DISTINCTIVENESS EFFECTS IN CHILDREN'S LONG-TERM RETENTION

ROXANA M. VERNESCU

DISTINCTIVENESS EFFECTS IN CHILDREN'S LONG-TERM RETENTION

by

Roxana M. Vernescu

A thesis submitted to the

School of Graduate Studies

in partial fulfillment of the

requirements for the degree of

Master of Science

Department of Psychology, Science
Memorial University of Newfoundland

03/1999

St. John's Newfoundland

Abstract

The bizarre imagery effect (BIE) was examined in children's long-term retention, as a manipulation of distinctiveness. In a mixed-list, free recall design kindergartners and Grade 2 children were presented with bizarre and common interactions of objects they were to remember. Toys were used to act out these interactions during acquisition. Children's recall was tested following a three-week retention interval. Main effects of grade and gender emerged at acquisition, but no significant item effect. As expected, 2nd graders made fewer errors at acquisition than kindergartners, but surprisingly, boys made fewer errors than girls. A 2 (item: C vs. B) x 2 (grade: K vs. 2) x 2 (gender: boys vs. girls) x 4 (trial) analysis of covariance was used for long-term retention analyses. Main effects of grade, item, and trial emerged, indicating that 2nd graders made fewer errors at retention than kindergartners, all children made fewer errors on the bizarre than common items, and errors declined across the 4 retention test trials. A significant Item x Trial interaction also emerged with bizarre items enjoying fewer errors across trials. Trace-integrity model analyses revealed an age difference. The primary effect of distinctiveness was found to be at reminiscence for kindergarten children, facilitating both storage- and retrieval-based parameters. A similar pattern was seen for 2nd graders, who, in addition, enjoyed a reduction in retrieval-based forgetting. That is, bizarre/distinctive items were much less likely to have a retrieval-based forgetting component over common items, for this age group. It is concluded that not only does the BIE arise in younger ages, but

developmentally, this effect seems to increase both quantitatively (probability of storage-based reminiscence doubled between kindergarten and Grade 2) and qualitatively (Grade 2 children also benefited from distinctiveness in terms of retrieval-based forgetting).

Acknowledgements

I would like to convey my sincerest thanks to a number of you whose support over the past couple of years has been most appreciated. I would like to extend a special thanks to the children at St. Peter's Primary who participated in this study. To the School of Graduate Studies and the Department of Psychology for the financial support provided during the course of this degree. To my supervisor, Dr. Mark Howe, for your guidance and motivational spirit (especially when mine dwindled). Thank you for having more faith in my independence than I gave myself credit for. I can truthfully say I have learned a lot. To my committee members, Drs. Julia O'Sullivan and Michael Rabinowitz, for your invaluable suggestions and contributions towards the finished product. To my internal and external examiners, Drs. Tammy Marche and Mary Courage, for your constructive criticisms and positive feedback. Dr. John Evans and Jamie Drover, I thank you for your most needed help with the statistical analyses. Jamie, thanks for attempting to answer each and every stats question I threw your way.

Last, but definitely not least, I thank you Cary, for your unconditional support and your limitless generosity. You have afforded me endless hours in which to vent all the frustrations associated with graduate school (and life in general). I honestly couldn't have done it without you.

Table of Contents

List of Tables

Page

List of Figures

Page

Distinctiveness Effects in Children's Long-Term Retention

The finding that stimuli that are in some way unusual or unique are remembered better than regular, more commonplace ones has been quite robust (Hunt & Elliot, 1980; Schmidt, 1985). Recently, this pattern has been labeled the distinctiveness effect (Waddill & McDaniel, 1998). Such effects are evident across a variety of manipulations, including semantic distinctiveness of individual words (e.g., Hunt & Mitchell, 1982; Schmidt, 1985), orthographic distinctiveness of verbal items (e.g., Hunt & Elliot, 1980), visual distinctiveness of the components of faces (e.g., Winograd, 1981), odor-evoked memories (e.g., Herz, 1997; Herz & Cupchik, 1995), bizarre imagery effects (e.g., Cox & Wollen, 1981; Einstein, McDaniel, & Lackey, 1989; Fritsch & Larsen, 1990; O'Brien & Wolford, 1982), and memory for atypical actions in script-based stories (e.g., Davidson & Hoe, 1993; Davidson & Jergovic, 1996; Hudson, 1988).

In what follows, I summarize the concept of distinctiveness, including a brief history, current definitions, findings within the area, and theoretical accounts of the effects of distinctiveness on memory. Next, research with bizarre imagery is outlined in terms of its use in the area of distinctiveness. Although sparse, research on the effects of distinctiveness on children's memory is examined and methodological and conceptual limitations are noted. Finally, a theoretical overview of a model of long-term retention

(the trace-integrity framework) is provided, and is subsequently used in this study to isolate the storage and retrieval loci of distinctiveness effects in children's memory.

Brief History

According to Schmidt (1991), the distinctiveness effect hypothesis is an offspring of investigations in the early 1950s and 60s with the von Restorff effect (also see Hunt, 1995; Reed & Richards, 1996). In a detailed review, Schmidt (1991) summarizes three ways in which this effect was studied and, consequently, defined. First, an item may be manipulated such that it is perceptually distinct or unique within a given context (e.g., a distinct colour of ink for the unique item versus the other items in a list). When comparing memory for the unique item it is found to be superior to memory for the same item in a homogenous list (e.g., all items printed in the same colour). The second manipulation involves the introduction of a conceptually different item within a list of conceptually similar items (e.g., inserting an animal name in a list composed entirely of countries). Here also, memory for the unique item is superior to memory for the other items in the list. While in these first two paradigms uniqueness is manipulated in terms of the current context, a third paradigm involves manipulation of uniqueness independent of this context (i.e., incongruent in terms of one's greater knowledge base). Specifically, instead of isolating one item, both distinctive and common items are represented in equal proportions within the entire list to-be-remembered. Distinctive items stray from one's general knowledge whereas common items do not. Although the effects of this

manipulation have not been consistent, the usual finding is better memory for the distinct than for the common items.

As Schmidt (1991) suggests, early research viewed the beneficial effect of distinctiveness on memory as a phenomenon to be explained. In recent years, however, this concept has evolved into an explanation of memory phenomena, illustrated in the memorial benefits of orthographically unusual words (e.g., Hunt & Einstein, 1981; Hunt & Elliot, 1980; Hunt & Mitchell, 1982), bizarre imagery (e.g., McDaniel & Einstein, 1986), word concreteness (e.g., Marschark, & Hunt, 1989), emotionally traumatic events (flashbulb memories)(e.g., McCloskey, Wible, & Cohen, 1988), and many others. Further, Schmidt (1991) brands these explanations as circular in that they use good memory performance as an index of distinctiveness while at the same time invoking distinctiveness to explain good memory performance.

Definition of Distinctiveness

Determining what it is that differentiates distinctive from common items is necessary in order to avoid such circularity. First, distinctiveness has been defined in terms of features of items that are shared in memory. Schmidt (1985, 1996) characterizes distinctiveness as being contingent on the number of features of a particular memory trace that are common to other items in memory. If a memory trace is unique, it shares very few features with other traces and, conversely, common traces share many features with

other traces in memory. Of course item feature overlap is highly subjective and context-bound. That is, whether trace features are deemed unique or common differs from person to person, even within the same individual on two different occasions, and is affected by expertise, culture, and context[1] (Schmidt, 1991).

Second, physiological responses can be used to define distinctiveness, where links are sought between distinctiveness and orienting responses (see Schmidt, 1991, for brief summary). These are indicative of increased attention to stimuli. The two factors important in eliciting this type of response are novelty and significance, the former based on feature overlap between the item in question and preceding stimuli and the latter on feature overlap between the item and other items of significance to the attendee (Schmidt, 1991). Unfortunately, it is not obvious how we can determine the independent contributions of item novelty and item significance to distinctiveness (Schmidt, 1991).

A third way of defining distinctiveness involves ratings of similarity, or conceptual relatedness, to estimate the importance of item features (for brief summary see Schmidt, 1991). As with item feature overlap, the importance or weight given any particular feature is also context-bound. For example, Hunt and Elliot (1980) showed that orthographic distinctiveness ratings are higher for distinctive words being rated in the

[1] Medin, Goldstone, and Gentner (1993) make this argument for similarity. Given the inherent polarity of these concepts, however, its application to distinctiveness is more than clear. Reports abound that similarity changes both in context-specific ways, affected by contextual cues, linguistic context, analogy and relational structure, and with experience. Evidence for the later is seen in developmental changes for similarity judgments as well as the effects of knowledge and expertise (for review, see Medin et al., 1993).

presence of orthographically common words than for the same items being rated in the presence of other orthographically distinct words.

Evidently, it is very difficult to establish a context- or subject-free operational definition of distinctiveness. An alternative would be to view distinctiveness as a hypothetical construct and thus define it in terms of converging experimental manipulations and their effects on memory. Accordingly, Schmidt (1991), after reviewing 16 different operational definitions of distinctiveness, defines four classes of phenomena that can be used to define distinctiveness: the emotional response to the stimulus (i.e., activation of the sympathetic nervous system), the depth and type of processing at encoding (i.e., greater rehearsal, more attention, greater-depth processing), a comparison of stimulus features with other items within a current context (i.e., primary distinctiveness), and a comparison of stimulus features with those in the larger background knowledge of the person (i.e., secondary distinctiveness). Although these are generally treated as independent operationalizations of distinctiveness, it appears that there is, in fact, extensive overlap between them. For instance, it seems that both the emotional response and the type of processing afforded a stimulus are contingent on primary and secondary distinctiveness. Both, the current context and the greater background knowledge, are likely to determine whether an event or situation elicits an orienting response or distinctive processing. Moreover, it seems that although secondary distinctiveness is independent of primary distinctiveness, the converse does not necessarily

hold. Because distinctiveness has to be defined in relation to a context, whether it be a recently encountered one or one accessible via the long-term store, a decision of distinctiveness within a local context will generally be influenced by the greater knowledge base. With this consideration in mind, a more detailed description of primary and secondary distinctiveness is deemed necessary (see following section - Research on Distinctiveness).

Several variables influence the effect of distinctiveness on memory, including experimental design (i.e., within-subject or mixed-list designs include the presentation of both types of items - common and distinctive - to all participants, whereas between-subjects or unmixed-list designs include the presentation of common items to one group of participants and distinct items to another), type of memory measure used (i.e., cued vs. free recall or recognition), and the effect of the distinctive item on memory for the surrounding common items. For example, distinctiveness not only has a beneficial effect on memory for distinctive items themselves, but also has a detrimental effect on memory for the surrounding, common items. Similarly, emotional distinctiveness manipulations (e.g., involving nude pictures, traumatic material, flashbulb memories, humour) have produced somewhat puzzling effects on memory, with beneficial effects on some items and detrimental effects on others. For instance, inserting a nude picture in a series of other pictures not only causes better memory for the nude picture, but also suppresses recall and recognition of the items following the nude picture (see Schmidt, 1991, for review). Last,

distinctive processing of information at encoding can lead to memory traces varying in distinctiveness (Schmidt, 1991). Specifically, these processes, elicited by different tasks or materials, promote greater rehearsal of, more attention paid to, or more in-depth processing of distinctive items. These types of processes appear to facilitate memory in both within-subject (mixed- list) and between-subjects (unmixed-list) designs.

Research on Distinctiveness

In Schmidt's (1991) description of the primary and secondary classes of distinctiveness he proposed that unique items may be conceptually incongruous in that they do not fit active cognitive structures. Primary distinctiveness occurs when there is a mismatch between the features of an item and features in primary memory, as defined by other recently presented items. In secondary distinctiveness the mismatch occurs between the features of the item and features already stored in secondary memory, or those in the larger context of previous experiences. Importantly, these are not independent phenomena. Primary distinctiveness will generally be contingent on secondary distinctiveness, albeit in varying degrees.

Manipulations such as perceptual distinctiveness (e.g., Bruce and Gaines, 1976), category distinctiveness (e.g., Hunt & Mitchell, 1982; Schmidt, 1985), high-priority events (e.g., Christianson & Loftus, 1987), and the consistency effect[2] (e.g., Pezdek, Whetstone,

[2] Items are presented in a scene that can be either consistent or inconsistent with one's general knowledge, or schema, of the scene. The consistency effect refers to the finding that inconsistent items are generally recalled and recognized better than consistent items.

Reynolds, Askari, & Dougherty, 1989) are viewed as exemplary of primary distinctiveness. Although these effects are dependent on greater knowledge base, they are interpreted as primary distinctiveness effects because the relative decisive context is the current one (Schmidt, 1991). These types of manipulations have exhibited a continuum of effects, such that distinctive items are either recalled but not recognized better (e.g., Pra Baldi, de Beni, Cornoldi, Cavedon, 1985; Riefer & Rouder, 1992), both recalled and recognized better with poorer recall (but not recognition) of surrounding information (Schmidt, 1985), or both recalled and recognized better with poorer recall and recognition of background information (see Schmidt, 1991 for review).

Secondary distinctiveness is determined relative to information in long-term storage. Either distinctive items are atypical with respect to a pre-formed conceptual class, or the semantic relation between items is atypical with respect to what is deemed typical for items within that class (Schmidt, 1991). Such manipulations include orthographic distinctiveness (e.g., Hunt & Elliot, 1980; Hunt & Mitchell, 1982), unusual faces (e.g., Winograd, 1981), and bizarre imagery (e.g., Einstein, et al., 1989; McDaniel & Einstein, 1986). Although distinctiveness within this class can be manipulated in either within-subject or between-subjects designs, the beneficial effect on memory seems to be confined to within-subject (i.e., mixed-list) designs. Furthermore, within a mixed-list design the effect of distinctiveness is often to enhance recall and recognition for certain

manipulations (e.g., orthographically distinctive words) but is confined to free recall for other manipulations (e.g., bizarre imagery).

For example, Hunt and his colleagues (e.g., Hunt & Elliot, 1980; Hunt & Mitchell, 1982) examined orthographic distinctiveness effects and found that irregular orthographic patterns (i.e., visually uncommon words) were rated as distinctive and facilitated memory performance, but only when presented in the context of regular orthographic patterns[3] (i.e., visually common words; Hunt & Elliot, 1980). These beneficial effects of orthographic distinctiveness were found on both recognition and recall tasks. Furthermore, the presence of the distinctive items did not have a detrimental effect on the recognition (Hunt & Elliot, 1980) or recall (Hunt & Mitchell, 1982) of orthographically common items within the list. The authors conclude that "if all but one item of a list are similar on some dimension, memory for the different item will be enhanced" (Hunt, 1995, p. 105). To produce this isolation effect it is not sufficient that the isolate and surrounding items are different from each other, but rather, that the surrounding items be similar on some dimension governed by the greater knowledge base, that is different from the isolate. They must provide a context. An event that is distinctive then, is one that violates that prevailing context (Hunt, 1995).

[3] The converse did not hold (i.e., orthographically common words presented against a background of orthographically distinctive words were not recalled/recognized better), underscoring that although distinctiveness develops relative to recently presented information, uniqueness within the greater knowledge base is also important (Hunt & Mitchell, 1982).

As mentioned, research with bizarre imagery is categorized as a manipulation of secondary distinctiveness. Bizarre imagery is, by definition, unique, atypical, and distinctive. By depicting unusual relations common objects become unique with respect to one's background knowledge and memory (i.e., secondary distinctiveness; McDaniel et al., 1995). The observation that bizarre events (when presented concurrent with common events) have a facilitative effect on memory has been termed the bizarre imagery effect (BIE).

Results in this area have been complicated and diverse, with varying manipulations generating varying conclusions. However, identifying the conditions that enhance the BIE may prove useful in understanding the role of distinctiveness in memory (Einstein et al., 1989), and memory development. A typical paradigm involves the presentation of sentences (usually fairly short and simple), either in mixed- or unmixed-list designs. Generally, the ratio of bizarre to common sentences is equal (Mercer, 1996). Subsequently, participants are given some sort of distractor task, followed by a retention test. To a great extent these tasks have concentrated on immediate or short-term recall (see Einstein & McDaniel, 1987; Mercer, 1996), with the few long-term retention studies showing no clear trends. Similar to other research on distinctiveness, results have revealed that the BIE depends on three variables. First, there are differential effects observed with the use of different designs. Specifically, the BIE has been observed consistently under conditions of mixed- as opposed to unmixed-list designs (although see

Einstein et al, 1989; Marshall, Nau, & Chandler, 1980). Second, type of memory task seems to be of consequence. Bizarre images have proved to be more memorable than common images on tests of free recall, in mixed-list designs. Not only do cued recall and recognition tests generally fail to exhibit a BIE (e.g., Riefer & Rouder, 1992), but these sometimes show a commonness effect (e.g., Pra Baldi et al., 1985; Wollen & Cox, 1981b). Third, the type of orienting task is important in these studies. Explicit imagery instructions (i.e., imagining an interaction and providing vividness ratings) tend to produce the BIE while semantic processing instructions (i.e., providing ratings of bizarreness for the interactions) do not (e.g., Einstein & McDaniel, 1987). Although explicit imagery is important, however, the degree of bizarreness of the events, depicted by their interactions or associations, is vital for the BIE.

Current research has concentrated largely on tests of immediate and short-term retention. As such, factors identified as determining or controlling the BIE are, for now, limited to short-term effects. These include individuals' imageability speed (e.g., Hacuk, Walsh, & Kroll, 1976), sentence complexity and list length (e.g., Marshall, Nau, & Chandler, 1980; McDaniel & Einstein, 1989; Robinson-Riegler & McDaniel, 1994), presentation rate (e.g., Kline & Groninger, 1991), level of association (e.g., Lang, 1995), processing level and delay of retention tests (e.g., O'Brien & Wolford, 1982; Weber & Marshall, 1978), interference (e.g., Einstein, McDaniel & Lackey, 1989), and differing nature of bizarre events (i.e., atypical vs. illogical). As such, bizarreness has been

operationalized and defined in a number of different ways. Thus, the varied results obtained from these manipulations are not entirely surprising.

Theoretical Mechanisms of Distinctiveness

Several theoretical mechanisms have been proposed as underlying the facilitative effect of distinctiveness on memory. These mechanisms are difficult to test and none can explain, entirely, the distinctiveness effect. Waddill and McDaniel (1998) divide these accounts into two general classes: differential processing views and representation views. The former hold that distinctive and common items somehow receive differential processing during encoding such that the former receives more (or better) processing than the latter, which aids subsequent recall. The *depth of processing* or *spontaneous elaboration* theorists (e.g., Schmidt, 1991; Waddill & McDaniel, 1998) assume that when encoding to-be-learned material we attempt to interpret and understand the material by incorporating it into our greater knowledge base. If the information is commonplace, it does not require excessive elaboration or activation of knowledge structures. We are familiar with it and understand it quite quickly. However, if the material is unique or unusual, it will require activation of more of our knowledge base until we are able to assimilate it into some sort of generic representation or until we adequately interpret and understand it (for brief review see Waddill & McDaniel, 1998).

A second differential processing assumption is that in addition to differential encoding, distinctive items elicit more attention than common items causing more

elaboration and more time allotted to distinctive than common materials. This is known as the *selective displaced rehearsal* account (as reviewed in Waddill & McDaniel, 1998). Hence, atypical information receives more processing time (more rehearsal) than common information and, therefore, will be better recalled on subsequent tests.

Third, Hirshman (1988) proposed the *expectation-violation* hypothesis as an account of the differential processing view. According to this hypothesis atypical stimuli elicit surprise from participants, who fail to find an association between features of the stimulus and general contextual cues. In turn, this creates an element of surprise (i.e., expectation-violation), which improves memory for those items. Green (1956) initially presented this notion, claiming that it is not isolation which favours recall, but instead, an unexpected change (i.e., a surprise factor; but see Hunt, 1995). Conversely, common items do not elicit this element of surprise, and are recalled more poorly (Hirshman, 1988).

Howe (1997) also suggests that distinctive items or events are remembered better due to a violation of our expectations in a particular situation. In relation to emotional distinctiveness, an event triggering an emotional response will function as distinct within the greater background of personal experiences (Howe, 1997). Such an event will be unexpected and remembered better than the mass of other experiences against which it will be gauged.

In contrast, representational theorists maintain that properties of the to-be-learned information map directly onto properties of memory records (Waddill & McDaniel, 1998).

Accordingly, atypical items create records that are unique and unusual. These representations contrast with the mass of other records in memory, while common items produce records that are common and fairly similar to the other records in memory.

According to one representational assumption, unique items are organized and clustered into their own category at encoding, one which is different from the greater mass of common items (Bruce & Gaines, 1976). This category is then given "privileged status" at retrieval (Schmidt, 1985). In this way distinctiveness influences initial access of the trace (Einstein & McDaniel, 1987) but not its redintegration[4].

A second assumption of the representational view is that distinctiveness is not decided upon at encoding and therefore can not influence how items are represented in memory. Instead, distinctiveness is associated with retrieval (McDaniel, Einstein, DeLosh, May, & Brady, 1995). When trying to remember, one forms a retrieval set of likely possibilities and it is relative to this retrieval set that items are defined (e.g., common or unique). Studies exploring the von Restorff effect (e.g., Hunt, 1995; Reed & Richards, 1996) are delineated as prime examples of the retrieval hypothesis. In one paradigm, an item which is deemed distinct or unusual is presented right at the beginning of a list of to-be-learned items. Hence, participants should not have time to define a context in which

[4] Redintegration is dependent on the strength of the bonds among particular features. That is, the extent to which a trace can be recovered depends on how strongly connected are its primitive features. Although redintegrative alterations (referred to as reminiscence) can occur simply because of external or internal cues present during the testing condition (Howe & O'Sullivan, 1997), distinctiveness is believed to allow for initial discrimination only, between targets and contextual items.

the isolate could be perceived as such. Despite this manipulation, the isolates are still recalled better than the other items on the list, "suggesting that the distinctiveness of the item emerged at retrieval" (McDaniel et al., 1995, p. 433). This being said, however, it is possible that an item's distinctiveness may be determined retroactively (i.e., at the end of the acquisition phase, once a context has been established), a possibility which certainly weakens the retrieval view (i.e., confounds position with process).

Theoretical Mechanisms of the Bizarre Imagery Effect

Similar difficulties are encountered in explaining the BIE. It has not yet been established why bizarre imagery improves memory, or why the effect is stronger under some conditions than others. One hypothesis takes a differential processing approach: more of a subject's attentional resources or capacities are focused on processing bizarre events than common events which, in turn, results in greater recall (Mercer, 1996). What this suggests is that the BIE should occur only in mixed, not unmixed, lists, a finding that is often obtained (Sharpe & Markham, 1992). Here, greater emphasis is placed on the encoding stage of processing which, as Sharpe and Markham (1992) point out, should result in a BIE for both number of sentences and number of items within a sentence (i.e., both the trace itself, and items within that trace). However, bizarre imagery has only been beneficial to sentence access, and not to access of items within sentences (for review see Einstein & McDaniel, 1987).

Although bizarreness is used as a manipulation of distinctiveness, a distinctiveness hypothesis is also put forth as an account for the BIE, and cited as more consistent with the overall patterns reported (Sharpe & Markham, 1992). It is important to note a couple of considerations. First, there is great difficulty in operationalizing distinctiveness. As previously discussed, although an item may be distinctive with respect to an individual's background knowledge, it may become functionally distinctive only in the presence of common items within the learning context (Sharpe & Markham, 1992). Both primary and secondary contexts are important. Being a manipulation of distinctiveness, "bizarreness" is prone to similar limitations, although, perhaps, not to the same degree. What is deemed as bizarre will vary between and even within individuals, is bound by context, culture, and expertise, and is relative to both primary and secondary contexts. Second, it is not sufficient to state that bizarre materials enhance memory performance simply because these are distinctive. Irrespective of the way distinctiveness is manipulated, the locus of its beneficial effects on memory performance need to be uncovered.

At its core it seems that the distinctiveness theory of the BIE is based on the organization-redintegration hypothesis (Sharpe and Markham, 1992). In contrast to the attentional hypothesis, this theory emphasizes the role of retrieval rather than encoding or storage (akin to McDaniel et al.'s, 1995 representational view of distinctiveness). In particular, the retrieval of stored memories consists of three stages: trace contact or access, trace use, and trace decoding. Bizarreness is hypothesized to increase the

discriminability of a memory trace making it more accessible during retrieval (but see earlier point on the possibility of distinctiveness being determined at encoding). Since bizarreness is not assumed to be directly related to trace organization, it should not influence retrievability of trace elements. Riefer and Rouder (1992) for instance, using a multinomial model, concluded that distinctiveness somehow improved the retrievability of traces from memory but not their storage. The authors admit, however, that the effect could also have been due to the differential encoding (and hence storage) of bizarre and common images: "Successful retrieval inherently depends on a sufficient level of encoding"(Riefer & Rouder, 1992, p. 602), and this level might not have been constant across groups or experimental conditions.

It is clear, then, that storage factors cannot be ruled out and may play an important role in the bizarreness effect. The finding that bizarre images are rated as less vivid than common images is quite common (see Riefer & Rouder, 1992, for review). Consequently, any potential storage benefits of bizarre images could have been muted by a less effective image quality. Providing an opportunity of consistent image quality across items (i.e., by having the experimenter provide to-be-remembered images) might curb these differential effects. All participants then, would be presented with the same imaginal cues from which to form their images.

Riefer and LaMay (1998) proposed a two-factor theory with respect to bizarre and common items which is based on the contrast between storage and retrieval[5]. An interaction is proposed between storage and retrieval stages of memory such that these favour common and bizarre materials differentially. That is, common items have a beneficial effect on storage, while bizarre items have a beneficial effect on retrieval. The authors further proposed that this two-factor theory can account for a variety of bizarre imagery effects (i.e., strong, weak, and nonexistent) and a wide range of other findings (Riefer & LaMay, 1998).

Whereas Riefer and Rouder (1992) had previously shown that bizarre stimuli facilitate retrieval, Riefer and LaMay (1998) showed that common stimuli can facilitate storage. In essence, the authors manipulated variables known to produce a bizarreness effect in such a way as to ensure the production of a commonness effect (i.e., used an unmixed- vs. mixed-lists design, and immediate cued recall vs. free recall). Applying the Riefer-Rouder (1992) multinomial model, Riefer and LaMay (1998) determined that the locus of the commonness effect was at storage. The authors concluded that the retrieval benefit of bizarre items is sometimes offset by the storage benefit of common items (Riefer & LaMay, 1998).

---

[5] However, others (e.g., Howe & Brainerd, 1989; Howe & O'Sullivan, 1997) view storage and retrieval as opposite ends of the same continuum, implications of which will be shortly discussed.

Research with Children

The effect of distinctiveness in children's memory performance has been investigated infrequently. Indirectly, it has been shown that children (as well as adults) focus on event features that are "surprising, novel and unique, ones that often violate one's knowledge and expectations" (Howe, 1997, p. 163). Examples in child research come from studies on memory for typical and atypical actions in script-based stories (e.g., Davidson & Jergovic, 1996; Davidson & Hoe, 1993; Hudson, 1988). Davidson and Jergovic (1996) explain that the development of script knowledge requires an understanding of temporal and causal sequences which typically occur within a given situation. Although it is important to examine the development of this ability very early on, it is also important to examine deviations from scripts, which can provide a more profound understanding of how these deviations are remembered (Davidson & Jergovic, 1996). In essence, these deviations are atypical, and thus distinct from the current context (primary distinctiveness). However, children understand these events to be unique because of the background knowledge they bring to the situation (secondary distinctiveness).

Hudson (1988) found that young children's recall and recognition of atypical actions within a given story was better than their recall and recognition of typical actions. This finding was obtained with both immediate and 24 hr retention intervals. Furthermore, the author reports that atypical actions which disrupt the goal of a particular story (e.g., at

the grocery store, they dropped some eggs on the floor) are recalled better than atypical actions that have no relevance to the goal of the story (e.g., at the grocery store, John tied his shoe). However, this pattern was restricted to recall performance only, as recognition performance was equivalent across the two types of sentences (disruptions and irrelevancies; Hudson, 1988).

Davidson and Hoe (1993) reported developmental evidence of an isolation effect. They found that children's memory (recall and recognition) of atypical actions was better than their memory for typical actions, on both immediate and delayed (1 day) testing. Further, recall performance for implausible atypical actions was better than that of plausible atypical actions on both days of testing. Recognition performance across the different types of atypical sentences (i.e., plausible, plausible within a sentence, and implausible), however, showed little difference. These results are consistent with those reported by Hudson (1988) as well as research on adults' memory for implausible sentences (e.g., Cox & Wollen, 1981; Einstein et al., 1989). The authors suggest that their results can be accounted for by the distinctiveness hypothesis. As such, these atypical actions were remembered better because they were presented within the context of common, script sentences (Davidson & Hoe, 1993). "Implausible sentences should orient attention and produce more distinctive encodings in the context of common sentences" (Davidson & Hoe, 1993, p. 122). The latter form a homogeneous background within which atypical sentences stand out. If the atypical sentences are plausible within

the context, it is possible that they amalgamate with the rest of the script actions, and become part of the script itself. If however, the atypical sentences are implausible within the context, the possibility of their amalgamation with script sentences is slim or nonexistent. Hence, these will be remembered better.

Davidson and Jergovic (1996) examined the disruption effect in two recall experiments with 6- and 8-year-old children. This effect refers to the finding that actions that are disruptive to the goal of a story are remembered better than atypical actions that bear no relevance to the goal of a story. The task involved both immediate and delayed recall of a scripted event containing different types of actions (i.e., script, pallid irrelevant, vivid irrelevant, obstacles, and distractions). The authors found a memory advantage for distractions that were disruptive (disruption effect) versus distractions that were not disruptive. Further, disruptive distractions were recalled better than any of the other actions, both immediately and after the 24 hr delay, across both age groups, and across all stories; obstacles were not recalled better than vivid irrelevant actions; vivid irrelevant actions were recalled better than pallid irrelevant actions; and pallid irrelevant actions were recalled most poorly across all events and time delays (i.e., even worse than script actions). Additional analyses revealed that after the 24 hr delay, recall of atypical actions declined somewhat, while recall of script actions improved.

Howe (1997) looked at the isolation effect in children's list learning. Children were shown lists of related black and white pictures that contained either (a) a coloured

related item, (b) a number, or (c) an unrelated item. Hence, both perceptual and conceptual distinctiveness was manipulated. Interestingly, perceptual distinctiveness (coloured related item) was beneficial (at acquisition and long-term retention) for both older (7-yr-olds) and younger (5-yr-olds) children, whereas conceptual distinctiveness (unrelated item) was beneficial only for the older children. Manipulating perceptual and conceptual distinctiveness simultaneously (i.e., presenting a number of a different colour) resulted in poorer recall performance. Howe (1997) suggests that this pattern of recall is consistent with a distinctiveness hypothesis and further proposes that there exists an *inverted U* relationship between distinctiveness and memory. That is, everyday common experiences do not create strong representations in memory; "too distinctive" experiences exhibit a negative effect on memory; and there exists an optimal, mid-range, level of distinctiveness that facilitates memory. Importantly, this was also the first study in which age differences at acquisition were controlled and the effects of distinctiveness at long-term retention in children's memory development were examined.

Distinctiveness effects in children's memory have also been examined in a handful of bizarre imagery studies (e.g., Emmerich & Ackerman, 1979; Merry & Graham, 1978; Tomasulo, 1982). For example, Emmerich and Ackerman (1979), using kindergarten children, employed a paired-associate learning task in which pairs of nouns were elaborated through either normal or bizarre interactions. No imaging instructions were provided during the acquisition procedure and retention tests consisted of cued-recall and

recognition tasks. Results were consistent with those of adult based cued-recall and recognition tests. The BIE was not evident. In fact, the opposite was true, with normal elaboration of noun pairs resulting in significantly better cued-recall performance. The recognition task failed to unveil any differences. These results should not be surprising, given that bizarreness is believed to only aid trace access.

In addition to examining memory for bizarre material, Tomasulo (1982) investigated the effects of age in a recall task. The author compared two groups of preschool children, mean ages of 4 yrs 7 mths and 3 yrs 6 mths, on their ability to recall normal, low bizarre, and high bizarre, line-drawn pictures of interacting object pairs. On the recall task, the older children were found to perform equally well across the three conditions but the younger children had significantly lower recall for the low bizarre and high bizarre conditions. Unfortunately, it is not clear what we can conclude from these findings because (a) the object line-drawings were neither accompanied by sentences describing the interactions nor by explicit imagery instructions at onset of the manipulation, (b) the objects were not matched for associative strength across the three conditions (e.g., Normal = matches lighting a pipe; Low Bizarre = pipe in a frying pan; High Bizarre = fish smoking a pipe), and (c) there was no control of learning across items or participants at acquisition (lack of criterion learning).

Last, Merry and Graham (1978) looked at imagery bizarreness and children's recall of sentences. The authors tested 12-13-year-olds on words recollected from sentences

they had previously *pictured in their minds* and rated as bizarre or common. The authors examined the effects of bizarre imagery on both immediate and long-term (one week) recall. In a group setting, children were asked to rate short sentences on the images produced (e.g., ordinary, unlikely, don't know, and can't imagine) followed by one of three recall conditions: (1) expected immediate recall with subsequent unexpected long-term recall, (2) unexpected immediate recall with subsequent unexpected long-term recall, and (3) unexpected recall at long-term retention only[6]. In all cases, children were requested to write down everything that they could remember. Recall was found to be significantly higher for words from bizarre sentences than for words from common sentences for all recall conditions.

It should be noted that Merry and Graham (1978) failed to: (a) control learning at acquisition, (b) investigate developmental trends, and (c) hold materials constant across conditions (e.g. Normal = The man smoked a cigar; Bizarre = The man pecked the worm). Thus, any observed differences between bizarre and common items could be due to differential learning or item artifacts. Importantly, each condition should have been associated with materials that were equivalent but for the interaction, or juxtaposition, between the items within the sentence, such that a common sentence (e.g., the man

[6] Participants were asked to rate the images they had formed from the sentences rather than the verbal meaning of the sentences, a task which also served as an incidental learning task. A set of abstract sentences further assessed if imagery was being used. Recall performance on this type of sentences was poorest for conditions 1 and 2, but not significantly different from recall performance of common sentences in condition 3.

smoked a cigar) would be matched to a bizarre sentence (e.g., the man ate a cigar). In this way, the inherent associative strength between the two nouns would have been constant across conditions (e.g., Lang, 1995), and subsequent results more easily ascribed to the experimental manipulation.

Methodological and Conceptual Limitations

As briefly discussed, distinctiveness research manipulating bizarre imagery has been limited and those studies that do exist have serious methodological problems that serve to limit their interpretability. Identifying current problems seems to be the first step toward increasing applicability of bizarre imagery to the concept of distinctiveness in children's memory. Moreover, identifying methodological and conceptual limitations of current research will serve to clarify the circumstances in which the distinctiveness effect is evident and potentially uncover the theoretical mechanisms behind it.

First, most investigators have concentrated on immediate, or short-term retention measures of memory, almost to the exclusion of long-term retention. Consequently, the locus (i.e., encoding, storage and retrieval) of the BIE (if reported) is frequently limited to initial acquisition. Failure to examine long-term retention performance has another consequence, namely, we do not know whether initial advantages (be they reported for common or bizarre items) are sustained over longer intervals. As acquisition advantages do not always translate into retention advantages, additional research is needed to

ascertain whether there are any long-term gains in memory associated with manipulating bizarre imagery.

Second, failure to equate item learning at acquisition characterizes the literature (e.g., Emmerich & Ackerman, 1979; Merry & Graham, 1978; Tomasulo, 1982). This becomes particularly important for interpretations of retention performance, be it short- or long-term. Specifically, a failure to ensure that a common criterion is used for all items, across conditions and participants, creates uncertainty in stating whether significant effects are due to actual differences in retention or to differential learning during the acquisition phase.

Third, there is a failure to control other factors (e.g., associative strength) that may vary between bizarre and common materials. The bizarre and common materials employed in many studies reviewed here were generally not equivalent across participants or conditions (see above discussion of limitations of Merry & Graham, 1978; Tomasulo, 1982). Again, this compromises the interpretation of the results from these studies. In addition, the issue of controlling image formation across participants and items has been problematic. The tendency of investigators to move from experimenter-generated images to self- or subject-generated images is a contributing factor to this problem. Clearly, the former allows one to gain more control over the experimental situation. Even with the implementation of this control, it would be difficult to ascertain whether young children in particular, are imaging the to-be-remembered materials as required. Although imaging

ability, per se, does not seem to be a factor in obtaining a BIE (see Anderson & Buyer, 1994), the effect is often greater when participants are given explicit imagery instructions than semantic processing instructions. Thus, a more consistent method should be devised to ensure consistent across-subject and across-condition imaging.

Fourth, a single, all-encompassing definition of bizarreness or distinctiveness is lacking. There are many different definitions, with these concepts being operationalized in many different ways. This is of particular concern in research with children. What we as adults deem common or bizarre does not necessarily map onto a child's notion of common and bizarre. This is clearly contingent on a number of variables such as the child's background knowledge and the child's current interpretation of the information presented (i.e., a particular interaction may be bizarre in certain instances and not in others, and the degree of bizarreness may depend on which of these instances the child focuses)[7].

Chosen because it maps onto most of the available research with younger children, the current experimental paradigm was designed with all of these considerations in mind. First, children were required to learn a mixed list of bizarre and common items to a strict acquisition criterion of two consecutive errorless trials. This ensures that learning should be reasonably well equated across item type (and age) at acquisition. Because easier items

[7] An anecdotal example comes from the present study. The example "The *snowflake* lands on the *mountain*" was given to a kindergarten child, and he was asked to label it as strange/weird or common/normal. The answer he gave was "Well, if it is in the winter then it's OK, but if it is in the summer then it's really strange". Clearly this manipulation of the presented information was not expected, yet it would make a world of difference to the child's interpretation of the event.

are learned faster than more difficult items and because younger children learn at a slower rate than older children (Howe & O'Sullivan, 1997), use of criterion learning procedures ensures these differences are not present at the end of the acquisition session. Interestingly, it has been argued that the use of this procedure creates a new problem, namely, over-learning confounds (see Brainerd & Reyna, 1995; Howe & Courage, 1997). Here, differences in over-learning might emerge such that easier items would be learned more completely (i.e., would have stronger traces) than more difficult items, and older children would learn any list of items more fully (i.e., more materials will be encoded) than younger children. As such, differences at retention would not be independent of differences at acquisition. However, over-learning effects have been found to be unrelated to memory durability. In two experiments involving 12-, 15- and 18-mths olds, Howe and Courage (1997) have shown that, " ... neither learning rate (mean trials to first recall) nor over-learning opportunities (mean trials after first correct recall or mean trials after last errors) predicted forgetting" (p. 156). Not confined to infants, similar results are noted with older participants (up to 11-year-olds; Brainerd & Reyna, 1995). Both of these studies used a stringent learning criterion (e.g., 2 consecutive errorless trials; see Howe & O'Sullivan, 1997, for a review of studies using criterion learning). As such, although some information may have been over-learned, the opportunity for under-learning was eliminated. Similarly, the 2-trial criterion used in the current study should serve to control

for differences in learning at acquisition, thus later recall should only reflect events that have transpired over the retention interval, not those at acquisition.

Second, as seen earlier (Lang, 1995), inherent item associations must either be removed or controlled in order to unambiguously measure the effects of bizarreness on memory. In the current study, the materials employed were equivalent across participants and conditions. All children were exposed to all the word pairs (only interactions differed) and objects were depicted in bizarre and common interactions an equal number of times. As such, objective (i.e., normed) associations between items were identical across participants and conditions. Thus, the BIE (if found) would be due to the bizarreness provided by the interaction specified between items and not to preexisting associative differences.

Third, a concerted attempt was made to control images across participants and conditions by explicitly manipulating the interactions of toy objects during the acquisition phase of the experiment. As such, all participants viewed, and were instructed to image, the same items in the same experimenter-provided interactive manner. Importantly, all participants had the equal opportunity (i.e., were provided with the same referents) to construct images, such that bizarre and common item conditions should have differed only in the interaction between the items and not the items themselves.

Fourth, children's long-term retention and the effects of distinctiveness using the bizarre imagery manipulation, were examined using a mathematical model based on the

trace-integrity framework (see Howe & O'Sullivan, 1997). This was done in order to evaluate whether bizarreness facilitates storage, retrieval, or both. Before describing the current experiment a brief overview of the trace-integrity framework is provided (for a more extensive review see Howe & O'Sullivan, 1997) including the storage and retrieval theory behind it.

Long-Term Retention Performance and the Trace-Integrity Framework

The processes reflected in changes observed across retention intervals merit more consideration than is apparent in the literature. Are these changes due to storage- or retrieval-based failures? More specifically, are they due to alterations in the memory traces, alterations in the accessibility of memory traces, or both? Retrieval-based theorists (e.g., Miller, Kasprow, & Schactman, 1986) often assume that forgetting is due to retrieval errors (i.e., memory storage is "static" and traces remain "intact") that can be produced, for example, by output interference between earlier and later stored traces (Howe & Brainerd, 1989). Essentially, forgetting depends on a failure to recollect information because of impoverished retrieval conditions (i.e., optimal retrieval cues are not provided) where participants are unable to access or activate otherwise intact traces. Although all traces survive, they do so at different strengths. Consequently, a decline in performance across time is explained in terms of the subjects' becoming less able to distinguish target from competing traces, particularly under non-optimal retrieval conditions (Howe & Brainerd, 1989). Because information remains intact in storage,

forgetting is said to be a matter of losing information "in memory" rather than "from memory" (Brainerd, Reyna, Howe, & Kingma, 1990).

On the other hand, storage-based theorists (e.g., Reynolds, 1977; Richardson, 1985) propose that declines in performance across time can also be due to storage failures. Forgetting occurs as the originally encoded material is "unlearned", or reorganized by newly acquired, interfering information (due to a weakening of bonds or loss of traces; Howe & Brainerd, 1989). Furthermore, original traces may become transformed by spontaneous internal restructuring of traces or bond strengths. In contrast to retrieval-based hypotheses, storage based theorists hold that there is a degeneration or transformation of traces "from memory" rather than "in memory" (Brainerd et al., 1990).

According to Howe and his colleagues (e.g., Howe & Brainerd, 1989; Howe & O'Sullivan, 1997), long-term retention performance has to be dependent on both storage and retrieval processes. Changes to memory traces appear to occur at encoding, storage, and retrieval over both short- and long-term intervals (Howe & Brainerd, 1989). Thus, it becomes important to distinguish those instances when accurate information in memory is inaccessible (i.e., retrieval failures) from those when it is unavailable (i.e., storage failures). Merely observing global performance changes across variable retention intervals is not sufficient to distinguish these possibilities. We need to identify and measure the factors that could produce such performance, and the disintegration/redintegration theory provides one method whereby this can be accomplished (Howe & Brainerd, 1989).

This theory builds on the trace-integrity model of long-term retention, and states that memory traces consist of primitive elements that are bound together into cohesive structures (Howe & O'Sullivan, 1997). During trace acquisition, these elements create representations in storage of to-be-remembered items which are then retrieved by later processes. Memorability is determined by the degree of cohesion among trace elements, or bond strength. *Disintegrative* types of changes (e.g. trace decay, interference, cross-talk among items) lead to the breakdown of bonds that hold primitive trace elements together. This does not necessarily imply that these traces are erased or that they vanish from memory. Rather, they can be reconstructed and modified in storage, and become indiscernible from the "background noise" of other trace elements, leaving open the possibility of trace reactivation or reinstatement during retention tests (Howe & O'Sullivan, 1997). Thus, although disintegrative changes may alter bonds between trace elements over time (i.e., threatening trace integrity), it is possible that redintegrative alterations (at test) can lead to recovery of some of the original information in storage through the redintegration of bonds. Within this trace-integrity framework, storage and retrieval processes are viewed as ends of a single continuum.

The implications of this model (see Howe & Brainerd, 1989; Howe & O'Sullivan, 1997) are (a) there exist processes that erode performance over the retention interval, such that forgetting could be due to changes in accessibility of information that is still intact in storage, or in the availability of that information in storage and (b) there also exist

processes that lead to a reinstatement of the material, which in turn, leads to the observation of an increase in retention performance during the retention trials themselves. An important contribution of this model is the idea that storage failures may not be permanent and that traces can be reinstated or restored. At the same time, it unifies storage- and retrieval-based forgetting, viewing amnesia and hypermnesia as a result of a single mechanism (i.e., the strength of bonds between primitive elements, or trace integrity; Howe & Brainerd, 1989).

Because this framework is implemented as a mathematical model, with independent measurements of storage- and retrieval-based failures, the relationships between the theoretical concepts of storage and retrieval, and empirical (error-success) outcomes on long-term retention tests are made explicit (Howe & O'Sullivan, 1997). The trace-integrity model described by the authors is consistent with other models with respect to trace features, bonds, mutability in storage, and the close connectedness of storage and retrieval, but is unique with respect to operationalizing these assumptions in the form of a testable mathematical framework (but see Riefer & Rouder, 1992). As such, assumptions are stated precisely (in the form of equations), are open to falsification (fit of model to data), and provide precise rules for linking observable performance to underlying theoretical constructs (Howe & O'Sullivan, 1997; see upcoming discussion in Results section of current study). Because of these advantages, this model will be used here to localize the effects of bizarreness in children's long-term retention.

Present Research

In the current experiment, participants were required to learn to-be-remembered items on a series of acquisition trials, with four subsequent recall tests of the items. Retention can be appraised by observing "global performance" and applying a number of comparisons. These can be made between the number of items recalled on the last acquisition trial and the number recalled on the first retention test, between the number of items recalled on the first and subsequent tests at retention, or between the number of items recalled on any earlier and later test at retention (Howe & Brainerd, 1989). Performance enhancing factors (i.e., reminiscence and hypermnesia) will be differentiated from performance limiting factors (i.e., amnesia and forgetting), by observing any of a number of possible outcomes which will be contrasted between common and bizarre/distinctive items. Howe and his colleagues (Howe, Kelland, Bryant-Brown, & Clark, 1992) define these factors as follows: *amnesia* (an observed net decrease in performance on retention tests), *hypermnesia* (an observed net increase in performance), *forgetting* (a failure to recall or recognize a particular item successfully recalled/recognized on a previous trial), and *reminiscence* (test-induced redintegrative process referring to successful recall/recognition of a particular item that had been unsuccessful on a previous trial). Further, two types of reminiscence are identified: *storage-based reminiscence* and *retrieval-based reminiscence*. Amnesia occurs when forgetting rates are greater than reminiscence rates, and hypermnesia occurs when

reminiscence rates are greater than forgetting rates (Howe et al., 1992). Consequently, retention performance (amnesia and hypermnesia) will be measured by considering both forgetting and reminiscence (for both types of items). Since long-term retention performance depends directly on these two variables and their interaction, when the two are "disentangled", Age X Retention interactions should be evident[8] (Howe et al., 1992). Importantly, these effects will be evaluated for both common and distinctive items.

The role of distinctiveness in children's long-term retention was directly examined and analyzed according to the trace-integrity model. Furthermore, developmental trends were examined by looking at recall by participants in two age groups, 5-year-olds (kindergarten) and 7-year-olds (2nd grade) across a 3-week retention interval. A free recall test was used instead of a cued recall with paired associate learning for two important reasons. First, bizarre imagery paradigms have shown a BIE with free recall versus cued recall or recognition (see Einstein & McDaniel, 1987). As discussed, this may be because distinctiveness helps with trace access rather than access of individual items within a trace (i.e., not redintegration). Second, the insensitivity of retention measures such as recognition and cued recall to developmental changes has been widely documented (e.g., Brainerd et al., 1990). Specifically, these tests generally ensure high

[8] This effect has been, in fact, emerging. Howe & O'Sullivan (1997) summarize that, a) forgetting is dominated by storage- rather than retrieval-based failures, b) trace recovery is a by-product of retrieval- not storage-based operations, and c) storage failure rates show significant declines with age, whereas development of retrieval recovery processes is only moderate.

levels of accuracy across all age groups, making these relatively poor choices when examining developmental trends.

It was expected that older children would exhibit greater total recall than younger children (see prior discussion). Along with a BIE, an Age X Bizarreness interaction was expected. Two possibilities exist: bizarre/distinctive items could either facilitate younger versus older children's recall, or older versus younger children's recall. The former interaction would be expected because younger children have poorer memory than do older children. Hence, any factor that improves these children's memory should raise their performance more than older children's, who may already be relying on other mnemonic strategies. The latter interaction however is also a possibility due to older children's more comprehensive background knowledge base (relative to younger children's), which would serve as a stronger secondary contextual comparative (again, relative to younger children). Since bizarre imagery paradigms are, in essence, manipulations of secondary distinctiveness, older children may perceive bizarre juxtapositions as more distinctive than would younger children. Hence, the effect would be more pronounced in the older age group.

Method

Participants

A total of 40 (20 female, 20 male) kindergarten children (mean age = 5 years 10 months, SD = 3.5 months) and 40 (25 female, 15 male) Grade 2 children (mean age = 7

years, 9 months, SD = 4 months) participated in this study. All of the children came from a predominantly middle class district and participated with school board and parental consent.

Design and Materials

A mixed-list, free-recall design was employed. The stimuli for this experiment consisted of 8 noun pairs adapted from an adult-normed list provided by Riefer and Rouder (1992). That is, sentences were rated on scales for bizarreness and level of interactions. Bizarre sentences were rated significantly higher for bizarreness than were common sentences, but the two types of sentences did not differ in their degree of interactiveness. Only the items deemed appropriate to our particular age groups were utilized[9]. For instance, nouns like *dog* and *bicycle* were used, not items like *lawyer* and *minister*. The former are more concrete than the latter, which is an important factor in the BIE (see Einstein and McDaniel, 1987 for review). More importantly, kindergarten children would be familiar with words such as *dog* and *bicycle*, whereas they might not be familiar with words like *lawyer* and *minister* (e.g., Cycowicz, Friedman, & Rothstein, 1997). For each pair of nouns, two sentences depicted either a common or bizarre interaction, for a total of sixteen possible interactions. For example, for the word pair dog-bicycle, the common sentence was, The *dog* chased the *bicycle*, whereas the bizarre

[9] To further ensure the adequacy of these items for our age groups these were normed on the children. Specifically, all children were asked to provide ratings of "common" or "bizarre" for each of the items, subsequent to the retention tests. Ratings obtained were similar to those of Riefer and Rouder (1992).

sentence was, The *dog* rode the *bicycle* (see Appendix A). The use of identical nouns for both bizarre and common interactions controlled for the problem of item dissimilarity common to previous studies of this nature. In addition, a set of toys (appropriate to both age groups, and both genders) were used to demonstrate each interaction concurrently with its verbal presentation. The same toys were used to depict both the bizarre and the common interaction for that particular noun pair. For example, children viewed either a toy *dog* chasing a toy *bicycle*, or a toy *dog* riding a toy *bicycle*. The use of the same toys to depict the interactions in a constant way was done in an attempt to provide perceptual-imaginal referents for the children and to promote the formation of similar, perceptually vivid images across participants and conditions. Moreover, the presentation of toys ensured that the acquisition task of memorizing the list was quite easy (i.e., for both sets of items), which allowed a clearer picture of retention effects, the focus of this study.

The stimuli were divided into four sets to which participants were randomly assigned. For each subject, 4 (half) items were bizarre and 4 were common. Further, there were 4 combinations of bizarre/common (see Appendix B). That is, which 4 items were bizarre and which were common was determined using 4 different combinations, such that a particular object-pair interaction was bizarre equally often as it was common across participants. The order of presentation of noun pairs during the learning session was also randomized from trial to trial to eliminate systematic effects of serial positioning.

Procedure

Participants were tested individually. Testing occurred over two sessions and was conducted in a quiet area in the children's school over a 3-week period. The to-be-remembered items were presented using a standard study-distractor-test procedure. The following instructions were given to each child at the onset of the experiment: "I am going to show you some toys, two at a time, and tell you something about them. I would like you to try to picture in your head what you see and hear and try to remember how the toys go together." Children heard the eight different interactions (as determined by the set, one to four, to which they were assigned), one at a time, along with the respective toy demonstration, for ten seconds each. The researcher enacted each interaction with the corresponding pair of toys for each of the different noun pairs. The manner of presentation was consistent from child to child, as well as the mode of delivery from condition to condition (i.e., bizarre and common interactions were delivered consistently within, as well as between, participants).

Following each study trial, participants were given a 30s distractor task during which they were required to match symbols and shapes according to a predesigned sample. This was introduced to minimize short-term memory (e.g., rehearsal, recency) effects. Immediately following the distractor task, children were asked to recall out loud as many of the noun pairs, within their respective interactions, as possible. The free-recall test was terminated when 20s had elapsed without a verbal response from the participants.

This study-distractor-test procedure continued until the children had learned the entire list of nouns to a criterion of two consecutive errorless test trials.

After the three-week retention interval, children were given a free recall test, using a sequence of test-distractor trials without further study opportunities. That is, each child was asked to recall what s/he could from the original learning session followed by 30s on the same distractor task used at acquisition, followed by a second recall-distractor trial, and so on, for a total of four test trials. Finally, each participant was administered a manipulation check to determine the level of bizarreness (as deemed appropriate by the child) of each of the presented interactions. The participants were asked what they thought of each noun pair interaction: "I would like you to tell me what you thought about what the toys were doing. Did you think what they were doing was really strange/weird, a little strange/weird, a little normal/common, or really normal/common?". Two clarifying questions were used. First, the child was asked "Do you think that what these toys were doing was weird/strange or common/normal?" Depending on the child's response, the subsequent question was "Do you think it was really strange/weird, or a little strange/weird?" or "Do you think it was really common/normal, or a little common/normal?". This was done in order to determine whether the children's individual ratings of the items (hereinafter referred to as child-normed judgements or data) corresponded to the adult-normed judgements of Riefer and Rouder (1992). It is important to establish a good operational definition of bizarreness, particularly when using

child participants. Norming the materials used by Riefer and Rouder on the child participants of this study more than addressed this issue. Both sets of data, child- and adult-normed, were subsequently analyzed.

Results

Because the results of adult- and child-normed data proved to be virtually identical, only adult-normed results are discussed. A complete summary of analysis of variance (ANOVA) and analysis of covariance (ANCOVA) results based on child-normed data is provided in Appendix C. An ANCOVA was used in order to eliminate so-called under- or over-learning effects at long-term retention (e.g., Howe, 1995; but see earlier discussion of the general absence of these effects in designs such as these). Errors at acquisition (both trial 1 errors and errors across all acquisition trials) served as the covariate (i.e., number of errors made acquiring bizarre interactions were used as the covariate for errors made in recalling bizarre interactions and number or errors made in acquiring common interactions were used as the covariate for errors made in recalling common interactions) and errors per trial at retention as the dependent variable. Global results are reported first, followed by a discussion of analyses using the trace-integrity model.

Global trends

Acquisition - Trial 1. To facilitate comparison with past research, mean errors on the first trial of acquisition were analyzed using a 2 (item: common v bizarre) x 2 (gender:

boys v girls) x 2 (grade: kindergarten v grade 2) ANOVA. Means and standard deviations are provided in Table 1. A main effect of grade, $F(1, 76) = 35.82$, $p = .000$, $MS_e = .65$ was observed, with older children making fewer errors ($M = .78$) on the first acquisition trial than younger children ($M = 1.14$). No significant main effects of either gender, $F(1, 76) = 3.63$, $p = .061$, $MS_e = .65$ or item, $F(1, 76) = .12$, $p = .735$, $MS_e = .98$, and no interactions were observed.

Table 1

Mean Errors (SDs) for Adult-Normed Data on Trial 1 and across all Trials at Acquisition

	Trial 1		All Trials	
	Ks x(sd)	**2s x(sd)**	**Ks x(sd)**	**2s x(sd)**
Males	1.13(.37)	.65(.25)	2.50(1.03)	.88(.42)
Females	1.16(.26)	.86(.25)	2.85(1.04)	1.69(1.05)

Note. because there were no item effects, mean errors were collapsed across conditions.

Acquisition - All Trials. Mean number of errors across all trials of acquisition were also analyzed using a 2 (item: common v bizarre) x 2 (gender: boys v girls) x 2 (grade: kindergarten v grade 2) ANOVA (see Table 1 for means and standard deviations). Main effects of grade, $F(1, 76) = 40.54$, $p = .000$, and gender, $F(1, 76) = 7.03$, $p = .010$,

MS_e = 7.36 emerged. As anticipated, 2nd grade children made fewer errors on all trials at acquisition (M = 1.39) than did kindergarten children (M = 2.68). Unexpectedly, however, boys made fewer errors (M = 1.81) than girls (M = 2.21). Finally, there was no main effect of item, $F(1, 76) = .16$, $p = .689$, $MS_e = 3.09$, and no interactions.

Retention - Acquisition Trial 1 Errors as Covariate. First, total errors at retention were analyzed using a 2 (item: common v bizarre) x 2 (grade: kindergarten v grade 2) x 2 (gender: boys v girls) x 4 (trial) ANCOVA, with mean errors made on the first trial of acquisition serving as the covariate. This was done for comparison purposes with previous research which does not employ criterion learning. The covariate was found to be significant for common, $F(1, 75) = 5.59$, $p = .021$, $MS_e = 1.42$, but not for bizarre, $F(1, 75) = .08$, $p = .776$, $MS_e = 3.72$, associations, $R^2 = .06$ and .00, respectively. Significant main effects for grade, item, and trial were found [$F(1, 75) = 17.24$, $p = .000$, $MS_e = 1.42$, $F(1, 75) = 5.67$, $p = .020$, $MS_e = 3.72$, $F(3, 228) = 57.25$, $p = .000$, $MS_e = .26$], but not for gender, $F(1,75) = .79$, $p = .376$, $MS_e = 1.42$ (means and standard deviations are the same for retention analyses, regardless of covariate [1 trial v all trials], and these are reported in Table 2). Specifically, older children made fewer errors at retention than younger children, all children made fewer errors in recalling bizarre than common items, and all children made fewer and fewer errors as the retention trials unfolded. A significant Item x Trial interaction was also observed, $F(3, 228) = 8.12$, $p = .000$, $MS_e = .31$. Since

this interaction is identical to the one observed when mean errors across all acquisition trials served as the covariate, it will be discussed in the next section.

Table 2

Mean Proportion (SDs) of Errors per Trial at Retention for Adult-Normed Data

Grade/Item Type	Trial			
	1 x(sd)	2 x(sd)	3 x(sd)	4 x(sd)
K				
Common	.57(.26)	.56(.23)	.49(.27)	.49(.24)
Bizarre	.57(.27)	.43(.25)	.35(.21)	.28(.21)
2				
Common	.42(.26)	.39(.27)	.28(.20)	.31(.23)
Bizarre	.39(.25)	.28(.21)	.23(.20)	.18(.16)

Retention - All Acquisition Trial Errors. Second, total errors at retention were analyzed using a 2 (item: common v bizarre) x 2 (grade: kindergarten v grade 2) x 2 (gender: boys v girls) x 4 (trial) ANCOVA, with total errors across acquisition trials serving as the covariate. The covariate was significant for both common, $F(1, 75) = 8.06$, $p = .006$, $MS_e = 1.38$, and bizarre, $F(1, 75) = 8.29$, $p = .005$, $MS_e = 3.35$, associations, $R^2 = .09$ and .08, respectively. Significant main effects for grade, item and trial were found

[$F(1, 75) = 14.33$, $p = .000$, $MS_e = 1.38$, $F(1, 75) = 6.99$, $p = .010$, $MS_e = 3.35$, $F(3, 228) = 57.25$, $p = .000$, $MS_e = .26$], but not for gender, $F(1,75) = .31$, $p = .576$, $MS_e = 1.38$ (see Table 2 for means and standard deviations). Grade 2 children made fewer errors than kindergarten children, all children made fewer errors in recalling bizarre than common items, and all children made fewer and fewer errors as the retention trials unfolded.

The significant Item x Trial interaction at long-term retention, $F(3, 228) = 8.12$, $p = .000$, $MS_e = .31$, is plotted in Figure 1. Post hoc analyses confirmed that errors tended to decrease at a faster rate for the bizarre items than for the common items. Newman-Keuls tests ($p < .05$) data indicated the following ordering for common and bizarre items, respectively (mean proportions in parentheses): Trial 1 (.49) = Trial 2 (.48) > Trial 3 (.39) = Trial 4 (.40) and Trial 1 (.48) > Trial 2 (.35) > Trial 3 (.29) > Trial 4 (.23).

At a global level then, 2nd graders made fewer errors at long-term retention than kindergartners, children made fewer errors recalling bizarre than common events, and children made fewer errors as the retention trials unfolded. Further, bizarreness of to-be-remembered events seemed to facilitate children's recall across trials beyond what was observed for common items. This was evidenced using both sets of norms (see Appendix C for effects using child-normed data). In other words, it looks as though bizarreness promotes hypermnesia at long-term retention. Next, the trace-integrity model is used to see if this should indeed be interpreted as a retrieval effect (e.g., Riefer & Rouder, 1992).

More specifically, the model will be used to isolate storage and retrieval contributions to forgetting and reminiscence.

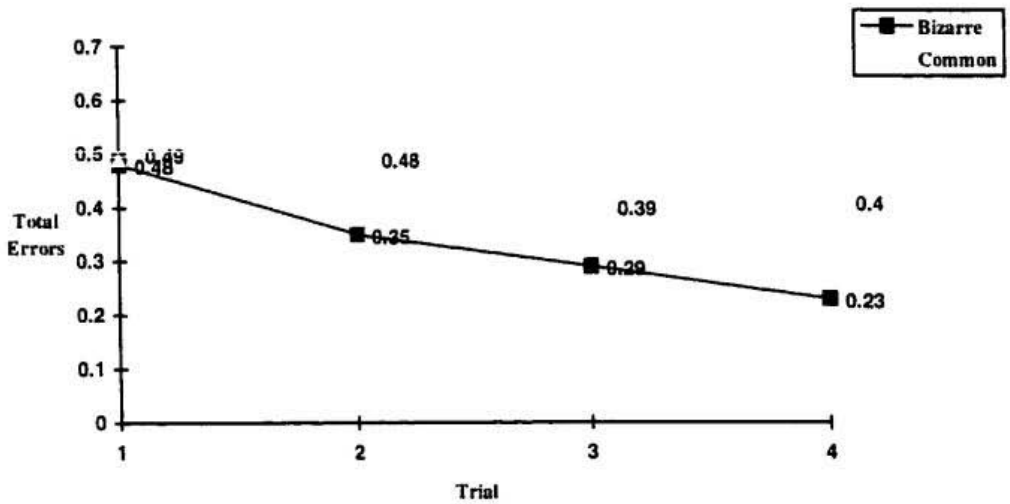

Figure 1. Mean proportion errors across retention trials for adult-normed bizarre and common materials (collapsed across age).

Trace-Integrity Analyses

Before applying the trace-integrity model, the degree of fit of the model to the data must be established and its parameters estimated. Standard likelihood-ratio procedures are available for parameter estimation and goodness-of-fit (see Howe & Brainerd, 1989; Howe, 1991; Howe & O'Sullivan, 1997). This task is accomplished using the following five-step sequence.

First, the data space is translated into an empirical probability space, and a function is derived which gives the a posteriori likelihood of that data. In the present experiment (and other similar four-trial experiments), this data space consists of 16 unique outcomes. The model provides independent estimates of the availability (in storage) and accessibility (retrievability) of memory traces, as well as estimates of relearning (restorage or retrieval-based) during the actual retention test. This is accomplished by separating forgetting and relearning components over this outcome space of correct (C) and incorrect (E) responses, or errors: $C_1C_2C_3C_4$, $C_1C_2C_3E_4$,.....,$E_1E_2E_3C_4$, $E_1E_2E_3E_4$, where subscripts 1-4 represent the four retention tests. Probabilities are then assigned to each outcome (i.e., from $p(C_1C_2C_3C_4)$ which represents the probability that an item is correctly recalled on all four trials, to $p(E_1E_2E_3E_4)$ which represents the probability that an item is incorrectly recalled on all trials) converting the data space into an empirical probability space. The a posteriori likelihood of the data can then be derived, one that contains 15 degrees of freedom (i.e., L_{15}; see Howe, 1995).

Second, this empirical probability space is converted into a mathematical space, by expressing the 16 empirical probabilities in terms of the model's nine parameters (Table 3). A second likelihood function with nine degrees of freedom is then derived based on these parameters. The nine parameters are composed of two forgetting (R and S) and seven relearning (a, r_i and f_i) parameters (see Table 4). R refers to retrieval-based forgetting (i.e., the conditional probability that although an item is in storage it is not accessible) and S refers to storage-based forgetting (i.e., the unconditional probability that an item is no longer available in storage). a is the restorage parameter (i.e., redintegration - or conditional probability that a trace can be restored) while the r_i and the f_i measure retrieval relearning, the former following successful recall and the latter following unsuccessful recall. More specifically, r_1, r_2, and r_3, provide measures of the probability of successfully recalling an item following one, two, or three preceding successes, respectively. f_1, f_2, and f_3, provide measures of the probability of successfully recalling an item following one, two, or three preceding errors, respectively (for a more detailed review see Howe & Brainerd, 1989; Howe & O'Sullivan, 1997).

Third, the number of times that each of the 16 possible outcomes occurred in the sample data is counted, numerical values of the model's nine parameters are estimated the values for both the fifteen- and the nine-parameter likelihood function are obtained. These values are then used to evaluate goodness of fit of the model and to test hypotheses regarding between- and within-condition differences.

Table 3

Mathematical Expressions Defining the Empirical Outcome Space

Outcome	Expression
p(CCCC)	$(1-S)(1-R)r_1r_2r_3$
p(CCCE)	$(1-S)(1-R)r_1r_2(1-r_3)$
p(CCEC)	$(1-S)(1-R)r_1(1-r_2)f_1$
p(CECC)	$(1-S)(1-R)(1-r_1)f_1r_1$
p(ECCC)	$Sa(1-R)r_1r_2 + (1-S)Rf_1r_1r_2$
p(CCEE)	$(1-S)(1-R)r_1(1-r_2)(1-f_1)$
p(CECE)	$(1-S)(1-R)(1-r_1)f_1(1-r_1)$
p(ECCE)	$Sa(1-R)r_1(1-r_2) + (1-S)Rf_1r_1(1-r_2)$
p(CEEC)	$(1-S)(1-R)(1-r_1)(1-f_1)f_2$
p(ECEC)	$Sa(1-R)(1-r_1)f_1 + (1-S)Rf_1(1-r_1)f_1$
p(EECC)	$S(1-a)a(1-R)r_1 + SaRf_1r_1 + (1-S)R(1-f_1)f_2r_1$
p(CEEE)	$(1-S)(1-R)(1-r_1)(1-f_1)(1-f_2)$
p(ECEE)	$Sa(1-R)(1-r_1)(1-f_1) + (1-S)Rf_1(1-r_1)(1-f_1)$
p(EECE)	$S(1-a)a(1-R)(a-r_1) + SaRf_1(1-r_1) + (1-S)R(1-f_1)f_2(1-r_1)$
p(EEEC)	$S(1-a)^2a(1-R) + S(1-a)aRf_1 + SaR(1-f_1)f_2 + (1-S)R(1-f_1)(1-f_2)f_3$
p(EEEE)	$S(1-a)^3 + S(1-a)^2aR + S(1-a)aR(1-f_1) + SaR(1-f_1)(1-f_2)$ $+ (1-S)R(1-f_1)(1-f_2)(1-f_3)$

Note. From "Misleading children's story recall: Forgetting and reminiscence of the facts", by M. Howe, 1991, Developmental Psychology, 27, p. 750.

Table 4

Theoretical Definitions of the Trace-Integrity Theory's Parameters

Process and parameter	Description
Forgetting	
S	Probability of storage failure
R	Probability of retrieval failure of information in storage
Reminiscence	
a	Probability that information in storage is redintegrated to a level above zero recall
r_1	Probability of two consecutive successes
r_2	Probability of three consecutive successes
r_3	Probability of four consecutive successes
f_1	Probability of a success after one error
f_2	Probability of a success after two consecutive errors
f_3	Probability of a success after three consecutive errors

Note. From "Misleading children's story recall: Forgetting and reminiscence of the facts", by M. Howe, 1991, Developmental Psychology, 27, p. 749.

Fourth, goodness of fit tests evaluate whether the trace-integrity model provides an adequate account of the data, and takes the form $\chi^2(6) = (-2\ln L_9) - (-2\ln L_{15})$. Goodness of fit is determined by establishing whether or not the nine-parameter model provides as good an account of the data (L_9) as the empirical model (L_{15}; e.g., see Howe & Brainerd, 1989). Goodness of fit was demonstrated (i.e., the model provided an excellent account of the data), with a value of $\chi^2(6)$s < 10.25 ($p > .05$), for all conditions.

Last, hypotheses regarding theoretical processes of retention performance are tested. A three-step likelihood-ratio procedure was employed beginning with an *experimentwise* test, followed by *conditionwise* and *parameterwise* tests. First, the experimentwise test established that parameter values varied across the experiment as a whole, with a value for this test statistic of $\chi^2(27) = 94.59$, $p < .001$. Second, four conditionwise tests were used to establish that parameter values varied between specific pairs of conditions. The numerical results of these tests showed that all pairwise age and item contrasts were significant: (a) age effects (kindergarten v Grade 2) for the common ($\chi^2(9) = 17.66$, $p < .05$) and bizarre ($\chi^2(9) = 26.62$, $p < .01$) items and (b) item effects (common v bizarre) for kindergartners ($\chi^2(9) = 38.83$, $p < .001$) and second graders ($\chi^2(9) = 29.37$, $p < .001$). Finally, parameterwise tests established whether a specific parameter is significantly different in value for any two conditions that differ significantly. Thus, for each of the significant conditionwise tests, the model's nine parameters were compared, for a total of 36 parameterwise tests. Significant findings are summarized for the age and

item effects. Table 5 provides numerical estimates of the model's theoretical parameters, which are used to analyze differences in forgetting and reminiscence rates between and within conditions, and storage and retrieval loci of these differences (Howe, 1995). Because results were identical when examining either child- or adult-normed data, only the latter are reported here.

Table 5

Estimates of the Trace-Integrity Model's Theoretical Parameters

	Parameter								
Condition	S	R	a	r_1	r_2	r_3	f_1	f_2	f_3
Kindergarten									
Common	.41	.25	.02	.79	.88	.88	.58	.74	.80
Bizarre	.40	.27	.12	.95	.99	1.0	.80	.98	1.0
Grade 2									
Common	.22	.26	.04	.86	.92	.95	.63	.67	.00
Bizarre	.30	.12	.25	.96	.99	1.0	.71	.00	.08

Age effects. Age differences were localized primarily at forgetting, not reminiscence. Kindergarten children exhibited more storage-based forgetting (parameter S) than grade 2 children on both common (.41 v .22, respectively) and bizarre (.40 v .30, respectively) conditions, and also more retrieval-based forgetting (parameter R) than grade 2 children on bizarre items (.27 v .12, respectively). Further, grade 2 children exhibited

more storage-based reminiscence on the bizarre items than kindergarten children (parameter a; .25 v .12, respectively), partially explaining the hypermnesia effects seen in the earlier, ANCOVA analyses.

Although there were trends in retrieval-based reminiscence age groups were differentiated only with respect to error-contingent retrieval-based reminiscence. Specifically, although the rate of success-contingent retrieval-based reminiscence (r_i values) was fairly stable across trials for both kindergarten and grade 2 children, error-contingent retrieval-based reminiscence (f_i values) increased across trials for kindergarten children and decreased for grade 2 children. That is, kindergartners were more likely to recall an item (whether bizarre or common) after one, two, and even three consecutive errors, whereas grade 2 children's recall performance was more stable (i.e., the probability of successful recall after one error was fairly high, but the probability of successful recall after two, or three, consecutive errors was extremely low).

Item effects. The main effect due to manipulating item distinctiveness was at reminiscence. In particular, children of both ages showed a higher probability of storage-based reminiscence (parameter a) on the bizarre than common items (kindergarten = .12 v .02, respectively, and grade 2 = .25 v .04, respectively). Further, success-contingent retrieval-based reminiscence (r_i values) was greater for bizarre than for common items for both age groups (see Table 5). However, error-contingent retrieval-based reminiscence (f_i values) parameters showed an item effect only for kindergartners such that bizarre items

exhibited a steeper increasing slope than common items. That is, the probability of recalling an item after one, two, or three consecutive errors was much greater for bizarre than for common items for these children. For grade 2 children, although error-contingent retrieval-based reminiscence was stable across conditions (i.e., f_i parameter was not reliable) retrieval-based forgetting (R) was twice as large for common (.26) than for bizarre (.12) items.

Discussion

It was suggested earlier that methodological and conceptual limitations of distinctiveness research with children (including its operationalization as bizarreness) were related to the different results reported in this literature. Specifically, failures to set a learning criterion at acquisition, to equate materials across participants and conditions (i.e., with respect to inherent associations and interactions between items), and to control images across participants and conditions, all limited the interpretation of recall performance particularly when developmental differences were examined. Moreover, few studies included an assessment of distinctiveness effects beyond immediate acquisition. As effects at acquisition are not always the same as those at long-term retention, additional research was required. Because of these shortcomings certain issues have not been clarified in the previous literature on the BIE, and more importantly, in the literature on distinctiveness. For example, it is not clear what the loci of distinctiveness effects are (particularly with respect to long-term retention) and if any developmental differences do

exist, what the locus of these differences might be. As noted in the current research, correction of these methodological and conceptual limitations, along with the application of the trace-integrity model, revealed both the locus of distinctiveness effects at long-term retention as well as the locus of developmental differences between the retention of distinctive (bizarre) and common information.

Several findings emerged from the global analyses in the current study. First, as expected, older children made fewer errors both on the first acquisition trial and in reaching criterion (i.e., required fewer trials overall) than younger children. Further, there was no observed item effect in either of these conditions This is in contrast with previous findings of a BIE or, for that matter, a commonness effect. Riefer and Rouder (1992), for example, report "a consistent advantage for bizarre over common sentences for free, ... immediate, ... recall" (p. 606), whereas Tomasulo (1982) reports greater immediate recall for common than bizarre items for the younger children in the study. Although at first blush a bizarreness effect on immediate free recall would seem obvious, a commonness effect also makes sense. Acquiring new information is, ultimately, an attempt to integrate this information into already existing knowledge structures. By definition, common events exploit an individual's prior experience with these events, and, as such, these events should be more easily assimilated into one's background knowledge and acquired more easily. The acquisition of an event depicting a common juxtaposition between two items

should, therefore, be favourable over the acquisition of an event depicting a bizarre juxtaposition.

Although the focus of the current study was not immediate retention, performance on the first acquisition trial (and each subsequent trial necessary to reach criterion) was, in essence, immediate, free recall. It may have been that the task here was too easy at acquisition to observe item effects. That is, in an attempt to control images across participants and item conditions by manipulating toys, the current learning task was made notably easier than incidental learning tasks of previous research (including Riefer & Rouder's [1992], or Tomasulo's [1982]), the consequence of which may have been the absence of item effects on immediate free recall performance.

Interestingly, a main effect of gender favouring the boys in the sample, was quite surprising. Typically, gender effects in learning tend to favour girls (e.g., Kramer, Delis, Kaplan, O'Donnell, & Prifitera, 1997). This effect disappeared when the acquisition analysis was conducted on first trial errors only (consistent with prior research) and was not apparent in the long-term retention analyses. An explanation for this effect is not readily available. As study items were largely gender-neutral this does not seem to be an item effect, and is most likely a sampling effect. However, albeit spurious in nature, it bears watching in future research.

Global analyses at long-term retention revealed results consistent with pre-experimental expectations. First, older children made fewer errors at retention than

younger children across all item conditions. Second, a BIE (i.e., an item effect) was observed at long-term retention. Global performance analyses showed that children in both age groups not only made significantly fewer errors in recalling bizarre than common items, but also made fewer and fewer errors as the retention trials unfolded. Theoretically, once these events were encoded, common juxtapositions became amalgamated into the larger background of the children's experiences, whereas bizarre juxtapositions somehow stood apart from the larger knowledge base. At long-term retention then, the distinctive features of the latter resulted in enhanced recall for these items. These findings are consistent with prior long-term retention research with older children (e.g., Merry & Graham, 1978), however, they are limited in that, they do not reveal any developmental differences regarding the effects of bizarreness on children's memory, nor do they show the locus of the BIE. Although improvement in recall across trials indicates the presence of hypermnesia, further analyses with the trace-integrity model were needed to reveal significant effects of reminiscence and forgetting.

Model-based analyses uncovered several developmental trends. First, kindergarten children showed a much higher probability of storage-based forgetting than grade 2 children on both common and bizarre items. This result is consistent with those of Howe and O'Sullivan (1997), who report that storage failure rates decline with age in childhood. That these effects were observed for both types of items suggests that developmental differences in storage failures are not affected by distinctiveness (unlike for instance

reminiscence and misinformation effects which generally affect traces at the level of storage). Generally, the effect of destructive and constructive elements that transpire during a retention interval is at *storage* (Howe & O'Sullivan, 1997). The distinctiveness or commonness of an event however, is first encountered during acquisition. It could have been that distinctiveness inoculates traces against storage-based failures by making these stronger against trace-altering factors. But, this did not prove to be the case. Bizarre events did not enjoy lower probabilities of storage-based forgetting than common events. Instead, it seems as though these items benefited from a reduction in retrieval-based failures, but only for the older, grade 2 children.

Second, rates of retrieval-based forgetting for common items were developmentally similar, which is consistent with previous reports (e.g., Howe & O'Sullivan, 1997) that not only is forgetting dominated by storage failure rates, but it is these rates that decline with age, while retrieval-based failures remain stable. In contrast, older children in the current study showed lower probabilities of retrieval-based forgetting for bizarre events than younger children. Further, bizarre events enjoyed lower probabilities of retrieval failures than common events for these older children. This is consistent with Riefer and Rouder's (1992) claim that distinctiveness influences retrieval operations. Subsequent analyses in the current study, however, showed that this is not the entire picture.

Last, developmental effects were observed on the storage-based reminiscence parameter. Specifically, older children showed greater probability of storage-based recovery than younger children, but only for bizarre events. Further, a trend was seen (albeit unreliable) such that retrieval stability rates tended to be uniform and fairly high, and storage recovery rates were lower in comparison to retrieval recovery rates.

Model-based analyses also revealed important item effects, otherwise not seen with the ANOVA and ANCOVA analyses. Perhaps the most important result noted was with respect to the storage-based reminiscence parameter. Bizarre relationships promoted greater storage-based recovery than common relationships. In contrast, Howe and O'Sullivan (1997) report that it is retrieval-based recovery operations which tend to dominate retention tests, with storage-based recovery operations being less reliable. As such, once information crosses the storage failure threshold it becomes very difficult to access that information on subsequent retention tests. Interestingly, an exception to this rule is akin to the effects of bizarreness observed in the current study. Specifically, easily categorized (semantically related) material "can be refurbished using storage, as well as retrieval, recovery operations, especially by older children" (Howe & O'Sullivan, 1997, p. 187). Similarly, bizarre relationships affected storage-based reminiscence (i.e., helped refurbishment of degraded information). Although both age groups showed this effect, older children enjoyed increased storage-based recovery over younger children (i.e., *a* parameter doubled) for the bizarre events. Significant effects on the retrieval-based

reminiscence parameters (both success- and error-contingent) however, were only evident for the younger children in the study.

Further, the effects of bizarreness contrast with the effects of manipulations such as reinstatement, misinformation, and retroactive interference which affect storage-based forgetting rates primarily, with little impact on retention testing (and occur during the retention interval; see Howe & O'Sullivan, 1997). Specifically their effects are linked directly to trace maintenance during the retention interval and indirectly to trace recovery during testing. Bizarreness, however, seems to have little effect on storage-based forgetting (although it does impact retrieval-based forgetting for older children) and a direct impact on storage-based recovery operations. Whereas the former variables primarily affect the storage component of forgetting, but not recovery, bizarreness primarily affects the storage component of recovery, but not forgetting.

These findings are quite interesting given prior hypotheses for this effect (i.e., bizarre materials benefit retrieval and common materials benefit storage; Riefer & LaMay, 1998). Although these authors conclude that memory for bizarre items is a matter of retrieval benefits (such that these traces are more accessible than common traces), they suggest that this may also be a matter of storage benefits (see prior discussion on quality of bizarre and common images). By manipulating vividness and controlling item acquisition across conditions, the current research shows storage to be an important factor for bizarre items (i.e., an increase in storage-based reminiscence). Importantly, Riefer and

Rouder's (1992) study lacked differentiation between components of storage and retrieval at acquisition and those at retention. Specifically, the authors' multinomial model separated storage and retrieval factors at acquisition only. Immediate free recall tests are, in essence, tests of acquisition which can not be compared with storage and retrieval results at retention (current model). Hence, statements such as "... the effect of distinctiveness is somehow to improve the retrievability of bizarre memory traces" (Riefer & Rouder, 1992) are not entirely definitive of the loci of distinctiveness at retention. Rather, a BIE effect in Riefer and Rouder's (1992) study would more precisely relate to item acquisition. Specifically, bizarre events seemed to have had an acquisition advantage over common events, the locus of which was at retrieval. Moreover, in the Riefer and Rouder (1992) study, the effect of distinctiveness on redintegrative changes was left unexamined. In fairness, however, the use of one free recall trial[10] by these authors eliminates the possibility of examining relearning across retention trials. Hence, reminiscence can only be considered in the present study. Whereas the multinomial model of Riefer and Rouder (1992) does not separate forgetting and relearning components of retention, the current model does (e.g., Howe & Brainerd.1989; Howe & O'Sullivan, 1997).

Thus, it seems that not only does the bizarre imagery effect arise in younger ages,

[10] The free recall trial was followed by a cued recall test which showed no difference between bizarre and common items. To ensure that the free recall trial did not serve as a learning trial, Riefer and Rouder (1992) conducted a separate study using only cued-recall. Here again, no significant differences were noted between items.

but developmentally, bizarreness effects seem to increase both quantitatively (probability of storage-based reminiscence doubles between kindergarten and Grade 2) and qualitatively (Grade 2 children exhibit a beneficial effect of bizarreness on the retrieval-based forgetting parameter whereas kindergarten children do not). Evidently, long-term retention performance of distinctiveness (at least when operationalized as bizarreness) can not be explained by storage- or retrieval-based theories alone. Only when considered in light of the disintegration/redintegration theory, did it become clear that distinctiveness affects both storage and retrieval components of forgetting and relearning.

Although these findings are fairly clear cut, consider some drawbacks to the current approach. First, Schmidt (1991) suggests that there is an inherent problem with mixed-list designs. Specifically, memory for distinctive items is compared to memory for common items within the same list. As such, the positive effects on memory of distinctive items can not be differentiated from the negative effects of these items on memory for common items. To avoid this problem, a homogenous or unmixed-list design needs to be used. Admittedly, this does pose a dilemma in the current research. Although the direct goal of this study was to examine the beneficial effects of distinctiveness on children's long-term retention, the effect of this on memory for common items is also important. It is possible, that the distinctive nature of the bizarre materials negatively affected memory for the common materials. Perhaps the distinctive nature of the former created a situation whereby these items became central features of the testing event and common items

adopted a more peripheral role. Although purely speculative, should this be the case, memory would be worsened for the peripheral items. Such a possibility would be consistent with memory for high-priority events (e.g., Christianson & Loftus, 1987; also see discussion in Howe, 1997).

Second, although establishing a good operational definition of bizarreness/distinctiveness is difficult (particularly in research involving children), the current experimental manipulation more than took this task into consideration. This was accomplished by norming the materials implemented herein (adult-normed by Riefer & Rouder, 1992) on the child participants. Although some children disagreed with some of the sentences deemed bizarre by adult standards, in the main the ratings were highly similar. Further, analyses (i.e., global and trace-integrity model) performed on these norms (i.e., involving acquisition and retention recall performance) produced virtually identical results. It was, indeed, reassuring to see the same trends emerge, for both loci of distinctiveness and developmental differences in the BIE effect. As such, these results enforce the reliability and validity of the materials employed.

Of course, future research should consider other aspects of distinctiveness effects in children's long-term retention. For example, concentrating on other ages would ascertain the appropriateness of generalizing these findings across childhood. Additional avenues could also lead to answers concerning the interaction between interference and distinctiveness in long-term retention, with child participants. It would, indeed, be

interesting to unveil the effects of manipulations such as misinformation and retroactive interference on distinctive events, and whether these are developmentally constant or variable. Further, longitudinal cross-sectional designs may also help in establishing developmental trends as well as uncovering individual differences within the BIE.

In sum, the research presented here has revealed several developmental trends that are consistent with prior research on children's memory (e.g., Howe & O'Sullivan, 1997). First, storage failure rates declined with age. Younger children showed more storage-based forgetting than older children, regardless of item condition. Second, retrieval-based failures were developmentally invariant for common items. However, retrieval-based forgetting did decline for older children when bizarre items were used. Concerning distinctiveness, there were no item effects at acquisition for either bizarre or common items. The absence of either a BIE or commonness effect at acquisition may have been due to the relative ease of the learning procedure used here, rather than the absence of such effects in children's learning. Regardless, distinctiveness effects were evident at long-term retention, particularly in terms of hypermnesia. With the application of the trace-integrity model it was shown that these effects were not due solely to retrieval factors but rather, to the redintegration of information that has undergone storage-based failure. Thus, although traces can be altered in storage, it seems that if the traces are distinctive the probability of their refurbishment in storage is significantly increased. This was particularly true for the older children. Unlike other manipulations, then, ones that

primarily reduce storage-based forgetting in younger children's retention performance, distinctiveness improves older children's storage-based redintegration (at least when operationalized as bizarreness).

References

Anderson, D. C., & Buyer, L. S. (1994). Is imagery a functional component of the "bizarre imagery" phenomenon? *American Journal of Psychology, 107*, 207-222.

Brainerd, C. J., & Reyna, V. F. (1995). Learning rate, learning opportunities, and the development of forgetting. *Developmental Psychology, 31*, 251-262.

Brainerd, C. J., Reyna, V. F., Howe, M. L., & Kingma, J. (1990). Development of forgetting and reminiscence. *Monographs of the Society for Research in Child Development, 55*, (3-4, Serial No. 222).

Bruce, D., & Gaines, M. T. (1976). Tests of an organizational hypothesis of isolation effects in free recall. *Journal of Verbal Learning and Verbal Behavior, 15*, 59-72.

Cox, S. D., & Wollen, K. A. (1981). Bizarreness and recall. *Bulletin of the Psychonomic Society, 18*, 244-245.

Cycowicz, Y. M., Friedman, D., & Rothstein, M. (1997). Picture naming by young children: Norms for name agreement, familiarity and visual complexity. *Journal of Experimental Child Psychology, 65*, 171-237.

Davidson, D., & Jergovic, D. (1996). Children's memory for atypical actions in script-based stories: An examination of the disruption effect. *Journal of Experimental Child Psychology, 61*, 134-152.

Davidson, D., & Hoe, S. (1993). Children's recall and recognition memory for typical and atypical actions in script-based stories. *Journal of Experimental Child Psychology, 55*, 104-126.

Einstein, G. O., & McDaniel, M. A. (1987). Distinctiveness and the mnemonic benefits of bizarre imagery. In M. A. McDaniel & M. Pressley (Eds.), *Imagery and related processes: Theories, individual difference, and applications* (pp. 78-102). New York: Springer-Verlag.

Einstein, G. O., McDaniel, M. A., & Lackey, S. (1989). Bizarre imagery, interference, and distinctiveness. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 15*, 137-146.

Emmerich, H. J., & Ackerman, B. P. (1979). A test of bizarre interaction as a factor in children's memory. *Journal of Genetic Psychology, 134*, 225-232.

Fritsch, T., & Larsen, J. D. (1990). Image-formation time is not related to recall of bizarre and plausible images. *Perceptual and Motor Skills, 70*, 1259-1266.

Green, R. T. (1956). Surprise as a factor in the von Restorff effect. *Journal of Experimental Psychology, 52*, 340-344.

Hacuk, P. D., Walsh, C. C., & Kroll, N. E. A. (1976). Visual imagery mnemonics: Common vs. bizarre mental images. *Bulletin of the Psychonomic Society, 7*, 160-162.

Herz, R. S. (1997). The effects of cue distinctiveness on odor-based context-dependent memory. *Memory & Cognition, 25*, 375-380.

Herz, R. S., & Cupchik, G. C. (1995). The emotional distinctiveness of odor-evoked memories. *Chemical Senses, 20*, 517-528

Hirshman, E. (1988). The expectation-violation effect: Paradoxical effects of semantic relatedness. *Journal of Memory and Language, 27*, 40-58

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

Howe, M. L. (1997). Children's memory for traumatic experiences. *Learning and Individual Differences, 9*, 153-174.

Howe, M. L., & Brainerd, C. J. (1989). Development of children's long-term retention. *Developmental Review, 9*, 301-340.

Howe, M. L., & Courage, M. L. (1997). The emergence and early development of autobiographical memory. *Psychological Review, 104*, 499-523.

Howe, M. L., & Courage, M. L. (1997). Independent paths in the development of infant learning and forgetting. *Journal of Experimental Child Psychology, 67*, 131-163.

Howe, M. L., Courage, M. L., & Bryant-Brown, L. (1993). Reinstating preschoolers' memories. *Developmental Psychology, 29*, 854-869.

Howe, M. L., Kelland, A., Bryant-Brown, L., & Clark, S. L. (1992). Measuring the

development of children's amnesia and hypermnesia. In M. L. Howe, C. J. Brainerd, & V. F. Reyna (Eds.), *Development of long-term retention* (pp. 56-103). New York: Springer-Verlag.

Howe, M. L., & O'Sullivan, J. T. (1997). What children's memories tell us about recalling our childhoods: A review of storage and retrieval processes in the development of long-term retention. *Developmental Review, 17*, 148-204.

Howe, M. L., O'Sullivan, J. T., & Marche, T. A. (1992). Toward a theory of the development of long-term retention. In M. L. Howe, C. J. Brainerd, & V. F. Reyna (Eds.), *Development of Long-Term Retention* (pp. 245-255). New York: Springer-Verlag.

Hudson, J. A. (1988). Children's memory for atypical actions in script-based stories: Evidence for a disruption effect. *Journal of Experimental Child Psychology, 46*, 159-173.

Hunt, R. R. (1995). The subtlety of distinctiveness: What von Restorff really did. *Psychonomic Bulletin & Review, 2*, 105-112.

Hunt, R. R., & Einstein, G. O. (1981). Relational and item-specific information in memory. *Journal of Verbal Learning and Verbal Behavior, 20*, 497-514.

Hunt, R. R., & Elliot, J. M. (1980). The role of nonsemantic information in memory: Orthographic distinctiveness effects on retention. *Journal of Experimental Psychology: General, 109*, 49-74.

Hunt, R. R., & Mitchell, D. B. (1982). Independent effects of semantic and nonsemantic distinctiveness. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 8*, 81-87.

Kline, S., & Groninger, L. D. (1991). The imagery bizarreness effect as a function of sentence complexity and presentation time. *Bulletin of the Psychonomic Society, 29*, 25-27.

Kramer, J. H., Delis, D. C., Kaplan, E., O'Donnell, L., & Prifitera, A. (1997). Developmental sex differences in verbal learning. *Neuropsychology (Abstract), 11*, 577-584.

Lang, V. (1995). Relative association, interactiveness, and the bizarre imagery effect. *American Journal of Psychology, 108*, 13-35.

Marshall, P. H., Nau, K., & Chandler, C. K. (1980). A functional analysis of common and bizarre visual mediators. *Bulletin of the Psychonomic Society, 15*, 375-377.

Marschark, M., & Hunt, R. R. (1989). A reexamination of the role of imagery in learning and memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 15*, 710-720.

McCloskey, M., Wible, C., & Cohen, N. (1988). Is there a special flashbulb memory mechanism? *Journal of Experimental Psychology: General, 117*, 171-181.

McDaniel, M. A., & Einstein, G. O. (1986). Bizarre imagery as an effective memory aid: The importance of distinctiveness. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 12*, 54-65.

McDaniel, M. A., Einstein, G. O., DeLosh, E. L., May, C. P., & Brady, P. (1995). The bizarreness effect: It's not surprising, it's complex. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 21*, 422-435.

Medin, D. L., Goldstone, R. L., & Gentner, D. (1993). Respects for similarity. *Psychological Review, 100*, 254-278.

Mercer, C. (1996). The bizarre imagery effect on memory. *Journal of Mental Imagery, 20*, 141-152.

Merry, R., & Graham, N. C. (1978). Imagery bizarreness in children's recall of sentences. *British Journal of Psychology, 69*, 315-321.

Miller, R. R., Kasprow, W. J., & Schactman, T. R. (1986) Retrieval variability: Sources and consequences. *American Journal of Psychology, 99*, 145-218.

O'Brien, E. J., & Wolford, C. R. (1982). Effect of delay in testing on retention of plausible versus bizarre mental images. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 8*, 148-152.

Pezdek, K., Whetstone, T., Reynolds, K., Askari, N., & Dougherty, T. (1989). Memory for real-world scenes: The role of consistency with schema expectation. *Journal of Experimental Psychology: Learning, Memory, & Cognition, 15*, 587-595.

Pra Baldi, A., De Beni, R., Cornoldi, C., & Cavedon, A. (1985). Some conditions for the occurrence of the bizarreness effect in free recall. *British Journal of Psychology, 6*, 427-436.

Reed, P., & Richards, A. (1996). The von Restorff effect in rats (Ratus norvegicus). *Journal of Comparative Psychology, 110*, 193-198.

Reynolds, J. H. (1977) Unavailable and inaccessible information in retroactive inhibition of paired associates. *Journal of Experimental Psychology: Human Learning and Memory, 3*, 68-77.

Richardson, J. T. E. (1985). The effects of retention tests upon human learning and memory: An historical review and experimental analysis. *Educational Psychology, 5*, 85-114

Riefer, D. M., & LaMay, M. L. (1998). Memory for common and bizarre stimuli: A storage-retrieval analysis. *Psychonomic Bulletin and Review, 5*, 312-317.

Riefer, D. M., & Rouder, J. N. (1992). A multinomial modeling analysis of the mnemonic benefits of bizarre imagery. *Memory & Cognition, 20*, 601-611.

Robinson-Riegler, B., & McDaniel, M. A. (1994). Further constraints on the bizarreness effect: Elaboration at encoding. *Memory & Cognition, 22*, 702-712.

Sharpe, L., & Markham, R. (1992). The effect of distinctiveness of bizarre imagery on immediate and delayed recall. *Journal of Mental Imagery, 16*, 211-220.

Schmidt, S. R. (1985). Encoding and retrieval processes in the memory for conceptually distinctive events. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 11*, 565-578.

Schmidt, S. R. (1991). Can we have a distinctive theory of memory? *Memory & Cognition, 19*, 523-542.

Schmidt, S. R. (1996). Category typicality effects in episodic memory: Testing models of distinctiveness. *Memory & Cognition, 24*, 595-607.

Tomasulo, D. J. (1982-83). Effects of bizarre imagery on children's memory. *Imagination, Cognition, and Personality, 2*, 137-144.

Waddill, P. J., & McDaniel, M. A. (1998). Distinctiveness effects in recall: Differential processing or privileged retrieval? *Memory & Cognition, 26*, 108-120.

Webber, S. M., & Marshall, P. H. (1978). Bizarreness effects in imagery as a function of processing level and delay. *Journal of Mental Imagery, 2*, 291-300.

Winograd, E. (1981). Elaboration and distinctiveness in memory for faces. *Journal of Experimental Psychology: Human Learning and Memory, 7*, 181-190.

Wollen, K. A., & Cox, S. D. (1981). The bizarreness effect in a multitrial intentional learning task. *Bulletin of the Psychonomic Society, 18*, 296-298.

Appendix A

Item-Pair Interactions

1. Girl-Doll

Common: The girl kissed the doll.
Bizarre: The girl boiled the doll.

2. Dog-Bicycle

Common: The dog chased the bicycle.
Bizarre: The dog rode the bicycle.

3. Car-Fence

Common: The car drove past the fence.
Bizarre: The car pets the fence.

4. Goldfish-Bowl

Common: The goldfish was swimming in the bowl.
Bizarre: The goldfish was eating out of the bowl.

5. Snowflake-Mountain

Common: The snowflake fell on the mountain.
Bizarre: The snowflake climbed the mountain.

6. Shoes-Milk

Common: The shoes were placed by the milk.
Bizarre: The shoes were filled with milk.

7. Plant-Radio*

Common: The plant rested on top of the radio.
Bizarre: The plant screamed at the radio.

8. Lamp-Book

Common: The lamp shined on the book.
Bizarre: The lamp read the book.

Note.* Radio was substituted here for Riefer and Rouder's (1992) original television, as a suitable toy television could not be found for acting out the interactions.

Appendix B

Sets 1-4: Item Pairs and Corresponding Interactions

Item Pairs/Set	Set 1	Set 2	Set 3	Set 4
1. Girl-Doll	B	C	B	C
2. Dog-Bicycle	B	C	C	B
3. Car-Fence	B	C	B	C
4. Goldfish-Bowl	B	C	C	B
5. Snowflake-Mountain	C	B	B	C
6. Shoes-Milk	C	B	C	B
7. Plant-Radio	C	B	B	C
8. Lamp-Book	C	B	C	B

Note. B: bizarre interaction presented, C: common interaction presented

Appendix C

Child-Normed Results

Table C1

Summary of ANOVA and ANCOVA Results Based on Child-Normed Data

Source	Results
Errors at Acquisition (trial 1)	
Grade	$F(1, 76) = 33.46$, $p = .000$, $Ms_e = .05$
Gender	$F(1, 76) = 2.58$, $p = .112$, $Ms_e = .05$ NS
Item	$F(1, 76) = .18$, $p = .674$, $Ms_e = .06$ NS
Errors at Acquisition (total)	
Grade	$F(1, 76) = 36.13$, $p = .000$, $Ms_e = .50$
Gender	$F(1, 76) = 6.02$, $p = .016$, $Ms_e = .50$
Item	$F(1, 76) = .00$, $p = 1.000$, $Ms_e = .15$ NS
Errors at Retention (with errors on the 1st acquisition trial as a covariate)	
Covariate 1	$F(1, 75) = 1.36$, $p = .247$, $Ms_e = .12$ NS*, $R^2 = .01$
Grade	$F(1, 75) = 18.18$, $p = .000$, $Ms_e = .12$
Gender	$F(1, 75) = .07$, $p = .796$, $Ms_e = .12$ NS
Covariate 2	$F(1, 75) = .50$, $p = .482$, $Ms_e = .24$ NS*, $R^2 = .01$
Item	$F(1, 75) = 5.36$, $p = .023$, $Ms_e = .24$
Trial	$F(3, 228) = 39.56$, $p = .000$, $Ms_e = .02$
Item x Trial	$F(3, 228) = 4.11$, $p = .007$, $Ms_e = .03$

Table C1 - continued

Errors at Retention (with total errors at acquisition as a covariate)	
Covariate 1	$F(1, 75) = 2.85$, $p = .096$, $Ms_e = .12$ NS*, $R^2 = .03$
Grade	$F(1, 75) = 15.53$, $p = .000$, $Ms_e = .12$
Gender	$F(1, 75) = .00$, $p = .993$, $Ms_e = .12$ NS
Covariate 2	$F(1, 75) = 2.46$, $p = .124$, $Ms_e = .24$ NS*, $R^2 = .03$
Item	$F(1, 75) = 5.35$, $p = .024$, $Ms_e = .24$
Trial	$F(1, 75) = 39.56$, $p = .000$, $Ms_e = .02$
Item x Trial	$F(3, 228) = 4.11$, $p = .007$, $Ms_e = .03$

Note. * although the covariate was not significant, scores were still adjusted in the retention analyses

Table C2

Mean Errors and Standard Deviations for Child-Normed Data on Trial 1 and across all Trials at Acquisition

	Trial 1		All Trials	
	Ks x(sd)	**2s x(sd)**	**Ks x(sd)**	**2s x(sd)**
Males	1.12(.38)	.67(.27)	2.47(1.11)	.92(.46)
Females	1.18(.27)	.84(.28)	2.84(1.08)	1.66(1.06)

Note. Because there was no item effect, mean errors were collapsed across conditions

Table C3

Mean Proportion (SDs) of Errors per Trial at Retention for Child-Normed Data

Grade/Item Type	Trial			
	1 x(sd)	2 x(sd)	3 x(sd)	4 x(sd)
K				
Common	.58(.28)	.54(.32)	.49(.33)	.48(.31)
Bizarre	.55(.25)	.45(.20)	.36(.18)	.33(.21)
2				
Common	.42(.29)	.39(.28)	.27(.22)	.35(.29)
Bizarre	.39(.23)	.28(.18)	.23(.19)	.17(.15)

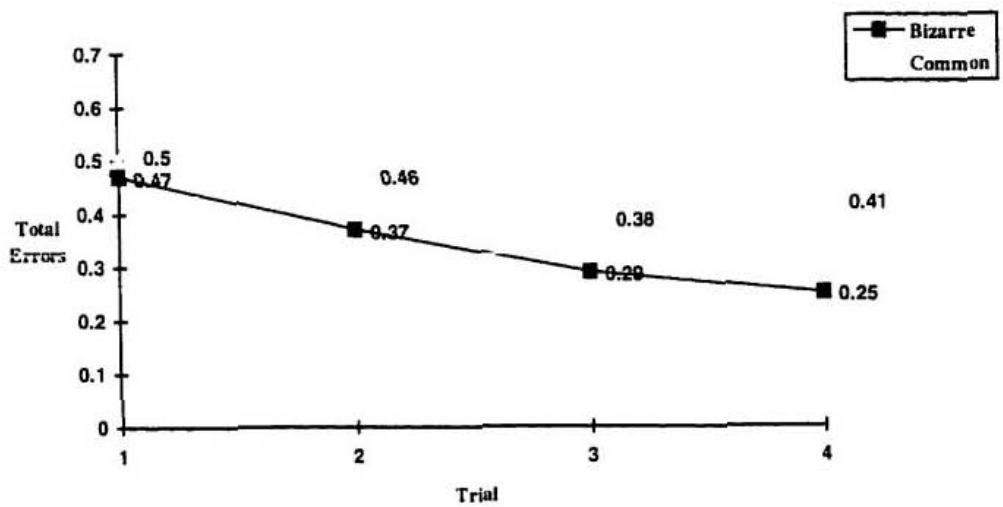

Figure C1. Mean proportion errors across retention trials for child-normed bizarre and common materials (collapsed across age).