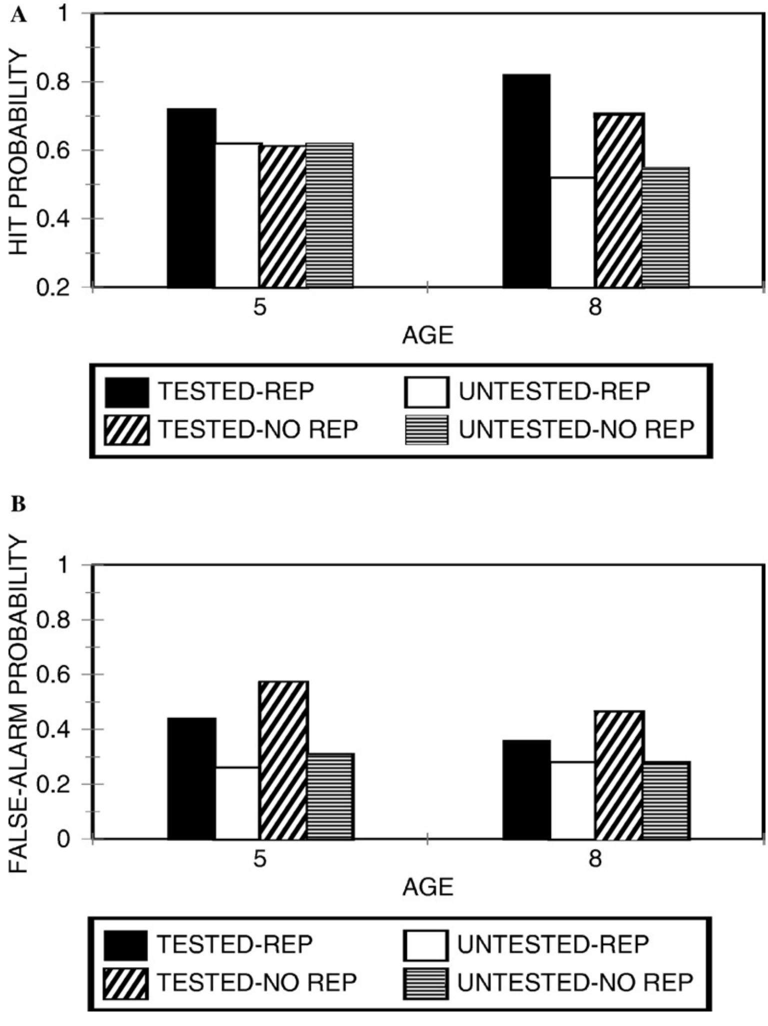

Fig. 8. Effects of prior testing upon children's subsequent memory performance. (Panel A) Target hit rates; (panel B) false-alarm rates for semantically related distractors; TESTED, items that appeared on the immediate test; UNTESTED, items that did not appear on the immediate test; REP, targets that were presented three times on the study list; and NO-REP, targets that were only presented once on the study list.

improved 5-year-olds hit rates only slightly, but it had a dramatic effect in older children. Concerning repetition, recall that this manipulation enhances verbatim retrieval, so that verbatim processing is more likely to occur on the immediate test for thrice-presented targets than for once-presented targets. Therefore, the initial test is more likely to provide verbatim retrieval practice for the former items, leading to the conclusion that the beneficial effects of repeated testing ought to be more marked for thrice-presented targets. This was the observed pattern for both younger and older children.

What about the more surprising predictions concerning how neutral memory tests affect false memory? A glance at panel B in Fig. 8 reveals that those predictions were also confirmed: for all the plotted comparisons of old versus new false-memory items, the false-alarm rate was higher for items that had been previously tested. Indeed, in 5-year-olds, the tendency of repeated testing to elevate false-memory was more pronounced than its tendency to elevate true memory, so that *repeated testing produced net reductions in accuracy*. Repeated testing's effects on false memory, like its effects on true memory, interacted with age and repetition. With respect to age, it can be seen in panel B that although repetition increased false-alarm rates at both age levels, the increase was roughly twice as large in the younger children. The parallel between this outcome and the standard developmental trend in the effects of suggestive questions on false memory (e.g., Ceci & Bruck, 1993) is obvious. With respect to repetition, note that at both age levels, repeated testing elevated false-alarm rates more when distractors' corresponding targets were presented only once. Of course, this is consistent with the aforementioned idea that superior verbatim memory for targets (thrice-presented items) helps children resist the tendency to accept semantically related distractors.

Summing up, these data show, as predicted on theoretical grounds, that neutral memory testing can elevate false-memory responses, though it also elevates true-memory responses. We conducted a follow-up experiment to determine whether the second effect might decline sharply, while the first might increase sharply, as a function of a factor that is commonplace in investigative interviews of children: delaying the first memory test for a few days. According to FTT's differential survival principle, this should produce a shift in the effects of repeated testing because verbatim traces will become rapidly inaccessible over the delay. Specifically, an initial memory test that is administered after, say, a week will necessarily provide a higher ratio of gist retrieval practice to verbatim retrieval practice than one that is administered immediately after target presentation. To explore this possibility, we conducted a point-for-point replication of the first experiment, except for one feature: the first memory test was not administered until one week after the targets were presented and the second was administered a week after that. This manipulation had the expected effects: the increase in true memory as a result of repeated tested was only about half as large as in the first experiment, but the increase in false memory was twice as large. Both effects interacted with age in the same manner as in the first experiment.

These basic patterns in the effects of repeated testing on children of different age levels have been replicated in other experiments in our laboratory, using different to-be-remembered materials and different delays between successive tests. To illustrate, Brainerd and Mojardin (1998) and Mojardin (1997) reported experiments in which children of three age levels (6-, 8-, and 11-year-olds) and adults were exposed to a series of short narratives. To measure repeated-testing effects, memory tests were administered immediately after narrative presentation, one week later, and one month later, following a design like that in Fig. 7 (except that the targets and distractors were sentences rather than words). Overall, these experiments produced four findings that bear on FTT's predictions about repeated testing. First, as in word-list experiments, repeated testing increased the hit rate for true memories but it also increased the false-alarm rate for false memories. Second, the elevation in false-alarm rates was larger (about 50% larger on average) than the elevation in hit rates, so that *the net effect of repeated testing was to reduce the accuracy of children's performance*. Third, the difference between false-alarm elevation and hit elevation increased as time went by, so that the net reduction in accuracy from repeated testing was greater on one-month tests than on one-week tests. Finally, the first three results all interacted with age, such that repeated testing's effect on hit rates increased with age while its effect on false-alarm rates decreased with age. This, in turn, meant that although repeated testing produced net reductions in accuracy at all age levels, the reductions were greater in younger children than in older children or adults.

Development of false persistence: Stability, superiority, and sleeper effects

We come now to a particularly surprising group of predictions that follow from the principle of differential retention rates for verbatim and gist memory. What would common-sense have to say about how specific false-memory responses (e.g., falsely reporting eating a hot dog and drinking a 7-Up at a baseball game) should vary as time passes? A defining feature of such responses is that the events were never experienced, whereas the events in true-memory reports were experienced. Because, strictly speaking, there are no experiences to be remembered, common-sense would say that children's false reports lack the palpable memorial support that true reports enjoy. It follows that false reports ought to be highly unstable as time passes, far more unstable than true reports, because there is no initial memory support that could be preserved over time. An interesting feature of these two common-sense ideas—that false reports should be highly unstable and that the stability of true reports, though far from perfect, ought to greatly exceed that of false reports—is that they figure prominently in the law, where they are known as the consistency rule of credibility. Witnesses testify about their memories of events, and different witnesses often give conflicting accounts of crucial features of events, leaving it up to juries to decide which version is most credible. Studies (e.g., Fisher & Cutler, 1992) have shown that juries most commonly base such decisions on the consistency rule: they assume when witnesses' memory reports are consistent over time, that those reports are more apt to be true than when they vary over time. Information on consistency of reporting over time is often available to juries because witnesses, especially in criminal cases, may be interviewed multiple times about crucial events before they testify at trial, and indeed, attorneys' decisions about whether to have particular witnesses testify may be based up on just how consistent their memory reports have been.

Contrary to common-sense and the consistency rule, FTT predicts that false reports can be highly stable over time and that their stability can sometimes exceed that of true reports. Yet another surprising prediction is that false reports can display *sleeper effects*, waxing over time even as true reports are waning. All of these predictions are by-products of three ideas that figured in the earlier discussion of theoretical principles: that far from lacking tangible memorial support, false reports are rooted in gist processing; that initial true reports are strongly weighted towards verbatim processing; and that as time passes, verbatim traces become rapidly inaccessible while gist traces remain reasonably accessible. Obviously, the prediction that false reports should display good stability follows from the first and last ideas. The prediction that false reports can sometimes be more stable than true ones is a consequence of all three ideas: false reports ought to be more stable over time than true reports whenever the former are overwhelmingly due to gist processing (rather than guessing, lying, social compliance, etc.) and the latter are overwhelmingly due to verbatim processing. The prediction of sleeper effects follows from the first and third ideas. Remember, in this connection, that the reporting of false memories can be suppressed by retrieving verbatim traces of the corresponding true memories (e.g., verbatim traces of eating a hamburger and drinking a Coke at the baseball game). Because verbatim traces become rapidly inaccessible, children will be less able to engage in verbatim-based suppression as time passes. Thus, under conditions in which children are able to perform such suppression at high levels on initial memory tests, the false reports that are being suppressed are apt to increase in frequency over time. Developmental studies have produced supportive findings for all three predictions.

Persistence of children's false memories

The earliest research that provided data on the first prediction appeared in papers by Brainerd, Reyna, and Brandse (1995a) and Poole (1995). The Brainerd et al. experiments implemented an obvious method of measuring the persistence of false-memory responses: if such responses were highly unstable over time, the fact that some event X is falsely reported (or not reported) on an initial test should neither increase nor decrease the chances that it will be reported (or not reported) on a later test. This prediction is easily tested by computing correlations between individual false-memory responses on initial and delayed memory tests, with zero correlations or very small ones demonstrating instability. The design of the experiments was like Fig. 7, except that the one-week delayed test consisted of exactly the same items as the immediate test. In other words, children of two-age levels (5- and 8-year-olds) were exposed to a long list of familiar words, and then they responded to a test list consisting of presented targets (e.g., CAT, CHAIR), false-memory items that were the category labels of targets (e.g., ANIMAL, FURNITURE), and unrelated distractors. One week later the same test was readministered. The statistics that were used to assess the persistence of false memories over the one-week interval were $P(\mathrm{FD}|\mathrm{FI})$, the *conditional* probability of making a false alarm to a category label on the delayed test given a false alarm to that distractor on the immediate test, and $P(\mathrm{FD})$, the *unconditional* probability of a false alarm to that distractor on the delayed test. Statistically speaking, specific false memories are *not* persistent over time if $P(\mathrm{FD}|\mathrm{FI}) = P(\mathrm{FD})$, but they *are* persistent if $P(\mathrm{FD}|\mathrm{FI}) > P(\mathrm{FD})$. A glance at the white bars in Fig. 9 reveals the latter pattern. For both younger children (panel A) and older children (panel B), subjects were about twice as likely to make a false-memory response on the delayed tests if they had made that particular response on the immediate test than if they had not.

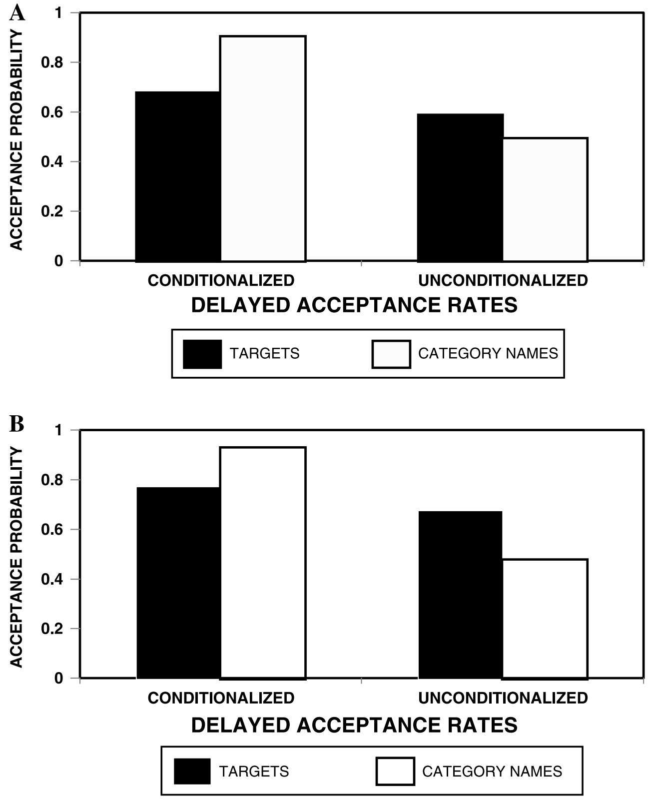

Fig. 9. Stability of true memories (targets) and false memories (category names) over a one-week retention interval. (Panel A) 5-year-olds and (panel B) 8-year-olds.

Poole (1995) observed analogous persistence of false memories. In an earlier study, Poole and White (1991) has exposed 4-, 6-, and 8-year-olds to an ambiguous sequence of events involving a male and a female character that was intended to foment false-memory reports of an assault. The children's memories were tested, using a combination of recall and recognition, immediately after the sequence of events and after a one-week delay. Poole computed two measures of false-memory persistence, a within-session measure and a between-session measure. With respect to the within-session measure, the immediate test included two questions about some of the key information (e.g., the man's appearance). For such repeated queries, Poole computed the conditional probability of a false report to the second question given a false report to the first, and she also computed the unconditional probability of a false report to the second question. She found that the conditional probability (.76) was quite high, and like Brainerd et al. (1995a), she found that it was much higher than the unconditional probability of a false report to the second question. With respect to the between-session measure, Poole computed the conditional probability of a false report to specific questions on the one-week test given a false report on the immediate test, and she also computed the unconditional probability of a false report on the one-week test. Once again, this conditional probability was quite high (.80) and was larger than the unconditional probability of a false report on the one-week test.

The finding of good levels of false-memory persistence has been replicated in later developmental research. To illustrate, Newcombe and Siegal (1997); Bjorklund and Muir (1988); Salmon and Pipe (2000); and Sussman (2001) have all reported studies of children's false-memory reports for everyday experiences over intervals ranging from a week to a year, and there was evidence of persistence across all of these intervals. The aforementioned experiments on repeated testing effects by Brainerd and Reyna (1996) and Brainerd and Mojardin (1998) also produced evidence of false-memory persistence across intervals of one week, two weeks, and one month.

Superior persistence of false memories

The second prediction is that it should be possible to demonstrate that, under certain conditions, children's false memories are more persistent than their true memories. This prediction has been studied in two ways: (a) by comparing intraitem correlations between performance on immediate and delayed tests for false-memory items to the same intraitem correlations for true-memory items and (b) by comparing changes in levels of performance between immediate and delayed tests for false-memory items to changes in performance for true-memory items. In the first type of design, which is exemplified by the Brainerd et al. (1995a) and Poole (1995) experiments, one merely computes the same immediate versus delayed correlations for recognition (or recall) of targets versus recognition (or recall) of unpresented meaning-preserving items and compares the relative magnitudes of the correlations. The preponderance of developmental studies have implemented this design. Many of them were reviewed a few years ago by Brainerd and Poole (1997), who concluded that most studies had shown that children's false memories were *at least as persistent* as their true memories and that, under some conditions, their false memories were more persistent. The Brainerd et al. (1995a) and Poole (1995) studies were again the earliest ones to produce pertinent data. Brainerd et al. reported two key results. First, the immediate versus delayed correlation in 5-year-olds was higher for false-memory responses than for true-memory responses. Second, the same was true for 8-year-olds. (A depiction of these results can be seen in Fig. 9, where the discrepancy between conditional and unconditional probabilities—the measure of persistence—is larger for false-memory responses than for true-memory responses.) Similarly, Poole found that the correlation between immediate and one-week delayed performance was higher for recall of incorrect details than for recall of correct details (.80 versus .71).

The second method measuring the relative stability of true and false memories has been used by Brainerd and Mojardin (1998) and Forrest (2002). In the latter experiment, younger and older children, 6- and 11-year-olds, listened to several of DRM lists. After each list, they were given either a free-recall test or an irrelevant activity (letter shadowing). After all the lists had been completed, the children responded to a recognition test containing target probes (e.g., NURSE), generating words (e.g., DOCTOR), other semantically related distractors (e.g., DISEASE), and unrelated distractors. Only half of the presented DRM lists were tested on this immediate test. One week later, all of the previously tested lists were retested, and the previously untested lists were tested for the first time. Forrest's findings about relative stability of true and false memory are shown in Fig. 10, which also includes a sample of adult data from another experiment in which the subjects were tested under very similar conditions. In this figure, the relative stability of true and false memories can be determined by simply comparing the heights of the immediate and delayed test bars for targets and the two types of false-memory items. For younger children (panel A), it can be seen that performance declined between the immediate and delayed tests for both targets and generating words and that the amounts of these declines were roughly equivalent Thus, true and false memory were equally stable for this comparison. However, there was no decline in performance for the other false-memory measure, so that this type of false memory showed the predicted pattern of greater stability than true memory. For older children (panel B), the same comparisons reveal that both types of false memories were more stable than true memories: while the true-memory measure declined sharply, there were only very small declines in the false-memory measures. For adults (panel C), the pattern was the same as in older children in that the true-memory measure declined considerably but neither false-memory measure declined.

False-memory sleeper effects

Although the notion that false memories can be more persistent over time than their true counterparts is incongruous, perhaps the most puzzling of the three predictions about persistence is that false memories may wax as time passes. Actually, some of the earliest developmental studies of the relation between forgetting and false memory were concerned with this possibility, particularly the research of Poole and White (1991, 1993) and Reyna and Kiernan (1994), and so have some of the most recent studies (Jones & Pipe, 2002). In most instances, the motivation has been forensic relevance: the normal field conditions in criminal investigation can result in delays of a few days to a few weeks between the occurrence of crimes and the initial gathering of reports from child witnesses/victims, and the normal conditions of criminal prosecution can result in further delays of months or years before children's sworn testimony is taken. Although the accuracy of children's statements will be impaired by the fact that true memory declines over such intervals, the chances of eventually convicting innocent defendants are heightened if the passage of time also causes their statements to be increasingly infected with false-memory responses.

Returning to some theoretical points that were noted earlier, if verbatim and gist traces are forgotten at different rates and if verbatim traces trump gist traces when both are accessible, allowing children to suppress responses to false-but-gist-consistent information, false memory sleeper effects could occur when forgetting intervals produce high levels of verbatim forgetting, coupled with little or no gist forgetting. Whether such sleeper effects, when they occur, are more marked for younger or older

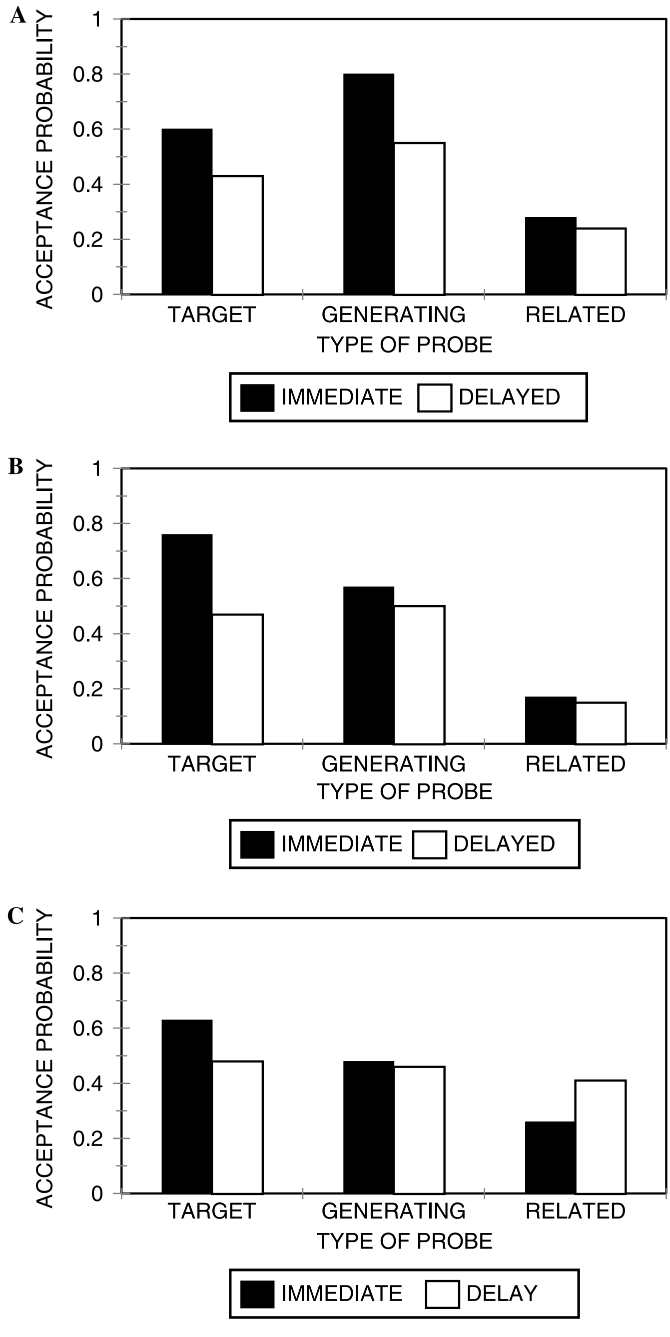

Fig. 10. Stability of true memories (targets) and false memories (generating words and related distractors) over a one-week retention interval in 6-year-olds (panel A), 11-year-olds (panel B), and adults (panel C).

children will depend upon other considerations. Because verbatim forgetting rates will normally be more rapid in younger children than in older ones, this consideration suggests that, other things being equal, sleeper effects will more pronounced in younger children. However, other things are not necessarily equal. Forgetting curves are negatively accelerated, which means that the higher the initial level of any form of memory, the larger its decline over any subsequent forgetting interval. Naturally, initial levels of verbatim memory will be higher in older children than in younger ones, leading to the possibility of more pronounced sleeper effects in older children. Because there are two underlying factors that control the magnitude of sleeper effects (average verbatim forgetting rate, initial level of verbatim memory), with one factor being higher in younger children and the other being higher in older children, the only circumstances in which one can make unambiguous predictions about developmental differences in sleeper effects are ones in which initial levels of verbatim memory are equated in younger and older children, which can be done, for instance, by administering tasks that yield equivalent initial levels of true memory performance. However, given that the designs of most studies have not attempted to equate either of the two factors, it would not be surprising to discover that false-memory sleeper effects have been found to increase with age in some studies and to decrease with age in others, and indeed, this is the pattern in the literature.

Some readers will have noticed that an example of a false-memory sleeper effect has already been presented in Fig. 10: the tendency of the adults in Forrest's (2002) experiment to make false alarms to related distractors *increased* by approximately 50% over the one-week delay. However, the earliest developmental demonstrations of sleeper effects appeared in two papers that have figured in previous discussions, the Poole and White (1991) and Reyna and Kiernan (1994) articles. In both instances, children received immediate and one-week delayed tests for target information, which consisted of a live staged event sequence in Poole and White and a series of narrative statements in Reyna and Kiernan. Also in both instances, the false-memory items were statements that described unpresented information that was consistent with the gist of presented information. Poole and White's subjects were 4-, 6-, and 8-year-olds and adults, while Reyna and Kiernan's subjects were 6- and 9-year-olds.

Poole and White (1993) administered two-year follow-up memory tests to children who participated in their 1991 study and analyzed the results for sleeper effects. The overall pattern of their data can seen in Fig. 11, from which it is apparent that there were sleeper effects and that those effects decreased with age. The lower curve plots the probability that on the initial memory tests, subjects at each age level falsely recalled events that were not part of the staged sequence. It can be seen that these initial levels of false recall were quite low (8% on average) and were roughly the same for all age levels. The upper curve plots the probability that subjects at each age level falsely recalled events that were not part of the staged sequence *after two years.* There is clear evidence of sleeper effects because false recall levels were higher after two years (while, at the same time, true recall levels were dramatically lower). There was also a marked developmental interaction. Four- and 6-year-olds' false recall was three times higher after two years, but 8-year-olds' false recall was only twice as high. Adults did not show a sleeper effect at all.

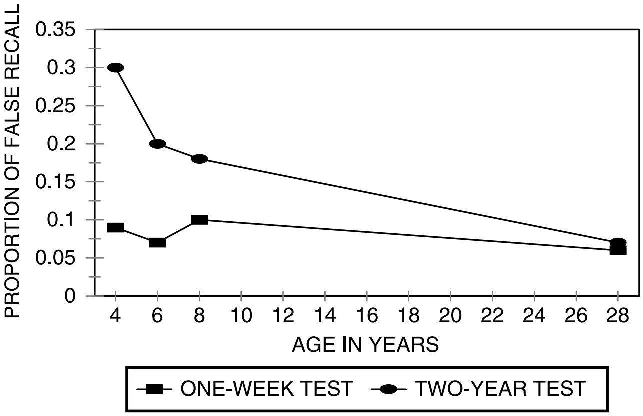

Fig. 11. Growth of false memories over a two-year interval (based upon Poole and White, 1993).

Reyna and Kiernan (1994) also detected sleeper effects, but the developmental trend was the opposite of Poole and White's (1993). The younger and older children responded to immediate and delayed tests on which the false-memory items were unpresented statements that preserved the gist of presented statements (e.g., "The coffee is hotter than the cocoa" preserves the gist of "The coffee is hotter than the tea" and "The tea is hotter than the cocoa"). There were marked sleeper effects at both age levels in that false-alarm rates for these unpresented statements were always higher on the delayed test than on the immediate test. Although sleeper effects were substantial at both age levels, they were more pronounced in older children. Over the one-week interval, the average increase in false-alarm rates was .24 for 6-year-olds and .51 for 9-year-olds. The likely explanation for Reyna and Kiernan's developmental trend can be found from a close inspection of their immediate-test data. Those data suggest that older children's initial verbatim memories were far superior to younger children's because older children's hit rates were much higher and their false-alarm rates to gist-preserving distractors were much lower.

The major source of more recent development findings on sleeper effects is an impressive series of experiments by Pipe, Salmon, and their associates (Gee & Pipe, 1999; Jones & Pipe, 2002; Pipe, Gee, Wilson, & Egerton, 1999; Pipe & Wilson, 1994; Salmon & Pipe, 1997, 2000). Some excellent illustrations may be found in the Pipe et al. article, which incorporates comparisons to data in the Gee and Pipe article and the Pipe and Wilson article. The Pipe et al. paper contains two studies, the first being a two-year follow up of a study that was reported by Pipe and Wilson and the second being a one-year follow up of a study that was reported by Gee and Pipe. In the latter article, children of two-age levels (6- and 9-year-olds) had participated in a live staged event sequence that involved a magician. The children had been interviewed about these events after 10 days and after 10 weeks. Pipe et al. repeated these interviews two years later for the younger children in Pipe and Wilson's study, and they repeated these interviews one year later for both age levels in Gee and Pipe's study. As the latter data provide a more comprehensive picture, we consider only the results for the one-year retest of Gee and Pipe's children.

There had been large developmental differences in true recall in the original Gee and Pipe study, with 6-year-olds recalling only about 7% of the target events on the 10-day test and 9-year-olds recalling 20%. Owing to the floor effect, 6-year-olds' true-memory responses could not decline much thereafter, though 9-year-olds were able to show the usual declines and did. More important, however, there were substantial sleeper effects at both age levels. On cued-recall tests, the absolute increases in false-memory responses over the three testing sessions were quite large: the average number of such responses was .5 at 10 days, 3.6 at 10 weeks, and 5.7 at one year. Another crucial datum that is diagnostic of sleeper effects is the accuracy of information that was recalled for the *first time* after the one-year delay. According to FTT, such information is more apt to be false than information that has been recalled on earlier tests because earlier reports are more likely to have been based upon verbatim processing (and therefore to be true). In line with this prediction, 30% of the information that children remembered for the first time after a year had passed was false on free-recall tests, and 62% was false on cued-recall tests. This finding, that memories that are reported for the first time after a delay are infected with large proportions of false information, has been obtained in other studies in which children responded to free- or cued-recall tests after long delays (Peterson, Moores, & White, 2001; Salmon & Pipe, 2000).

The developmental literature contains still other examples of studies in which children's false-memory responses increased appreciably over forgetting intervals, even as their true-memory responses decreased (e.g., Brainerd & Reyna, 1996; Brainerd et al., 1995b; Kiernan, 1993; Koriat, Goldsmith, Schneider, & Nakash-Dura, 2001; Quas et al., 1999; Reyna & Kiernan, 1995). Taken as a whole, the literature establishes, consistent with theoretical prediction, the existence of false-memory sleeper effects in children, and it also establishes that such effects can be detected with recognition tests, free-recall tests, or cued-recall tests. Although sleeper effects have been observed with all three types of tests, our review of this literature indicates that they have been more marked with recognition tests. Of course, this is consistent with the aforementioned idea that the presentation of false-but-gist-consistent memory probes, as is done on recognition tests, is an especially good way to induce retrieval of the gist traces that support false-memory responses.

Concluding comments

FTT is an interdisciplinary model of cognition that focuses on the interface between memory and higher reasoning and on developmental changes in these domains. In origin, it was designed to improve upon early theories of memory/reasoning relations, such as constructivism and working-memory capacity, that incorrectly predicted strong associations between the accuracy of reasoning and the accuracy of memory for problem information. The theory's recent applications to memory development can be organized under two broad headings.

The first is a nucleus of principles that constitute FTT's substantive perspective on memory development. Those principles are explanatory, rather than metaphorical, inasmuch as each evolved from a delimited set of experimental findings and is intended to account for those findings. For instance, the parallel storage principle is an extrapolation from various results that point to advanced storage of meaning, relative to storage of actual target events that instantiate meaning. Such results include situations in which the presence of memory for targets' meanings can be detected before the presence of memory for the targets' themselves, as well as situations in which the presence of memory for targets' meanings can be detected even though target exposures are so fleeting that subjects are completely unaware that anything was presented. Likewise, the dissociated retrieval principle grows out of a collection of findings that demonstrate high levels of statistical independence and experimental dissociation between tasks that are slanted towards memory for targets' exact surface forms and tasks that are slanted towards memory for subjects' understanding of targets' meanings.

Although FTT's principles are extrapolations from data, they are not mere restatements of findings. Rather, each is a specific theoretical proposal, involving memory processes rather than data, from which multiple distinct findings (e.g., the missing letter effect, the word superiority effect, subliminal semantic priming, and so forth in the case of parallel storage) would fall out as predictions if the proposal happened to be correct. Another important attribute of these principles is that although each is designed to explain a different family of results, there is a superordinate level of organization that unites them. They all implement a single distinction between contrasting and sometimes conflicting modes of representing experience, verbatim and gist traces. Indeed, the principles can be viewed as a series of elaborations on this basic representational theme—as elaborations of how verbatim and gist traces are stored in the first place, of how they are subsequently retrieved, of how they are preserved over time, of the characteristic phenomenologies that result when they are accessed, and of how they vary with development.

The other broad heading under which applications of FTT to memory development can be organized is the use of these principles for predicting new developmental effects. Prediction from theoretical principles and research that evaluates those predictions are essential tools for assessing the principles' validity. Although the principles are grounded in sets of experimental results, history teaches us that there are always other ways to explain any group of effects. But if a principle generates new predictions, especially ones that are counterintuitive from the perspective of alternative theoretical models, and if those predictions can be verified, this provides differential support for the principle, relative to the assumptions of alternative models. In that connection, we discussed a series of effects from four domains in which hypothesis-driven research has figured prominently: development of retrieval phenomenology, "reversed" developmental trends in false memory, development of mere-memory testing effects, and development of false persistence in memory. For the most part, each effect that was discussed followed from just one or another of the principles. The cumulative impact of these effects is to raise confidence in the validity of the parent principles. It should be added that there are other domains of research in which such prediction-oriented research has been concentrated. Examples include studies of output order during recall (Brainerd et al., 1993), studies of inverted-U functions for manipulations of storage and retrieval duration (Brainerd, Reyna, Wright, & Mojardin, 2003), and studies of accuracy enhancements from manipulations of similarity of study and test cues (Brainerd, Reyna et al., 2003). However, space considerations do not permit coverage of these other domains.

We would argue that by following this prediction-oriented path, not only have instructive new findings about memory development been generated that would not otherwise have been topics of research, but there has been a parallel gain in theoretical understanding from these findings owing to their close connection to theoretical principles. The fact that predicted effects necessarily make direct contact with the principles that do the predicting means that the principles are able to evolve rapidly in response to data. Specifically, when data fail to conform to prediction, they provide definite hints about how existing principles need to be modified, and they provide similar hints about new principles that may need to be formulated. Examples of both types of evolution can be found in the literature on FTT. For instance, the parallel storage idea has undergone substantial modification in response to results that conflicted with the original version of this principle (e.g., compare Brainerd & Reyna, 1990, to Reyna & Brainerd, 1995). Similarly, the retrieval phenomenology principle did not even figure in early formulations of FTT. It became necessary to introduce it both to acknowledge the close relationship between retrieval phenomenologies and memory performance that was apparent from experimental findings and to spell out the connections between retrieval phenomenologies and FTT's representational distinctions. All this is by way of acknowledging that FTT, like any active theory, is a work in progress. On the one hand, as we have discussed, there is a stable core of explanatory principles that have achieved a measure of success in forecasting new findings about memory development, some of them quite surprising. On the other hand, the close mapping of data with theoretical principles that is conferred by prediction-oriented research guarantees that core principles will continue to be modified and new principles will be continue to be added as data accumulate.

Acknowledgments

This research was supported by a National Science Foundation Grant (SBR-9730143), by National Institute of Health Grants (NIH31620 and P50AT00008), and by a Schering Plough/Integrated Therapeutics Group grant.

References

Abrams, R. L., & Greenwald, A. G. (2000). Parts outweigh the whole (word) in unconscious analysis of meaning. *Psychological Science, 11*, 118–124.

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

Ankrum, C., Palmer, J. (1989, November). *The perception and memory of objects and their parts*. Paper presented at Psychonomic Society, Atlanta, GA.

Baker-Ward, L., Hess, T. M., & Flannagan, D. A. (1990). The effects of involvement on children's memory for events. *Cognitive Development, 5*, 55–69.
Bauer, P. J., Burch, M. M., & Kleinknecht, E. E. (2002). Developments in early recall memory: Normative trends and individual differences. *Advances in Child Development and Behavior, 30*, 103–152.
Bauer, P. J., Wiebe, S. A., Carver, L. J., Waters, J. M., & Nelson, C. A. (2003). Developments in long-term explicit memory late in the first year of life: Behavioral and electrophysiological indices. *Psychological Science, 14*, 629–635.
Bjorklund, D. F. (1978). Negative transfer in children's recall of categorized materials. *Journal of Experimental Child Psychology, 26*, 299–307.
Bjorklund, D. F. (2004). *Children's thinking: Cognitive development and individual differences* (4th ed.). Pacific Grove, CA: Brooks/Cole.
Bjorklund, D. F., Cassel, W. S., Bjorklund, B. R., Brown, R. D., Park, C. L., Ernst, K., & Owen, F. A. (2000). Social demand characteristics in children's and adults' eyewitness memory and suggestibility: The effect of different interviewers on free recall and recognition. *Applied Cognitive Psychology, 14*, 421–433.
Bjorklund, D. F., & Jacobs, J. W. (1985). Associative and categorical processes in children's memory: The role of automaticity in the development of free recall. *Journal of Experimental Child Psychology, 39*, 599–617.
Bjorklund, D. F., & Muir, J. E. (1988). Children's development of free recall memory: Remembering on their own. *Annals of Child Development, 5*, 79–123.
Brainerd, C. J., & Gordon, L. L. (1994). Development of verbatim and gist memory for numbers. *Developmental Psychology, 30*, 163–177.
Brainerd, C. J., Holliday, R. E., & Reyna, V. F. (2004). Behavioral measurement of remembering phenomenologies: So simple a child can do it. *Child Development 75*, 505–222.
Brainerd, C. J., & Kingma, J. (1985). On the independence of short-term memory and working memory in cognitive development. *Cognitive Psychology, 17*, 210–247.
Brainerd, C. J., & Mojardin, A. H. (1998). Children's false memories for sentences: Long-term persistence and mere testing effects. *Child Development, 69*, 1361–1377.
Brainerd, C. J., Payne, D. G., Wright, R., & Reyna, V. F. (2003). Phantom recall. *Journal of Memory and Language, 48*, 445–467.
Brainerd, C. J., & Poole, D. A. (1997). Long-term survival of children's false memories: A review. *Learning and Individual Differences, 9*, 125–152.
Brainerd, C. J., & Reyna, V. F. (1990). Gist is the grist: Fuzzy-trace theory and the new intuitionism. *Developmental Review, 10*, 3–47.
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. (1995). Autosuggestibility in memory development. *Cognitive Psychology, 28*, 65–101.
Brainerd, C. J., & Reyna, V. F. (1996). Mere memory testing creates false memories in children. *Developmental Psychology, 32*, 467–476.
Brainerd, C. J., & Reyna, V. F. (1998a). Fuzzy-trace theory and children's false memories. *Journal of Experimental Child Psychology, 71*, 81–129.
Brainerd, C. J., & Reyna, V. F. (1998b). When things that never happened are easier to remember than things that did. *Psychological Science, 9*, 484–489.
Brainerd, C. J., & Reyna, V. F. (2001). Fuzzy-trace theory: Dual processes in memory, reasoning, and cognitive neuroscience. *Advances in Child Development and Behavior, 28*, 41–100.
Brainerd, C. J. & Reyna, V. F. (in press). *The science of false memory*. Oxford, England: Oxford University Press.
Brainerd, C. J., Reyna, V. F., & Brandse, E. (1995a). Are children's false memories more persistent than their true memories? *Psychological Science, 6*, 359–364.
Brainerd, C. J., Reyna, V. F., & Forrest, T. J. (2002). Are young children susceptible to the false-memory illusion? Child Development*, 73*, 1363–1377.
Brainerd, C. J., Reyna, V. F., Harnishfeger, K. K., & Howe, M. L. (1993). Is retrievability grouping good for recall? Journal of Experimental Psychology: General*, 122*, 249–268.

Brainerd, C. J., Reyna, V. F., & Kneer, R. (1995b). False-recognition reversal: When similarity is distinctive. *Journal of Memory and Language, 34*, 157–185.

Brainerd, C. J., Reyna, V. F., & Mojardin, A. H. (1999). Conjoint recognition. *Psychological Review, 106*, 160–179.

Brainerd, C. J., Reyna, V. F., Wright, R., & Mojardin, A. H. (2003). Recollection rejection: False-memory editing in children and adults. *Psychological Review, 110*, 762–784.

Brainerd, C. J., Stein, L., & Reyna, V. F. (1998). On the development of conscious and unconscious memory. *Developmental Psychology, 34*, 342–357.

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

Brewer, W. F., & Treyens, J. C. (1981). Role of schemata in memory for places. *Cognitive Psychology, 13*, 207–230.

Bruck, M., & Ceci, S. J. (1997). The nature of applied and basic research on children's suggestibility. In N. Stein, P. A. Ornstein, B. Tversky, & C. J. Brainerd (Eds.), *Memory for everyday and emotional events*. Hillsdale, NJ: Erlbaum.

Budson, A. E., Sullivan, A. L., Daffner, K. R., & Schacter, D. L. (2003). Semantic versus phonological false recognition in aging and Alzheimer's disease. *Brain and Cognition, 51*, 251–261.

Case, R. (1985). *Intellectual development: Birth to adulthood*. New York: Academic Press.

Ceci, S. J., & Bruck, M. (1993). Suggestibility of the child witness: A historical review and synthesis. *Psychological Bulletin, 113*, 403–439.

Ceci, S. J., & Bruck, M. (1995). *Jeopardy in the courtroom*. Washington DC: American Psychological Association.

Conway, M. A., Collins, A. F., Gathercole, S. E., & Anderson, S. J. (1996). Recollections of true and false autobiographical memories. *Journal of Experimental Psychology: General, 125*, 69–95.

Deese, J. (1959). On the prediction of occurrence of certain verbal intrusions in free recall. *Journal of Experimental Psychology, 58*, 17–22.

Dent, H. R., & Stephenson, G. M. (1979). An experimental study of the effectiveness of different techniques of questioning child witnesses. *British Journal of Social and Clinical Psychology, 18*, 41–51.

Dodson, C. S., & Schacter, D. S. (2002). When false recognition meets metacognition: The distinctiveness heuristic. *Journal of Memory and Language, 46*, 782–803.

Donaldson, W. (1996). The role of decision processes in remembering and knowing. *Memory & Cognition, 24*, 523–533.

Draine, S. C., & Greenwald, A. G. (1998). Replicable unconscious semantic priming. *Journal of Experimental Psychology: General, 127*, 286–303.

Eichenbaum, H. (2002). *The cognitive neuroscience of memory*. New York: Oxford University Press.

Esposito, N. J. (1975). Review of discrimination shift learning in young children. *Psychological Bulletin, 82*, 432–455.

Fisher, R. P. & Cutler, B. L. (1992, September). *The relation between consistency and accuracy of eyewitness testimony*. Paper presented at the third european conference on law and psychology, Oxford, England.

Forrest, T. J. (2002). *Memory errors in elementary school children*. Unpublished doctoral dissertation, University of Arizona, Tucson, AZ.

Gardiner, J. M., & Java, R. I. (1991). Forgetting in recognition memory with and without recollective experience. *Memory & Cognition, 18*, 617–623.

Gee, S., & Pipe, M. E. (1999). Helping children to remember: The influence of object cues on children's accounts of a real event. *Developmental Psychology, 31*, 746–758.

Gernsbacher, M. A. (1985). Surface information loss in comprehension. *Cognitive Psychology, 17*, 324–363.

Ghetti, S., Qin, J., & Goodman, G. S. (2002). False memories in children and adults: Age, distinctiveness, and subjective experience. *Developmental Psychology, 38*, 705–718.

Goldmann, R. E., Sullivan, A. L., Droller, D. B. J., Rugg, M. D., Curran, T., Holcomb, P. J., Schacter, D. L., Daffner, K. R., & Budson, A. E. (2003). Late frontal brain potentials distinguish true and false recognition. *Neuroreport, 14*, 1717–1720.

Gribben, J. (1995). *Schrodinger's kittens and the search for reality: Solving the quantum mysteries*. New York: Little Brown.

Hastie, R., & Park, B. (1986). The relationship between memory and judgment depends on whether the judgment task is memory-based or online. *Psychological Review, 93*, 258–268.

Haith, M. M., & Benson, J. B. (1998). Infant cognition. In D. Kuhn & R. S. Siegler (Eds.), *Handbook of child psychology. Cognition, perception, and language* (Vol. 2, pp. 199–254). New York: Wiley.

Heaps, C. M., & Nash, M. (2001). Comparing recollective experience in true and false autobiographical memories. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 27*, 920–930.

Howe, M. L. (2000). *The fate of early memories: Developmental science and the retention of childhood experiences*. Washington, DC: American Psychological Association.

Howe, M. L., Courage, M. L., Vernescu, R., & Hunt, M. (2000). Distinctiveness effects in children's long-term retention. *Developmental Psychology, 36*, 778–792.

Jacoby, L. L. (1991). A process dissociation framework: Separating automatic from intentional uses of memory. *Journal of Memory and Language, 30*, 513–541.

Jones, C. H., & Pipe, M. E. (2002). How quickly do children forget events? A systematic study of children's even reports as a function of delay. *Applied Cognitive Psychology, 16*, 755–768.

Johnson, M. K., Foley, M. A., Suengas, A. G., & Raye, C. L. (1988). Phenomenal characteristics of memories for perceived and imagined autobiographical events. *Journal of Experimental Psychology: General, 117*, 371–376.

Karibian, D. (2003). *Using fuzzy-trace theory to predict memory for word lists and sentences in children with learning disabilities*. Unpublished doctoral dissertation, University of Arizona, Tucson, AZ.

Katz, S., Lautenschlager, G. J., Blackburn, A. B., & Harris, F. H. (1990). Answering reading-comprehension items without passages on the SAT. *Psychological Science, 1*, 122–127.

Kendler, H. H., & Kendler, T. S. (1962). Vertical and horizontal processes in problem-solving. *Psychological Review, 69*, 1–16.

Kensinger, E. A., & Schacter, D. L. (1999). When true memories suppress false memories: Effects of aging. *Cognitive Neuropsychology, 16*, 399–415.

Kiernan, B. J. (1993). *Verbatim memory and gist extraction in children with impaired language skills*. Unpublished doctoral dissertation, University of Arizona, Tucson, AZ.

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

Koriat, A., Goldsmith, M., & Pansky, A. (2000). Toward a psychology of memory accuracy. *Annual Review of Psychology, 51*, 481–538.

Koriat, A., Goldsmith, M., Schneider, W., & Nakash-Dura, M. (2001). The credibility of children's testimony: Can children control the accuracy of their memory reports?. *Journal of Experimental Child Psychology, 79*, 405–437.

Koriat, A., Levy-Sadot, R., Edry, E., & de Marcas, S. (2003). What do we know about what we cannot remember? Accessing the semantic attributes of words that cannot be recalled. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 29*, 1095–1105.

Kuenne, M. R. (1946). Experimental investigation of the relation of language to transposition behavior in young children. *Journal of Experimental Psychology, 36*, 471–490.

Kulig, J. W., & Tighe, T. J. (1976). Subproblem analysis of discrimination-learning: Stimulus choice and response latency. *Bulletin of the Psychonomic Society, 7*, 377–380.

Lampinen, J. M., Copeland, S. M., & Neuschatz, J. S. (2001). Recollections of things schematic: Room schemas revisited. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 27*, 1211–1222.

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

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

Mandler, J. M. (1998). Representation. In D. Kuhn & R. S. Siegler (Eds.), *Handbook of child psychology. Cognition, perception, and language* (Vol. 2, pp. 255–308). New York: Wiley.

Marx, M. H., & Henderson, B. (1996). A fuzzy trace analysis of categorical inferences and instantial associations as a function of retention interval. *Cognitive Development, 11*, 551–569.

Merriman, W. E., Azmita, M., & Perlmutter, M. (1988). Rate of forgetting in early childhood. *International Journal of Behavioral Development, 11*, 467–474.

Mojardin, A. H. (1997). *Age differences in forgetting false memories*. Unpublished master's thesis, University of Arizona, Tucson, AZ.

Moravcski, J. E., & Healy, A. F. (1995). Effect of meaning on letter detection. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 21*, 82–95.

Murphy, G. L., & Shapiro, A. M. (1994). Forgetting of verbatim information in discourse. *Memory & Cognition, 22*, 85–94.

Nelson, D. L. (1999). *University of South Florida word association norms*. University of South Florida.

Newcombe, N. A., & Siegal, M. (1997). Explicitly questioning the nature of suggestibility in preschoolers' memory and retention. *Journal of Experimental Child Psychology, 67*, 185–203.

Ornstein, P. A., Ceci, S. J., & Loftus, E. F. (1998). Adult recollections of childhood abuse: Cognitive and developmental perspectives. *Psychology, Public Policy, and Law, 4*, 1025–1051.

Ornstein, P. A., & Haden, C. A. (2001). Memory development or the development of memory? *Current Directions in Psychological Science, 10*, 202–205.

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

Peterson, C., Moores, L., & White, G. (2001). Recounting the same events again and again: Children's consistency across multiple interviews. *Applied Cognitive Psychology, 15*, 353–371.

Piaget, J. (1967). *Biology and knowledge*. Chicago, IL: University of Chicago Press.

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

Pipe, M. E., Gee, S., Wilson, J. C., & Egerton, J. M. (1999). Children's recall 1 or 2 years after an event. *Developmental Psychology, 35*, 781–789.

Pipe, M. E., & Wilson, J. C. (1994). Clues and secrets: Influences on children's event reports. *Developmental Psychology, 30*, 515–525.

Poole, D. A. (1995). Strolling fuzzy-trace theory through eyewitness testimony (or vice-versa). *Learning and Individual Differences, 7*, 87–93.

Poole, D. A., & Lamb, M. E. (1998). *Investigative interviews of children*. Washington, DC: American Psychological Association.

Poole, D. A., & White, L. T. (1991). Effects of question repetition on the eyewitness testimony of children and adults. *Developmental Psychology, 27*, 975–986.

Poole, D. A., & White, L. T. (1993). Two years later: Effects of question repetition and retention interval on the eyewitness testimony of children and adults. *Developmental Psychology, 29*, 844–853.

Poole, D. A., & White, L. T. (1995). Tell me again and again: Stability and change in the repeated testimonies of children and adults. In M. Zaragoza, J. R. Graham, G. N. N. Hall, R. Hirschman, & Y. S. Ben-Porath (Eds.), *Memory, suggestibility, and eyewitness testimony in children and adults* (pp. 24–43). Thousand Oaks, CA: Sage.

Posner, M. I., & Keele, S. W. (1968). The genesis of abstract ideas. *Journal of Experimental Psychology, 77*, 353–363.

Quas, J. A., Goodman, G. S., Bidrose, S., Pipe, M. E., Craw, S., & Ablin, D. S. (1999). Emotion and memory: Children's long-term remembering, forgetting, and suggestibility. *Journal of Experimental Child Psychology, 72*, 235–270.

Reyna, V. F. (1981). *The animated word: Modification of meaning by context*. Unpublished doctoral dissertation, Rockefeller University, New York, NY.

Reyna, V. F. (1995). 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 in cognition* (pp. 29–61). New York: Academic Press.

Reyna, V. F. (1996). Conceptions of memory development, with implications for reasoning and decision making. *Annals of Child Development, 12*, 87–118.

Reyna, V. F. (2000, November). *Phantom recollection of narratives*. Paper presented at Psychonomic Society, New Orleans, LA, November, 2000.

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. (1991). Fuzzy-trace theory and the acquisition of scientific and mathematical concepts. *Learning and Individual Differences, 3*, 27–60.
Reyna, V. F., & Brainerd, C. J. (1995). Fuzzy-trace theory: An interim synthesis. *Learning and Individual Differences, 7*, 1–75.
Reyna, V. F., & Brainerd, C. J. (1998). Fuzzy-trace theory and false memory: New frontiers. *Journal of Experimental Child Psychology, 71*, 194–209.
Reyna, V. F., Holliday, R., & Marche, T. (2002). Explaining the development of false memories. *Developmental Review, 22*, 436–489.
Reyna, V. F., & Kiernan, B. (1994). The development of gist versus verbatim memory in sentence recognition: Effects of lexical familiarity, semantic content, encoding instructions, and retention interval. *Developmental Psychology, 30*, 178–191.
Reyna, V. F., & Kiernan, B. (1995). Children's memory and interpretation of psychological metaphors. *Metaphor and Symbolic Activity, 10*, 309–331.
Reyna, V. F., & Lloyd, F. (1997). Theories of false memory in children and adults. *Learning and Individual Differences, 9*, 95–123.
Roediger, H. L., & McDermott, K. B. (1995). Creating false memories: Remembering words not presented on lists. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 21*, 803–814.
Salmon, K., & Pipe, M. E. (1997). Props and children's event reports: The impact of a 1-year delay. *Journal of Experimental Child Psychology, 65*, 261–292.
Salmon, K., & Pipe, M. E. (2000). Recalling an event one year later: The impact of props, drawing and a prior interview. *Applied Cognitive Psychology, 14*, 99–120.
Schacter, D. L., Israel, L., & Racine, C. (1999). Suppressing false recognition in younger and older adults: The distinctiveness heuristic. *Journal of Memory and Language, 40*, 1–24.
Seamon, J. G., Luo, C. R., & Gallo, D. A. (1998). Creating false memories of words with or without recognition of list items: Evidence for nonconscious processes. *Psychological Science, 9*, 20–26.
Seamon, J. G., Luo, C. R., Schwartz, M. A., Jones, K. J., Lee, D. M., & Jones, S. J. (2002). Repetition can have similar or different effects on accurate and false recognition. *Journal of Memory and Language, 46*, 323–340.
Strong, E. K. (1913). The effect of time-interval upon recognition memory. *Psychological Review, 20*, 339–371.
Sussman, A. L. (2001). Reality monitoring of performed and imagined interactive events: Developmental and contextual effects. *Journal of Experimental Child Psychology, 79*, 115–138.
Tighe, T. J., Tighe, L. S., & Schechter, J. (1975). Memory for instances and categories in children and adults. *Journal of Experimental Child Psychology, 20*, 22–37.
Tulving, E. (1985). Memory and consciousness. *Canadian Psychologist, 26*, 1–12.
Tulving, E., & Thomson, D. M. (1971). Retrieval processes in recognition memory: Effects of associative context. *Journal of Experimental Psychology, 87*, 116–124.
Tversky, A., & Kahneman, D. (1981). The framing of decisions and the psychology of choice. *Science, 211*, 453–458.
Verfaellie, M., Schacter, D. L., & Cook, S. P. (2002). The effect of retrieval instructions on false recognition: Exploring the nature of the gist memory impairment in amnesia. *Neuropsychologia, 40*, 2360–2368.
Verfaellie, M., Rapcsak, S. A., Keane, M. M., & Alexander, M. P. (2004). Elevated false recognition in patients with frontal lobe damage is neither a general nor a unitary phenomenon. *Neuropsychology, 18*, 94–105.
Wallace, W. P., Stewart, M. T., Shaffer, T. R., & Barry, C. R. (1998). Are false recognitions influenced by prerecognition processing?. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 24*, 284–298.
Weismer, S. E. (1985). Constructive comprehension abilities exhibited by language-disordered children. *Journal of Speech and Hearing Research, 28*, 175–184.
Werner, H. (1948). *Comparative psychology of mental development*. Chicago, IL: Follet.

Whitcomb, D. (1992). *When the child is a victim* (2nd ed.). Washington, DC: National Institute of Justice.
Winer, G. A., Craig, R. K., & Weinbaum, E. (1992). Adults failure on misleading weight-conservation tests: A developmental analysis. *Developmental Psychology, 28*, 109–120.