Journal of Memory and Language 60 (2009) 229–251

Contents lists available at ScienceDirect

Journal of Memory and Language

journal homepage: www.elsevier.com/locate/jml

An associative-activation theory of children's and adults' memory illusions

Mark L. Howe [a,*], Marina C. Wimmer [a], Nadine Gagnon [b], Shannon Plumpton [c]

[a] Department of Psychology, Lancaster University, Center for Research in Human Development, Lancaster LA1 4YF, United Kingdom
[b] University of British Columbia, Vancouver, BC, Canada
[c] Simon Fraser University, Vancouver, BC, Canada

ARTICLE INFO

Article history:
Received 8 April 2008
Revision received 14 October 2008
Available online 25 November 2008

Keywords:
DRM paradigm
False memories
Associative-activation theory
Memory development
Children's false memory
Semantic density
Fuzzy-trace theory

ABSTRACT

The effects of associative strength and gist relations on rates of children's and adults' true and false memories were examined in three experiments. Children aged 5–11 and university-aged adults participated in a standard Deese/Roediger–McDermott false memory task using DRM and category lists in two experiments and in the third, children memorized lists that differed in associative strength and semantic cohesion. In the first two experiments, half of the participants were primed before list presentation with gist-relevant cues and the results showed that: (1) both true and false memories increased with age, (2) true recall was higher than false recall for all ages, (3) at all ages, false memory rates were determined by backward associative strength, and (4) false memories varied predictably with changes in associative strength but were unaffected by gist manipulations (category structure or gist priming). In the third experiment, both gist and associative strength were varied orthogonally and the results showed that regardless of age, children's (5) true recall was affected by gist manipulations (semantic cohesion) and (6) false recall was affected by backward associative strength. These findings are discussed in the context of models of false memory illusions and continuities in memory development more generally.

Introduction

Recent research has shown that children are less susceptible to false memories than adults especially when those false memories are generated spontaneously as in the Deese/Roediger–McDermott (DRM; Deese, 1959; Roediger & McDermott, 1995) paradigm (Brainerd, Reyna, & Forrest, 2002; Howe, 2005, 2006, 2008a; Howe, Cichetti, Toth, & Cerrito, 2004; Howe, Gagnon, & Thouas, 2008). Briefly, the DRM task consists of presenting lists of words that are related to an unpresented concept (e.g., *hot*, *snow*, *warm*, *winter*, *ice*, *wet*, *frigid*, *chilly*, *heat*, *weather*, *freeze*, and *air* are related to the unpresented item COLD). The unpresented item, also termed the critical lure, is often reported along with presented list items during subsequent recall or recognition tasks.

Although the false-memory illusion arising from the DRM paradigm is a robust phenomenon, it is not clear what mechanisms underlie children's false memories and why false memories increase across childhood. From the adult literature it is generally suggested that the items on the DRM list directly activate related but unpresented concepts during study and this happens automatically outside of conscious awareness. When it is time to recall or recognize the list items, the unpresented but activated concept is also falsely output because participants cannot discriminate them from the original list members.

Theoretical explanations of this effect, such as fuzzy-trace theory (FTT), have attempted to account for these findings by suggesting that there are two memory traces encoded during list presentation—verbatim and gist (Brainerd & Reyna, 2005). These traces are qualitatively different from each other: verbatim traces are concerned with item-specific surface information (e.g., the number of phonemes in a word) and gist traces with meaning-based information (i.e., the theme of a word list). It is this gist

* Corresponding author.
E-mail address: mark.howe@lancaster.ac.uk (M.L. Howe).

0749-596X/$ - see front matter © 2008 Elsevier Inc. All rights reserved.
doi:10.1016/j.jml.2008.10.002

trace that is thought to be responsible for false recall in the DRM paradigm particularly when verbatim traces (which fade more rapidly) are not available. Gist traces are particularly relevant in recall rather than recognition because the physical presence of the item in the recognition test is more likely to activate its verbatim trace. Although gist is not clearly defined in FTT (e.g., Howe, 2008b), in general it concerns the abstraction of a common "theme" contained within the presented material (e.g., see Brainerd, Reyna, & Ceci, 2008). Thus, for example, in the DRM paradigm, the gist is the theme of the list. When categorized or taxonomic items are used, the category itself serves as the gist. Not only do word lists contain gist, but so too do single words—that is, individual words can have their own unique meaning or local gist that can cause false memories (e.g., Reyna & Lloyd, 1997). It is the extraction of this gist that underlies false memory production, particularly when verbatim traces have faded. Developmentally, although young children are capable of extracting the gist, the ability to do so improves with age, especially their ability to extract thematic meaning across list items (Brainerd & Reyna, 2005). As this ability increases with age, so does the child's susceptibility to the DRM illusion.

An alternative explanation, associative-activation theory (AAT), states that children's false memories are a product of associative-activation processes (Howe, 2005, 2006, 2008b). Associative relations between list items and the critical lure are important contributors to the false memory illusion (Deese, 1959; McEvoy, Nelson, & Komatsu, 1999; Roediger, Watson, McDermott, & Gallo, 2001b; Underwood, 1965). This theory (as well as activation-monitoring theory, AMT, Roediger, Balota, & Watson, 2001a) derives from spreading activation models that suggest that the processing of one word activates a corresponding node (i.e., concept) in our mental lexicon (or more generally, our knowledge base) and this activation spreads to surrounding concept nodes (Collins & Loftus, 1975; Landauer & Dumais, 1997). In particular, one word (or concept) activates another and some of these activated concepts are items that have not been presented but have been "incorrectly" activated due to their connection with presented items in the knowledge base. Howe (2005, 2006, 2008b; Howe et al., 2008) suggests that increases in children's false memory with age may be due to changes in children's knowledge base (e.g., accretion of associations, reorganization of knowledge). This, combined with additional experience and practice utilizing concepts results in increases in the automaticity with which children access or activate associations in their knowledge base, including associations that mediate false remembering.

According to the AAT (Howe, 2005, 2006, 2008b; Howe et al., 2008), children's false memories, like those of adults, should increase as a function of increasing associative strength between list items and the critical lure. However, it is unlikely that young children's rates of false memories will be similar to those of older children and adults because concepts and the associative relations among them are not activated as automatically as they are for adults. As demonstrated by studies of memory organization, the quantity and quality of associative relations between old and newly acquired concepts continue to undergo significant changes with development (Bjorklund, 1987, 2004). As children gain knowledge and experience through exposure and development, their conceptual representations and the associative links among related concepts become better integrated in memory (Bjorklund,1987, 2004; Howe, 2000). Although these changes in the organization of conceptual representations strengthen the links between related concepts, children's ability to automatically process these relations and activate related concepts will depend on their developing cognitive abilities.

The production of false memories in the DRM paradigm is relatively automatic for adults, for whom associated but unpresented items, such as the critical lure, are primed through their associative links with the presented list items. Children's ability to process and activate these items through their associative links with list items is not as automatic (for a review, see Howe, 2000). The activation of these associative links is effortful for young children, but as children get older, less cognitive effort is required to process and activate associative links between related concepts, which in turn increases their susceptibility to the DRM false memory illusion (Howe, 2005; Metzger et al., 2008). For example, Howe (2005) found that young children are able to inhibit false memories suggesting that they are less automatic *per se*. In this study children were presented with two DRM lists. After the first list had been presented children either received remember instructions ("keep remembering the words on this list") or forget instructions ("this was just a practice list, you can forget it") and then told to remember the second list. A third control group received no instructions and served as a baseline for true and false memory rates. The results showed that children in the forgetting condition had significantly fewer false memories than children in the control condition whereas children in the remember condition had significantly more false memories than children in the control condition. If false memories occurred automatically at the time of encoding or retrieval, as they do in adults (see Kimball & Bjork, 2002), inhibition would have been extremely difficult. The fact that children were able to inhibit false memories suggests that they are less automatic than for adults. Therefore, consistent with the predictions of AAT, it is likely that as the automaticity of associative-activation increases with age so too does children's susceptibility to false memory illusions.

FTT contains no assumptions about the role of consciousness and automaticity in the development of the false memory illusion (Brainerd, Forrest, Karibian, & Reyna, 2006; Brainerd & Reyna, 2005). However, FTT differs from AAT on other important dimensions, ones in which distinct predictions can be derived and tested. One critical difference between AAT and FTT is that according to AAT, activation spreads from one word or concept representation to another whereas according to FTT it spreads from a word or list of words to an extracted gist (e.g., thematic representation) to additional representations consistent with the extracted gist. In other words, in AAT concepts directly activate each other in the knowledge base whereas in FTT there is an additional, mediated step involving a separate thematic representation that must be extracted prior to the activation of additional, nonpresented words or concepts. Unlike FTT, AAT does not require the separate activation of a theme or gist trace in order for false memories to occur.

What does the empirical research say about these two theories? Because gist effects are said to be more likely using recall than recognition measures (e.g., Brainerd et al., 2008), we focus primarily on studies that report recall data. First, consider studies on gist-cuing manipulations. Here, experiments have shown that when gist is manipulated in children's recall, false memory rates can be elevated, a finding that on the surface supports FTT (e.g., Brainerd et al., 2006; Holliday, Reyna, & Brainerd, 2008; Odegard, Holliday, Brainerd, & Reyna, 2008). For example, 11-year-olds were more likely to produce false memories when they received additional cue words at study that biased the context toward, rather than away, from the gist of a list (Odegard et al., 2008). However, the interpretation of this outcome is problematic for a number of reasons. For example, because there was no control group that did not receive any gist manipulation, it is difficult to say whether the toward-condition elevated false memories, the away-condition lowered false memories, or both. Moreover, the manipulation itself may not have been purely gist-based inasmuch as the words used as additional cues to meaning may simply have served to increase the associative-activation of the unpresented critical lure.

In another study, false memory rates increased when children were told beforehand that they would be presented with words that have similar meanings. This effect was larger for younger (6- and 11-year-olds) than for older (14-year-olds) children (Brainerd et al., 2006). This is what FTT would anticipate because younger children have more problems extracting gist than older children and, therefore, gist instructions should have a more pronounced effect at a younger age. It is not clear, however, how instructing children to look for something they cannot easily extract (i.e., gist) should increase their ability to do so. Alternatively, reminding children to look for meaningful relations in studied material is but one metamemory instruction that may affect children's ability to search for interitem associative connections, connections they already possess but may fail to activate automatically (e.g., see Schneider & Pressley, 1997).

Other studies have found that gist-like manipulations affect children's false memory rates, although these effects do not interact with age. For example, Holliday et al. (2008) found that, contrary to their expectation, false recall rates increased similarly for 7- to 13-year-olds when specific thematic labels were given before list presentation (e.g., "the words you will hear are all related in meaning, they are all medical words"). Howe (2006) found that presenting a label (e.g., animals) prior to categorized (taxonomic) lists did not increase false recall for 5- to 11-year-olds. Overall, then, the empirical evidence from gist-cuing studies reporting recall data is inconsistent and difficult to interpret due to methodological differences (e.g., the use of DRM versus category lists) across studies.

Second, the AAT interpretation is favored by dissociations of true and false recall. Hutchison and Balota (2005) with adults and Howe (2006) with children found that within-list thematic relations affected true but not false memories and that associative relations influenced both, but had a greater effect on false memories. Park, Shobe, and Kihlstrom (2005) also demonstrated that categorical relations elicited fewer false memories than associative relations in studies of adults' true and false recall. Moreover, children's true recall increases when categorical lists are presented in contrast to DRM lists, but there are no differences for false recall (Howe, 2006). Thus, contrary to predictions from gist-based theories, neither adults' nor children's false memories increase when provided with lists that emphasize a single, thematic relationship. Concerning manipulations of associative relations, lists that are high in backward associative strength produce more false recall than lists low in backward associative strength, but these lists do not exhibit different rates of true recall (e.g., Howe et al., 2004). Moreover, Hege and Dodson (2004) with adults and Howe (2006) with children have shown that when associative processing is significantly reduced but semantic processing remains intact (e.g., using pictorial stimuli rather than verbal stimuli; also see Chiarello, 1998), true recall rates are unaffected but false recall rates are significantly reduced and age effects in false recall rates disappear (also see Howe, 2008a).

Third, and perhaps the strongest evidence for an associative-activation model comes from research, already mentioned, that has shown that associative strength is crucial for false memory production. The strength of the associative links between list items (e.g., *bed*) and the critical lure (e.g., SLEEP), also termed backward associative strength (BAS), is indicative of a list's ability to produce false memories in children and adults. That is, lists with very little mean BAS yield low rates of false memories (Gallo & Roediger, 2002; Howe et al., 2004). Similarly, recent findings have indicated that the strength of associative links between the critical lure (e.g., SLEEP) and the list items (e.g., *bed*), also referred to as forward associative strength (FAS), can influence the production of false memory in adults, although these studies only examined recognition (Brainerd & Wright, 2005).

Strictly speaking, the finding that associative strength between list items and the critical lure, whether forward or backward, plays a key role in adults' production of false memories is difficult for gist-based theories to accommodate. For example, because BAS is a measure between list items and their *direct* activation of the nonpresented critical lure, it is unclear what advantage accrues to such models by imposing the added step of having to first extract the overall gist of the list before false memories are produced. If BAS is a measure of the extent to which a presented item directly activates nonpresented but related concepts, what advantage is associated with imposing an intermediary step (gist extraction) that makes this relationship indirect? On the face of it, this seems to be a more complex explanation than is demanded by the data.

Alternatively, gist-based theories could always counter with the assumption that BAS and FAS are proxy variables for gist. Although it is clear that global measures such as BAS and FAS can be further delineated (e.g., broken down into different meaning components; see Brainerd, Yang, Reyna, Howe, & Mills, in press; Roediger et al., 2001b), such arguments may still be problematic because associations are typically viewed in the broadest sense, referring to any set of concepts that can be mentally linked (e.g., contiguity, co-occurrence, taxonomic membership, conceptual, orthographic, or phonemic similarity, and so forth) (see Gallo, 2006; Howe, 2008b). Gist, on the other hand, is often viewed as being more restricted, referring instead to a theme "... that subjects can form (and remember) to summarize or categorize the words" (Gallo, 2006, p. 40). Thus, in contemporary models, gist is viewed as a subset of the more general class of relationships between concepts denoted by the term association. Even if gist was to be broadened to encompass what has traditionally been the province of associationism for well over a century and a half (e.g., Hamilton, 1859), such theories can still founder due to their insistence that the additional step of gist extraction is necessary for the occurrence of false memories.

Howe (2005, 2006, 2008b; Howe et al., 2008) has suggested that these same associative mechanisms that account for adults' false recall can also be used to understand children's false memory illusions. Although there is some support for this proposition, experiments that have directly and systematically compared the impact of variation in FAS or BAS on true and false rates have yet to be conducted with children, something that is one focus of the current research. Two studies have shown that variation in BAS does affect children's false recall and that these effects are more pronounced for older than younger children (e.g., Brainerd et al., 2006; Howe et al., 2004). These and other recent findings suggest that developmental trends in false memory illusions are related to increases in the activation of associative links that may already exist between concepts rather than children's growing ability to extract meaning across list items (i.e., gist-across-the-list).

As noted, associations are multidimensional and consist of many different stimulus elements that overlap (e.g., phonological, orthographic, temporal contiguity, spatial proximity, superordinate relations, property relations, physical or conceptual similarity; also see Brainerd et al., in press; Roediger et al., 2001b). Associative strength is determined by the relationship between items that exists on the basis of any one dimension of similarity or may involve many dimensions that overlap. Research on children's memory organization has confirmed that associative links between concepts appear at a relatively young age and are further strengthened and defined with increases in both knowledge and experience (Bjorklund, 1987; Frankel & Rollins, 1985). It has also been established that children can use associative relations very early in life, although the nature and type of association changes with age and experience (e.g., Lange, 1978). Research on the development of conceptual organization in memory has indicated that both younger and older children rely on associative (including categorical) links between concepts to guide the recollection process (Bjorklund, 2004). With development children gain more experience and knowledge and their ability to automatically activate and use associative relations to guide recollection improves (Bjorklund, 2004; Howe et al., 2008). Given that conditions are made favorable, even very young children can activate associative (including categorical) relations and use them to mediate recollection (e.g., Frankel & Rollins, 1985). For example, increasing the salience of categorical information by blocking similar category items together can facilitate the processing of such relations for young children (Howe, Brainerd, & Kingma, 1985). Of course blocking information by category is the same as presenting a DRM list—all items are related to a prototype (category lists) or the one associative item (DRM lists).

Research on children's false memories has shown that children as young as five years old, like older children and adults, can spontaneously extract categorical relations across word lists (Howe, 2006). In this study, young children's rates of true and false recall were not influenced by the provision of superordinate category cues prior to list presentation. Despite their ability to spontaneously extract the thematic relationship across lists' items (i.e., category lists were better recalled than DRM lists), young children's rates of true and false recall remained lower than that of older children and adults. Indeed, providing gist-consistent descriptive cues did not increase children's false recall. Even when categorical (e.g., *banana, pear, orange*) and DRM items (e.g., *core, red, pie*) were used to cue the same critical lure (e.g., APPLE), children's false recall rates are controlled by variations in associative strength, not the availability of an overarching, thematic cue (Howe, Wimmer, & Blease, in press). A tentative conclusion from research like this is that children's ability to extract meaning across word lists is unrelated to their lower susceptibility to false memory illusions. Instead, as proposed by theories relying on associative-activation, children's false memories, like adults', result from the activation of associative relations between list items and the critical lure.

The role of associative strength in the production of false memories has mostly focused on the influence of list items' backward associative links to the critical lure and has been limited to adults. Studies with adults have indicated that variations in a list's total, rather than mean, BAS can directly influence the occurrence of false memories in the DRM paradigm (Robinson & Roediger, 1997). More specifically, Robinson and Roediger (1997) found that false recall increased in parallel with the addition of related items to increase both list length and total associative strength. On the other hand, adding unrelated associates to keep list length constant did not change rates of false recall, but rather was limited to decreasing rates of true recall. Given these results, it appears that a list's total, rather than mean, BAS influences false recall. Veridical recall, on the other hand, is diminished when unrelated exemplars disrupt the associative coherence of a list. Again, the effects of associative relations between list items and the critical lure appear to be limited to false memories.

Unlike BAS, FAS has not been significantly associated with the production of false memories during free recall (although this may be an artifact of the restricted range or variability of measures of FAS). Indeed, previous studies on false recall have failed to find a significant relationship between FAS and false memories (see McEvoy et al., 1999; Roediger et al., 2001b). However, recent findings have indicated that, when manipulated within lists mean forward associative strength (MFAS), even when limited in range, can influence false memory production in adults. Using eight 4-item lists with varying BAS and FAS, Brainerd and Wright (2005) found that lists with high FAS or BAS produced higher rates of false recognition. More specifically, lists for which either FAS or BAS or both were high produced the highest proportions of false recognition. It appears that at least for recognition, adults' false memories can be influenced by the lure's associative links to the list items.

Overall, then, the effects of associative strength appear to be stronger when it comes to the production of false recall, as true recall is often less affected by the strength or direction of associative links between list items and the critical lure. In addition, true and false recall are affected differently by gist-like manipulations (e.g., ones that emphasize vertical connectivity such as categorical or thematic structure) and associative coherence. Categorical or thematic coherence of word lists has been found to influence rates of true but not false recall in both children (Howe, 2006) and adults (Hutchison & Balota, 2005). In fact, Hutchison and Balota (2005) found that increasing the number of meaning-consistent items in word lists increased accurate recall without affecting rates of false recall. This was also found with children, for whom category lists produced better true recall than DRM lists, but there was no effect on children's false memories (Howe, 2006; Howe et al., in press).

Although it is known that associative strength plays an important role in the production of adult's false memories, few studies have examined these effects with children. The goal of this article is to provide more definitive evidence of the primacy of associative-activation in the development of false memory illusions, linking children's memory development in this domain to that observed with adults. Although such memory continuity effects have been observed in the development of accurate remembering (for a review, see Howe, 2000), the same has not been true for false memory illusions. In what follows, we present three experiments that examine the role of associative strength in children's and adults' false recall patterns. Across the experiments we seek evidence of the continuity of memory development. Along the way, we provide a number of tests of FTT's and AAT's predictions about developmental patterns of false memory illusions. We begin by describing the first two experiments, ones that provide tests of explicit manipulations of gist (use of category lists and labels) and associative strength (different variations in associative strength). We then describe a third experiment, one in which gist manipulations are more implicit (variation in the semantic cohesion or density of the list) and are directly contrasted with variation in associative strength. To anticipate, the findings (a) are more consistent with AAT's predictions of developmental patterns in false memory illusions than FTT's and (b) provide strong evidence of developmental continuity in the emergence of false memory illusions, consistent with that known to exist for the development of accurate remembering.

Overview of experiments 1 and 2

As seen, developmental increases in false memory may be related to the development of automaticity in associative-activation (AAT) rather than an inability to extract gist across list items (FTT). In the first two experiments, we examine this more closely by varying associative strength. In the first experiment, we varied both BAS and FAS. In the second experiment, we manipulated associative strength by removing either the strongest BAS items from each list or several weaker ones that reduced list BAS by the same amount as removing the strongest associate. In order to derive different predictions for FTT and AAT, associative strength was varied across both categorical/thematic lists and DRM lists. Although previous findings on the dissociation of true and false memories have shown that categorical/thematic information primarily affects true recall (Howe, 2006; Hutchison & Balota, 2005), these effects have been demonstrated using a between-subjects design. In order to examine this dissociation with more precision, we examined the effects of category and associative (DRM) lists using a within-subject design.

Both FTT and AAT predict that (a) true and false recall should increase with age and (b) children's and adults' true memories should be greater for category than for DRM lists. However, the two theories differ when it comes to false memory production. The key prediction from AAT is that false memory should increase with increases in the list's BAS independently of the greater gist salience for category than DRM lists. Thus, lists with high FAS and BAS should generate more false memories than lists with lower FAS and BAS (Experiment 1) and lists with higher BAS, regardless of whether only a single strong item is removed or several weak items are eliminated, should generate more false memories than lists with lower BAS (Experiment 2). FTT, on the other hand, predicts the same outcome for false memory as it does for true memory. That is, to the extent that categorical lists contain a single gist, thereby increasing the probability that its gist will be more easily discerned, category lists should produce higher false memory rates than DRM lists which can contain multiple "gists" or associative relations. FTT also predicts a developmental interaction. Because older children and adults are more likely to apprehend these implicit cues, these effects should be more pronounced with younger children. To our knowledge, this will be the first time that the role of FAS and BAS (Experiment 1) and overall list strength (Experiment 2) on false memories in categorized and DRM lists has been examined developmentally.

A second manipulation was used to discriminate FTT and AAT, namely, the provision of explicit cues (labels) prior to list presentation, something that should prime across-the-list gist. This should be the optimum time to present gist cues as it has been shown that, at least for adults, encoding is a primary locus of false memory formation for both category and DRM lists (Dewhurst, Bould, Knott, & Thorley, 2009). For category lists, the category label was given prior to list presentation to cue the vertical gist (e.g., "fruit" was presented before reading the list "orange, banana, ..."). For DRM lists, a word or phrase describing what was common across the list was presented (e.g., for the "doctor" list, the cue was "medical things"). As noted earlier, AAT predicts that these labels should only increase false recall rates if they change the overall activation of the critical lure (i.e., that they too are associatively related to the critical lure). FTT, on the other hand, predicts that these labels should increase false recall rates, particularly in younger children who are said to be less able to spontaneously extract gist-across-the-list.

Overall, then, according to the AAT, associative-activation is the crucial mechanism that accounts for children's false memories and therefore false recall rates should be directly related to fluctuations in associative strength. For FTT, although false recall rates should increase with age, gist manipulations should increase false recall rates in general, but also exhibit two different developmental effects. (1) Category lists should yield higher false recall rates than DRM lists, particularly for younger children who are less able to extract gist spontaneously. This is because category lists contain a single, dominant thematic connectivity among items, something not always possessed by DRM lists. (2) Explicit gist cues at encoding should also increase false recall rates for both category and DRM lists, but particularly for younger children who are not as likely to extract gist spontaneously.

Experiment 1

In this experiment, BAS and FAS were varied systematically across both category and DRM lists, allowing us to examine the role of associative relations in the development of false memories independently of across-the-list-gist relations. The role of BAS in false memory production has been well established in adult research but the role of FAS is less clear. The few previous studies with adults have found small significant effects for FAS that are much smaller than BAS effects, but these findings have been usually limited to measures of recognition. Given that gist plays a more important role in recall, it is important to examine the relative influence of BAS and FAS using this more gist-sensitive response measure. Moreover, the role of FAS in relation to BAS has never been investigated in children's false memories and it is therefore crucial to explore this further. Although our implementation of this factor in this design was purely exploratory, it was nonetheless expected that FAS would influence rates of false memories. However, this effect may be limited to cases where explicit cues were provided, particularly for children. That is, although descriptive labels failed to increase rates of true and false recall in children (Howe, 2006), levels of FAS were not controlled in those studies. Results from research with adults has indicated that for category lists, output frequency of the list items during the production of category norms did influence rates of false recall when a descriptive cue was provided (Smith, Ward, Tindell, Sifonis, & Wilkenfeld, 2000). Thus, although it was expected that young children are capable of extracting gist across the list without explicit cues, descriptive labels of category lists, through their forward associative links to the critical lure, may be more effective in priming their list's top category exemplars (the critical, unpresented item), resulting in higher false memory rates for participants provided with descriptive labels. A similar effect might occur for DRM lists. More specifically, descriptive labels could prime the DRM list items as well as the critical lure in instances where backward (or forward) associative links were weak.

Method

Participants

A total of 250 children [48% male; 86 5-year-olds ($M = 5.5$, $SD = 5$ months), 80 7-year-olds ($M = 7.6$, $SD = 6$ months), and 84 11-year-olds ($M = 11.5$, $SD = 4$ months)] were recruited. All children (predominantly White and middle-class) were tested following parental consent and their own assent on the day of testing. A sample of 60 university students ($M = 18.6$, $SD = 9$ months) was also recruited and tested following informed consent. Participants were randomly assigned to the label and no label conditions, with half of the participants per age group (50% male) assigned to each condition.

Design, materials, and procedure

A 2(List: DRM vs. Category) × 3(Strength: High BAS-High FAS vs. High BAS-Low FAS vs. Low BAS-High FAS) × 4(Age: 5- vs. 7- vs. 11- vs. 18-year-olds) × 2(Labels: No labels vs. Labels) design was used where list and associative strength were manipulated within subject and age and labels were between-subjects variables. (Note that a Low BAS-Low FAS condition was not included because such conditions routinely produce floor effects in false memory rates).

Participants were tested individually in a quiet room. Twelve 14-item lists (6 DRM lists and 6 category lists; see Appendix A) were presented to each participant. Associative strength varied between lists with each participant being exposed to four High BAS-High FAS lists (2 DRM and 2 category), four High BAS-Low FAS lists (2 DRM and 2 category), and four Low BAS-High FAS DRM lists (2 DRM and 2 category). Mean word frequencies did not differ for list items, $t(10) = 1.56$, $p = .15$, or for critical lures between DRM and category lists, $t(10) = -1.41$, $p = .19$. Neither did mean word frequency differ for list items, $F(2, 11) = 1.91$, $p = .20$, or for the critical lures between the three conditions (High/High vs. High/Low vs. Low/High), $F(2, 11) = 1.18$, $p = .35$ (word frequencies were taken from Stuart, Masterson, Dixon, & Quinlan (1993-1996)).

Lists were read aloud by the experimenter at a 3-second rate per item. Half of the participants in each age group received a descriptive label prior to hearing each list whereas the other half did not. Descriptive labels for category lists consisted of the category name included in the norms (Van Overschelde, Rawson, & Dunlosky, 2004). Labels for the DRM lists (see Appendix A) were constructed to highlight the shared semantic relations across list items and were drawn from Roediger et al., 2001b. All lists have been used previously with children and adults in these age ranges (e.g., Brainerd et al., 2002; Howe, 2005, 2006, 2008a; Howe et al., 2008; Roediger et al., 2001b; Seamon, Luo, Schlegel, Greene, & Goldberg, 2000). In order to connect the present research with previous studies investigating label effects, it was also necessary to use labels that have already been implemented in prior research, ones that have been shown to affect false memory rates (e.g., Holliday et al., 2008). Finally, as in previous studies, the critical target for category lists consisted of the highest frequency word (most typical exemplar) for the selected category. DRM and category lists were selected on the basis of mean[1] BAS and FAS as determined using the University of South Florida norms (Nelson, McEvoy, & Schreiber, 1999). In cases where associative strength was not available, the Edinburgh Associative Thesaurus (EAT) was consulted (Kiss, Armstrong, Milroy, & Piper, 1973). DRM lists were also selected on the basis of mean BAS and FAS as indicated in the Roediger et al., 2001b norms. Mean FAS and BAS are shown in Table 1.

Prior to the presentation of the first list, participants were given general memory instructions to listen carefully to the words on the list as they would be asked to recall the words later. Following these instructions, participants in the label condition were given a descriptive label describing the theme of the list prior to list presentation. Participants in the no label condition simply received the word list. Following list presentation, participants were given a 30-second distractor task (circling randomized pairs of letters) prior to recall. Once recall was complete, the next cycle began. This procedure was repeated for the remaining 11 lists. Items were presented in the usual descending order of BAS. All presentation and recall was oral and list order was randomized within each age group.

Results and discussion

Preliminary analyses indicated no main effect or interactions involving gender so this variable was eliminated from subsequent analyses. Consistent with the hypotheses for this study, rates of true and false recall will be analyzed separately. Throughout this article, effect sizes (η^2) refer to partial eta[2].

[1] Although *total* backward associative strength is often preferred over *mean* backward associative strength, we used mean strength in order to provide a correspondence to the metric preferred for variation in forward associative strength. Because in our case mean and total backward associative strength were correlated, using the mean as the metric for backward associative strength did not serve to compromise the findings.

[2] Spreading activation models (e.g., Anderson, 1983; Anderson & Bower, 1973; Collins & Loftus, 1975) assume that presented items activate their memory representations and this activation spreads to other words via links throughout an associative network. Because list items are associated with the unpresented critical lure, this lure is activated numerous times (11 to 15 times in our experiments, depending on the list being studied) and is later falsely included in the recall set. Current models of false memory production in children and adults, while still relying on associative connectivity, contain more complicated, multiplicative encoding and retrieval rules than those of earlier models (e.g., see Kimball, Smith, & Kahana, 2007). Although the details of this model are not critical for current purposes, it is important to note that unlike earlier models in which associative strength between items and critical lures was the key determinant of false recall rates, these newer fSAM models also place importance on associations among studied items (intralist connectivity). Because of the importance of intralist connectivity, variation in strength across all of the items on the list, not individual items, becomes more central in determining false recall rates in these models. Thus, like the associative-activation theory, newer spreading-activation models would also predict no differences in false recall rates as a function of whether one strong item or several weaker items were removed from the list, just the overall associative strength of the list.

Table 1
Mean backward and forward associative strength for each variation of category and DRM lists.

Associative strength	Category		DRM	
	MBAS	MFAS	MBAS	MFAS
High BAS–high FAS	.150	.040	.234	.062
High BAS–low FAS	.080	.020	.235	.028
Low BAS–high FAS	.070	.030	.010	.061

Note. BAS = backward associative strength; FAS = forward associative strength; MBAS = mean backward associative strength of a list; MFAS = mean forward associative strength of a list.

True recall

The proportion of items correctly recalled was analyzed using a 2(List: DRM vs. category) × 3(Strength: High BAS-High FAS vs. High BAS-Low FAS vs. Low BAS-High FAS) × 2(Label: No label vs. label) × 4(Age: 5-, 7-, 11-, and 18-year-olds) analysis of variance (ANOVA). The results showed a main effect for list, $F(1, 302) = 242.23$, $p < .001$, $\eta^2 = .45$, where the mean proportion of true recall for category lists ($M = .44$) was greater than that for DRM lists ($M = .35$). There was also a main effect for age, $F(3, 302) = 487.47$, $p < .001$, $\eta^2 = .83$, where post-hoc Newman–Keuls tests ($p < .05$) revealed that 5-year-olds ($M = .18$) recalled less than 7-year-olds ($M = .30$) who recalled less than 11-year-olds ($M = .43$) who recalled less than 18-year-olds ($M = .67$). Finally, there was a main effect for strength, $F(2, 604) = 49.71$, $p < .001$, $\eta^2 = .14$, where post-hoc tests showed that BAS was the critical determinant for true recall, with High BAS-High FAS lists ($M = .41$) and High BAS-Low FAS lists ($M = .41$) yielding more true recall than Low BAS-High FAS lists ($M = .37$).

In addition, four first-order interactions emerged, three of which involved age. The one interaction that did not involve age was between List × Strength, $F(2, 604) = 42.63$, $p < .001$, $\eta^2 = .12$, where post-hoc tests confirmed that although High BAS yielded the best recall for both DRM and category lists, and that category lists were better recalled than DRM lists, each list type had a unique ordering of strength relations. That is, for DRM lists, the ordering was High BAS-High FAS ($M = .38$) > High BAS-Low FAS ($M = .34$) > Low BAS-High FAS ($M = .33$), whereas for the category lists the ordering was High BAS-Low FAS ($M = .48$) > High BAS-High FAS ($M = .43$) > Low BAS-High FAS ($M = .40$).

The three interactions involving age were as follows. First, there was an Age × List interaction, $F(3, 302) = 9.56$, $p < .001$, $\eta^2 = .09$. As can be seen in the left panel of Table 2, and was confirmed by post-hoc tests, although category

Table 2
Mean proportion of true and false recall as a function of age and list in Experiment 1 (standard errors in parentheses).

Age	True recall		False recall	
	DRM lists	Category lists	DRM lists	Category lists
5-year-olds	.14 (.009)	.22 (.010)	.10 (.020)	.17 (.023)
7-year-olds	.24 (.010)	.36 (.011)	.20 (.020)	.30 (.024)
11-year-olds	.37 (.010)	.48 (.011)	.30 (.020)	.37 (.024)
18-year-olds	.65 (.011)	.69 (.013)	.44 (.023)	.38 (.028)

lists were better recalled than DRM lists at all ages, this gap increased with age up until 11-year-olds and then decreased with adults. Second, there was an Age × Strength interaction, $F(6, 604) = 22.73$, $p < .001$, $\eta^2 = .18$. As can be seen in the left panel of Table 3, and was confirmed by post-hoc tests, there were no differences in true recall due to strength for 5- and 7-year-olds but the two High BAS lists were better recalled than the Low BAS list for 11- and 18-year-olds. Third and last, there was an Age × Label interaction, $F(3, 302) = 5.02$, $p < .01$, $\eta^2 = .05$. As can be seen in Table 4 and was confirmed by post-hoc tests, label conditions were better than no labels conditions for all child age groups but there were no differences for adults.

Although some of these patterns for FAS (Brainerd & Wright, 2005) and category versus DRM lists (Howe, 2006; Seamon et al., 2000) have been observed in prior studies, many have not been previously reported. For example, all age groups tended to recall more targets for category than DRM lists but only older children and adults true recall was affected by variation in associative strength. Interestingly, descriptive labels aided all children's true recall by approximately the same magnitude (also see Howe, 2006), but had no effect on adults' true recall. A full discussion of these trends will be deferred until after the false recall findings are presented.

False recall

The proportion of critical lures falsely recalled was analyzed using a 2(List: DRM vs. category) × 3(Strength: High BAS-High FAS vs. High BAS-Low FAS vs. Low BAS-High FAS) × 2(Label: No label vs. label) × 4(Age: 5-, 7-, 11-, and 18-year-olds) ANOVA. The results showed a main effect for list, $F(1, 302) = 10.29$, $p < .001$, $\eta^2 = .03$, where the mean proportion of false recall for category lists ($M = .31$) was greater than that for DRM lists ($M = .26$). Recall that previous research has not found differences in false memory rates across list type (Howe, 2006; Howe et al., in press). That such an effect was obtained here might cause some to interpret this as evidence for FTT, particularly as average BAS was lower for category than DRM lists, suggesting that something other than BAS (e.g., gist) was driving false memory rates. However, such an interpretation would be incorrect. Not only was there a main effect for label, $F(1, 302) = 7.14$, $p < .01$, $\eta^2 = .02$, where no labels resulted in lower false recall rates ($M = .26$) than labels ($M = .31$), but, more important, a List × Label interaction, $F(1, 302) = 7.32$, $p < .01$, $\eta^2 = .02$. Post-hoc tests showed that there were no differences in false recall rates for DRM ($M = .26$) and category ($M = .26$) lists in the no label condition, but rates of false recall were higher for category ($M = .35$) than DRM ($M = .27$) lists when labels were present. Importantly, as we show subsequently, consistent with AAT but not FTT, this effect was due to higher associative strength relations between labels and the critical lures in category than DRM lists.

There were two other main effects, one for age, $F(3, 302) = 40.59$, $p < .001$, $\eta^2 = .29$, where post-hoc Newman-Keuls tests ($p < .05$) revealed that 5-year-olds ($M = .14$) falsely recalled less than 7-year-olds ($M = .25$) who falsely recalled less than 11-year-olds ($M = .34$) who falsely recalled less than 18-year-olds ($M = .41$), and one for strength, $F(2, 604) = 114.45$, $p < .001$, $\eta^2 = .28$, where post-hoc tests showed that, like true recall, BAS was the critical determinant for false recall, with High BAS-High FAS lists ($M = .35$) and High BAS-Low FAS lists ($M = .35$) yielding more false recall than Low BAS-High FAS lists ($M = .15$). Three more first-order interactions emerged as well as a second-order interaction. First, there was a List × Strength, $F(2, 604) = 107.25$, $p < .001$, $\eta^2 = .26$, where post-hoc tests confirmed that although High BAS yielded the greatest false recall for both DRM and category lists, like true recall, each list type had a unique ordering of strength relations. That is, for DRM lists, the ordering was High BAS-High FAS ($M = .45$) > High BAS-Low FAS ($M = .26$) > Low BAS-High FAS ($M = .08$), whereas for the category lists the ordering was High BAS-Low FAS ($M = .45$) > High BAS-High FAS ($M = .25$) > Low BAS-High FAS ($M = .22$). These patterns for false recall are the same as those obtained for true recall. Second, there was an Age × List interaction, $F(3, 302) = 6.16$, $p < .001$, $\eta^2 = .09$. As can be seen in the right panel of Table 2, and was confirmed by post-hoc tests, category lists yielded higher false recall than DRM lists up to 11-year-olds and then no difference for adults. Third, there was an Age × Strength interaction, $F(6, 604) = 19.17$, $p < .001$, $\eta^2 = .16$. As can be seen in the right panel of Table 3, and was confirmed by post-hoc tests, there were no differences in false recall due to strength for 5-year-olds but

Table 4
Mean proportion of true recall as a function of age and label in Experiment 1 (standard errors in parentheses).

Age	No labels	Labels
5-year-olds	.17 (.012)	.19 (.012)
7-year-olds	.28 (.012)	.32 (.012)
11-year-olds	.41 (.012)	.45 (.012)
18-year-olds	.69 (.014)	.64 (.014)

Table 3
Mean proportion of true and false recall as a function of age and associative strength in Experiment 1 (standard errors in parentheses).

Age	True recall			False recall		
	High BAS High FAS	High BAS Low FAS	Low BAS High FAS	High BAS High FAS	High BAS Low FAS	Low BAS High FAS
5-year-olds	.18 (.010)	.19 (.010)	.19 (.010)	.12 (.026)	.19 (.027)	.10 (.018)
7-year-olds	.30 (.010)	.30 (.010)	.30 (.010)	.28 (.026)	.35 (.028)	.12 (.019)
11-year-olds	.44 (.010)	.45 (.010)	.40 (.010)	.36 (.026)	.45 (.027)	.21 (.018)
18-year-olds	.71 (.012)	.70 (.012)	.59 (.012)	.63 (.031)	.43 (.033)	.18 (.021)

the two High BAS lists yielded higher false recall than the Low BAS lists for 7-, 11-, and 18-year-olds.

Finally, there was an Age × List × Strength interaction, $F(6, 604) = 8.78$, $p < .001$, $\eta^2 = .08$. As can be seen in Panels A and B of Fig. 1, and was confirmed by post-hoc tests, both the High BAS-High FAS and High BAS-Low FAS lists produced greater false recall than Low BAS-High FAS lists for all age groups. The interaction occurs because (a) for category lists, High BAS-Low FAS lists produced greater false recall for children but the reverse held for adults and (b) for DRM lists, false recall rates were greater for High BAS-High FAS lists than High BAS-Low FAS lists for all age groups except the youngest (5-year-olds). Thus, false memories increased with age like true memories. As well, BAS was strongly predictive of false recall levels. That is, strong associative relations in both directions (High BAS-High FAS) increased false recall rates for DRM and category lists for adults and increased false recall rates for children on DRM lists. The pattern is less clear for category lists where High BAS-Low FAS but not High BAS-High FAS increased false recall rates for children. However, despite this, the overall pattern is very consistent: in all cases, high BAS had a significant effect on children's false memory production and this was independent of gist manipulations.

Together, these results are more consistent with the predictions from AAT's conceptualization of false memory and its development than with FTT's. That is, false recall

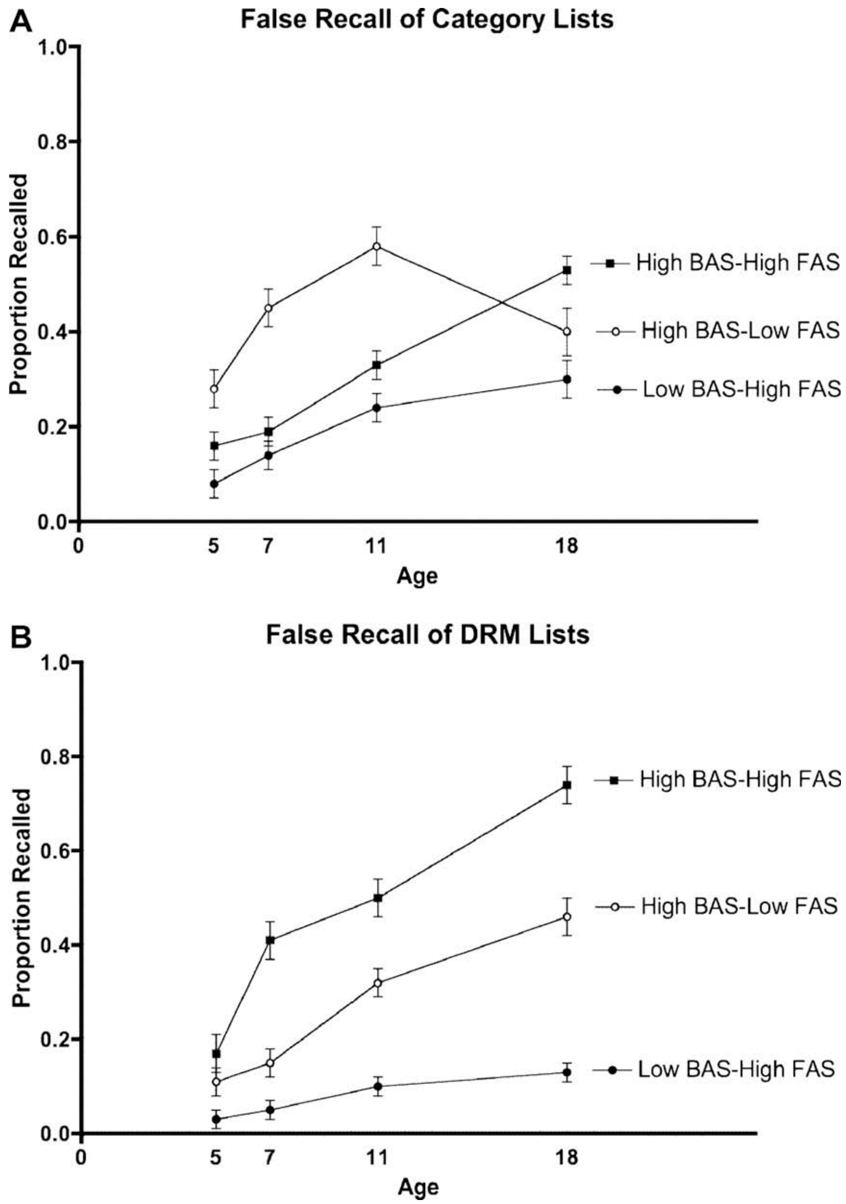

Fig. 1. Mean proportion falsely recalled as a function of Age and Strength for Category lists (Panel A) and DRM lists (Panel B) for Experiment 1.

rates were directly related to the list's associative strength, particularly BAS, regardless of whether those lists were mainly associative (DRM) or categorical. In fact, whereas true recall was negatively (but not significantly) related to BAS ($r(11) = -.125$, $p = .35$; category $r(5) = -.112$; DRM $r(5) = -.145$), false recall was positively (and significantly) related to BAS ($r(11) = .505$, $p < .05$; category $r(5) = .408$; DRM $r(5) = .534$). This dissociation was observed regardless of whether labels were present or not and whether lists were DRM or categorical.

Interestingly, when false recall rates were higher in category than DRM lists, this effect was restricted to conditions in which category labels were explicitly presented prior to the category list itself. That is, for false recall, regardless of age, descriptive labels selectively increased (primed) rates of false recollection for category but not DRM lists. Although at first blush this may be seen as consistent with FTT's claim that labels aided children's and adults' processing of gist across the list, it is not. First, if such were the case, we would have anticipated the same effects for DRM lists when given an appropriate organizing label. This was not observed. Second, FTT anticipates an interaction of labeling with age. That is, because younger children are particularly deficient in gist processing, the labeling effect should have been larger for younger children than older children and adults. Again, no such interaction emerged, a finding consistent with that of Holliday et al. (2008).

Third, and finally, it turns out that that this finding is consistent with AAT's prediction that labels can serve to increase false memory rates when they increase associative-activation of the critical lure. That is, this effect appears to be most likely related to the strong associative relations between the critical lure and descriptive labels for category, but not DRM, lists. In fact, further investigation revealed stronger associative relations between the critical lures (top exemplars from category norms) and the descriptive labels for category lists ($M = .34$) than for DRM lists ($M = .13$) and this difference is significant, $t(10) = 2.06$, $p < .05$. Using a partial correlational analysis and holding list BAS constant, associative strength from the label to the lure was a significant predictor of false recall in category lists ($r(3) = .814$, $p < .05$) but not DRM lists ($r(3) = .229$, $p = .36$). Thus, rather than providing support for FTT these findings provide further evidence for the important role that associative strength plays in the false memory illusion, regardless of age. Consistent with the earlier suggestion that labels may increase false memory because they increase associative-activation of the critical lure, and not because they increase gist extraction rates *per se* (e.g., Holliday et al., 2008; Odegard et al., 2008), these findings show that both list-to-lure and label-to-lure associative strength contribute independently to false memory illusions.

Interestingly, when descriptive labels were not provided, false recall for DRM and category lists did not significantly differ and, as suggested by AAT, false recall varied solely as a function of associative strength. Overall, then, for all lists associative strength, particularly BAS, played a key role in determining false memory rates. Although FAS was also important, particularly for DRM lists, developmental trends were controlled mainly by BAS. These results are consistent with previous findings with children (Howe,2005, 2006) and tend to favor an associative-activation approach to the development false memory illusions.

Experiment 2

The second experiment was designed to further examine the role of associative strength in the development of false memories by implementing a different type of manipulation. Specifically, we varied whether the list contained all items (normal or intact lists), all of the items except the strongest BAS term (strongest item removed), or all of the items except for several weaker items, ones that summed to the same BAS value as the strongest item removed (weak items removed), for both DRM and category lists. According to AAT, if BAS plays the primary role in false recall, then regardless of whether a single strong item is removed or several weaker items are removed that sum to the same BAS value as the strongest one, then there should be no difference in false memory rates for strong and weak lists but both should have lower false memory rates than lists with higher associative strength (the intact list). Although neither FTT nor AAT makes specific predictions about different rates of spreading activation[2], some spreading activation models suggest that removing stronger items would produce lower activation levels of the critical lure because there may exist higher rates of spreading activation from a single, strong item than the summative strength of removing several weaker items. If so, then removing a stronger item should have more of an impact on false recall than removing several weaker items whose individual strengths do not carry as much weight in an associative network (for a more in-depth analysis of priming effects, see Tse & Neely, 2007).

This manipulation also provides a test of Deese's (1959) and Underwood's (1965) different hypotheses about the importance of associative strength versus the importance of number of associates creating implicit associative responses (IARs) (also see Robinson & Roediger, 1997). For Deese (1959), false recall rates are correlated with BAS of the list words to the critical lure—lists with higher BAS should have higher false recall than lists with lower BAS regardless of the number of items in the list. Implicit in Underwood's (1965) theory, false recall rates although affected by BAS are also correlated with the number of IARs generated to the unpresented critical lure. The more items presented that generate IARs to the critical lure the more likely it is that the critical lure will be falsely remembered. Thus, according to Underwood (1965), lists with more items removed (our weaker items removed conditions) should result in fewer false recollections than lists with one item removed (our stronger item removed conditions) despite no overall differences in BAS (assuming that even the weakest item removed would have produced an IAR). For Deese (1959), no differences should emerge in false recollection rates between these two conditions because BAS remains constant. The only difference that should appear is between the intact lists and those with reduced BAS.

In previous studies, researchers have examined variations in list length on false recollection rates, although they have not controlled for levels of BAS. For example, Robinson and Roediger (1997) varied the number of items presented from 3 to 15 and found a linear increase in adults' false memory rates with list length (ranging from 3% to 30%). Although they left the three highest BAS items in each list, they did not measure the variation in BAS systematically across list length. Sugrue and Hayne (2006) also varied list length (7 items vs. 14 items), leaving the top seven BAS items in each list, but did not systematically vary BAS with list length. They found that longer lists increased false recall rates by 127% for adults and 48% for children (9- and 10-year-olds).

To our knowledge, this is the first experiment to simultaneously vary list length while keeping BAS constant. This methodology not only permits a test of the Deese (1959) and Underwood (1965) hypotheses, but also of AAT's claim that BAS, not number of items, is critical to false memory production in children and adults. As in Experiment 1, we also examined FTT's claims about the importance of across-the-list-gist by examining the effects of labels (present versus absent) and list type (DRM vs. category). That is, we examined the claim that younger children should produce more false memories when explicit gist cues were available than when they were not as well as whether younger children were more likely to produce false recall when categorized lists were presented than when DRM lists were presented. In order to provide a better controlled test of this hypothesis in Experiment 2, we equated the associative-activation of the labels to the lure across DRM and category lists.

Method

Participants

A new sample of 210 children and adults participated (53% male) in this experiment. There were 47 5-year-olds ($M = 5.4$, $SD = 3$ months), 50 7-year-olds ($M = 7.4$, $SD = 4$ months), 53 11-year-olds ($M = 11.4$, $SD = 4$ months), and 60 18-year-olds ($M = 18.6$, $SD = 8$ months). The children were White, came from middle-class homes, and were tested following parental consent and their assent on the testing day. Adults were tested following signed, informed consent. Participants were randomly assigned to the label and no label conditions, with half of the participants per age group (50% male) assigned to each condition.

Design, materials, and procedure

A 2(List: DRM vs. Category) × 3(Item removed: none vs. strongest vs. weaker) × 4(Age: 5- vs. 7- vs. 11- vs. 18-year-olds) × 2(Labels: No labels vs. Labels) design was used where list and item removed were manipulated within subject and age and labels were between-subjects variables.

A different set of 12, 15-item lists (6 DRM lists drawn from Roediger et al., 2001b, and 6 category lists drawn from van Overschelde et al., 2004; see Appendix B) were presented to each participant. As in Experiment 1, half of the participants received a label before list presentation and half received no label. Unlike in Experiment 1, the associative strength between the label and the critical lure did not differ significantly between DRM ($M = .11$) and category lists ($M = .25$), $t(10) = 1.16$, $p > .10$. The item(s) removed varied between lists with each participant being exposed to 4 lists in which nothing was removed (2 DRM and 2 category), four lists in which the strongest BAS item was removed (2 DRM and 2 category), and four lists in which the weakest BAS items were removed (2 DRM and 2 category). For the latter two types of lists, the cumulative BAS for each list was held constant such that the BAS associated with the strongest item removed equaled that of the BAS associated with the weakest items removed (see Appendix B). Although list length varied, particularly when weak items were removed (i.e., from a high of 13 presented items to a low of 11 presented items), all lists exceeded 10 items and are above the threshold necessary to produce false recollections in both children and adults (see Sugrue & Hayne, 2006).

Mean word frequency did not differ for list items, $t(10) = 1.27$, $p = 0.23$, or for critical lures, $t(10) = -1.03$, $p = 0.33$, between DRM and category lists. In addition, there were no differences for word frequency for list items, $F(2, 35) = 2.70$, $p = 0.75$ between the three conditions (intact vs. strongest removed vs. weaker removed) (word frequencies were taken from Stuart et al. (1993-1996)). Thus, the removal of list items did not affect mean word frequency values of the lists.

Each list, DRM and category, appeared in every possible condition. That is, each DRM and category list appeared as an intact list, as a strongest item removed list, and as a several weaker items removed list either with or without a label (varied across participants). Because of these controls, ones that were not possible in Experiment 1 (because lists were constrained by their associative strength values), we can exclude the possibility of lists effects in Experiment 2.

Participants were tested individually in a quiet room. Items were read aloud by the experimenter at a 3-second rate. Half of the participants in each age group received a descriptive label prior to hearing each list whereas the other half did not. Descriptive labels for category lists consisted of the category name included in the norms (van Overschelde et al., 2004) and for the DRM lists were constructed to highlight the shared semantic relations across list items and were drawn from Roediger et al. (2001b) (see Appendix B). As before, all lists have been used previously with children and adults in these age ranges (e.g., Brainerd et al., 2002; Howe, 2005, 2006, 2008a; Howe et al., 2008; Roediger et al., 2001b; Seamon et al., 2000), the labels have been used previously (e.g., Holliday et al., 2008; Howe, 2006), and, as is the norm with category associate lists, the critical target for category lists was the highest frequency word (most typical exemplar) for the selected category. DRM and category lists were selected on the basis of mean BAS as determined using the University of South Florida norms (Nelson et al., 1999). Like Experiment 1, in cases where associative strength was not available, the Edinburgh Associative Thesaurus was consulted (Kiss et al., 1973). DRM lists were also selected on the basis of mean BAS as indicated in the Roediger et al. (2001) norms.

Prior to the presentation of the first list, participants were given general memory instructions to listen carefully to the words on the list, as they would be asked to recall the words later. Following these instructions, participants in the label condition were given a descriptive label describing the theme of the list prior to list presentation. Participants in the no label condition simply received the word list. Following list presentation, participants were given a 30-second distractor task (circling randomized pairs of letters) prior to recall. Once recall was complete, the next cycle began. This procedure was repeated for the remaining 11 lists. Items were presented in the usual descending order of BAS. All presentation and recall was oral and list order was randomized within each age group.

Results and discussion

Preliminary analyses showed that gender was not a source of significant variation and was eliminated from subsequent analyses. As before, true recall findings are presented first followed by false recall.

True recall

The proportion of correct responses was analyzed using a 2(List: DRM vs. Category) × 3(Item removed: none vs. strongest vs. weaker) × 4(Age: 5- vs. 7- vs. 11- vs. 18-year-olds) × 2(Labels: No labels vs. Labels) ANOVA. There was a main effect for list, $F(1, 202) = 145.33$, $p < .001$, $\eta^2 = .42$, where category lists ($M = .40$) were better recalled than DRM lists ($M = .34$), and a main effect for age, $F(3, 202) = 379.08$, $p < .001$, $\eta^2 = .85$, where post-hoc tests showed that 5-year-olds ($M = .17$) recalled less than 7-year-olds ($M = .27$) who recalled less than 11-year-olds ($M = .40$) who recalled less than 18-year-olds ($M = .64$). These effects were modified by an Age × List interaction, $F(3, 202) = 3.94$, $p < .01$, $\eta^2 = .06$. As seen in Table 5, and confirmed by post-hoc tests, although category lists were recalled better than DRM lists at all ages, the absolute magnitude of these differences increased with age. Importantly, there was also an Age × Item removed interaction, $F(6, 404) = 10.45$, $p < .001$, $\eta^2 = .13$. As can be seen in the left panel of Table 6, and confirmed by post-hoc tests, there were no differences in recall as a function of items removed for 5- and 7-year-olds, whereas 11-year-olds recalled significantly fewer items from the weak-items-removed list than the strong-item-removed and intact (normal) lists. This trend continued for adults who recalled the least from the weak-items-removed lists than the strong-item-removed lists and the most from intact lists. Thus, rather than being a list strength effect *per se*, amount recalled was positively correlated with list length for both older children and adults. Finally, there were no effects associated with label. It would seem that when lists are better controlled than in the first experiment, the label effect disappears.

Table 5
Mean proportion of true recall as a function of age and list in Experiment 2 (standard errors in parentheses).

Age	DRM lists	Category lists
5-year-olds	.15 (.012)	.18 (.013)
7-year-olds	.24 (.012)	.30 (.012)
11-year-olds	.36 (.011)	.44 (.012)
18-year-olds	.61 (.011)	.67 (.011)

False recall

The proportion of critical lures falsely recalled was analyzed using a 2(List: DRM vs. Category) × 3(Item removed: none vs. strongest vs. weaker) × 4(Age: 5- vs. 7- vs. 11- vs. 18-year-olds) × 2(Labels: No labels vs. Labels) ANOVA. There were three main effects. First, there was a main effect for age, $F(3, 202) = 14.79$, $p < .001$, $\eta^2 = .18$, where post-hoc tests confirmed that false recall increased significantly across age. That is, 5-year-olds ($M = .17$) falsely recalled fewer critical items than 7-year-olds ($M = .25$) who recalled fewer false items than 11-year-olds ($M = .33$) who recalled fewer false items than adults ($M = .40$). Second, there was a main effect for list, $F(1, 202) = 88.31$, $p < .001$, $\eta^2 = .30$, where there were more false memories for DRM lists ($M = .36$) than for category lists ($M = .22$) a finding more typical in this literature for false recall and the opposite pattern to that for true recall both for children (Howe, 2006; Howe et al., in press) and adults (Dewhurst et al., 2009). Finally, there was a main effect for item removed, $F(2, 404) = 144.79$, $p < .001$, $\eta^2 = .42$, where post-hoc tests confirmed that intact lists ($M = .44$) exhibited more false recall than either the strong-item-removed lists ($M = .21$) or weak-items-removed lists ($M = .21$), these latter two conditions did not differ. Importantly, unlike true recall, the effect for false recall would appear to be strength based rather than length based inasmuch as removing the same amount of BAS from a list, regardless of whether it was a single strong item or several weaker ones, halved average false recall rates. A 50% decline in false recall rates is consistent with previous research with children and adults (e.g., Sugrue & Hayne, 2006).

Table 6
Mean proportion of true and false recall as a function of age and items removed in Experiment 2 (standard errors in parentheses).

Age	True recall			False recall		
	Intact	Strong Removed	Weak Removed	Intact	Strong Removed	Weak Removed
5-year-olds	.18 (.014)	.16 (.015)	.16 (.013)	.28 (.041)	.11 (.030)	.10 (.029)
7-year-olds	.27 (.013)	.26 (.014)	.27 (.013)	.38 (.040)	.19 (.029)	.18 (.028)
11-year-olds	.43 (.013)	.41 (.014)	.36 (.012)	.49 (.039)	.25 (.029)	.24 (.027)
18-year-olds	.68 (.012)	.64 (.013)	.61 (.012)	.60 (.036)	.30 (.027)	.30 (.025)

However, it should be noted that this 50% decline occurred with only a 24% reduction in BAS. What this shows is that changes in BAS do not map one-to-one onto changes in false memory rates, at least not for recall. In addition, this result shows that when BAS is systematically controlled and list length is varied, it is activation strength that determines false recall rates in children and adults not list length. This finding is consistent with AAT's predictions as well as Deese's (1959) hypothesis concerning the importance of BAS not list length in determining false recollection rates. Importantly, the finding that strong and weak lists exhibit exactly the same false memory rates excludes the possibility that variation in list length may account for the findings. This is inconsistent with the interpretation of Underwood's (1965) implicit hypothesis where the number of IARs determines false recollection rates but consistent with the interpretation that stronger items are more likely to evoke an IAR than weaker items. Interestingly, as can be seen in the right panel of Table 6, the difference between intact and changed lists tends to increase with age. However, this interaction was not significant. Finally, unlike the predictions of FTT, false recall rates were not affected by label, a finding consistent with previous research (e.g., Howe, 2006).

Summary and comparison of experiments 1 and 2

The pattern for true recall across both experiments was extremely consistent and showed that category lists were better recalled than DRM lists. This is consistent with previous findings (e.g., Howe 2006; Howe et al., in press). Differences occurred between experiments in the label manipulation where Experiment 1 showed increased children's (but not adults') true recall when a label was presented, whereas Experiment 2 did not reveal any differences between label and no label conditions.

For false recall, both experiments showed that BAS between list items and the critical lure is a key determinant for false memory production in children and adults. However, like true recall, differences occurred for the label manipulation. Experiment 1 showed a higher level of false recall for category lists but not for DRM lists when a label was presented. In contrast, Experiment 2 revealed no differences in false memory rates between the label and no label conditions. Further analyses showed that this label effect in Experiment 1 was due to the higher level of associative-activation between the label and the critical lure in category than DRM lists. It seems that when the associative strength between the label and the critical lure is better equated across DRM and category lists, as it was in Experiment 2, the label effect disappears.

Interestingly, for the no label conditions, the findings of the two experiments differ slightly in how BAS affects false recall. That is, in Experiment 1 variation in BAS did not have as strong an effect on 5-year-old's false recall as it did on all of the other age groups. This may be due to the lower false memory rate for this younger age group in Experiment 1 than Experiment 2. That is, both high BAS/high FAS and high BAS/low FAS conditions in Experiment 1 showed lower false memory rates than intact lists in Experiment 2. Although false recall rates were generally higher for all children in Experiment 2 than Experiment 1, they were not significantly different across experiments ($F < 1$) and were not attributable to differences in word frequency across lists ($t(22) = 0.79$, $p = .44$) or across critical lures ($t(22) = -1.42$, $p = .17$). However, the effects of BAS may have been obscured for the youngest children in the first experiment by possible floor effects, effects that were not present in 5-year-old's false recall in the second experiment.

Despite these modest differences in results across experiments, the overall pattern is very clear. As previous research suggests, children are better at remembering category lists than DRM lists. False memories are driven by the amount of associative strength between the list items and the critical lure (BAS). Providing a descriptive label for the list items overall does not, in and of itself, increase false memory levels. Labels only have an effect on false memory production when they show high levels of association to the critical lure. Again, this indicates that the level of association to the critical lure is the significant determinant for children's false memory production.

Moreover, both experiments demonstrate that true recall is affected more by gist factors whereas false recall is affected by associative strength. These outcomes are more consistent with AAT than FTT particularly inasmuch as gist manipulations (i.e., category vs. DRM lists, gist-integrating labels vs. no labels) primarily affect true, not false, recall. Indeed, in some strong versions of FTT, specific gist manipulations are said to primarily affect false, not true, recall: "FTT predicts that the effects of the semantic cue instructions should only be found in false recall of critical lures because true recall of list words is based on verbatim traces" (Holliday et al., 2008, p. 70). To the contrary, both Experiments 1 and 2 show that variation in false recall rates are controlled by associative strength, not gist, manipulations.

Despite the clear pattern of results, one question that arises is whether a direct comparison between category lists and DRM lists is appropriate since they have different semantic structures. According to Wu and Barsalou (2007) there are six types of semantic relations that can occur between words: antonymy (words of opposite value, e.g., cold and hot), entity (entity-property relationships between words, e.g., chair and wood), introspective (mental state-property relationships between words, e.g., happy and sun), situational (words are related to the same association, e.g., medical treatment), synonymy (words that have the same meaning, e.g., bunny and rabbit), and taxonomy (words that belong to the same taxonomic category, e.g., cats and dogs are both animals). Although it is often difficult to discriminate associative strength and semantic overlap (see Hutchison, 2003), Brainerd et al. (in press) suggest that DRM lists contain many of these semantic relations whereas categorical lists contain only taxonomic relations. The meaning connections in categorical lists are therefore less varied than in DRM lists. If so, one could argue that extracting the overall meaning from word lists is easier when there is a larger variety or different types of semantic relations available. Alternatively, and more likely, such lists may lead to multiple gists, leading to more confusion than when only a single gist is present. Thus, it may be easier to extract the overall meaning from lists when only one type of semantic relation (e.g., a taxonomic relation) is available. This might be particularly true for children who may have more difficulty juggling multiple relations in a list. That is, children may be better with lists containing a single relation, especially when it comes to extracting the gist of a list. Because categorized lists contain only one semantic relation (i.e., taxonomy) they are thematically more consistent than DRM lists. For this reason extraction of a separate thematic representation should be easier for thematically more consistent material (i.e., categorized lists) than for materials that have different types of relations (i.e., DRM lists). If extraction of thematic representations (i.e., gist extraction) of presented material is responsible for false memories as suggested by FTT, then a higher level of false memories in categorized lists that are thematically more consistent than DRM lists would be expected. However, this is not consistent with the present findings.

Given findings from these experiments as well as other recent findings (e.g., Howe, 2006; Howe et al., 2008, in press), a reasonable conclusion is that at a minimum, gist-like manipulations can and do affect children's true recall and associative strength can and does affect their false recall. In Experiments 1 and 2 we manipulated the availability of gist processing by varying whether participants studied DRM or categorized lists and by explicitly cuing or not cuing key semantic relationships across list items using labels. As just discussed, these are powerful manipulations that have as their locus of effect, true recall. In fact, across a number of experiments now it appears as though effects at false recall may be limited to variation primarily in BAS. Of course, some might argue that gist manipulations, despite having substantial effects on true recall, did not affect false recall rates because of lack of power. However, such arguments founder given the sheer number of participants for which such effects have been relatively weak or absent (*N*s are now well over 1000) and the number of times these effects have failed to appear in false recall relative to their general robustness in true recall. More importantly, these arguments founder because when these effects do emerge, they are consistent with AAT's prediction that categorized lists and theme-relevant labels can and do affect false recall rates, but only when they contribute to the overall associative-activation of the critical lure.

Of course, it is prudent to test assumptions using a variety of operationalizations of key constructs. Indeed, any one manipulation by itself has potential limitations and, testing gist-based versus associative-activation explanations by only manipulating explicit variables such as list content and cuing has its own set of concerns (e.g., differential sensitivities of measuring recollection with recall versus recognition, presenting cues only at encoding versus at encoding and retrieval), ones that may limit the degree to which findings can be generalized. To correct this situation, we conducted a third experiment, one that examined a different, more implicit test of gist versus associative-activation.

Experiment 3

In Experiments 1 and 2 it has been established that children's false memories are dependent on associative-activation processes indexed by BAS and that this is independent of gist manipulations. In Experiment 3, we take a different tactic and examine what might be considered a more implicit operationalization of gist-across-the-list, one that has to do with the semantic coherence of the studied material, namely, semantic density (SD). *SD* is an index of associations among list items themselves rather than between the list items and the critical lure. Specifically, if the thematic cohesion of a list is related to a child's ability to access the list's gist, then lists whose cohesion index (semantic density) is high should evince higher levels of false recollection (particularly recall measures as they clearly involve gist recollective processes; see Brainerd & Reyna, 2005) than lists whose cohesion index is low (also see discussion in Gallo, 2006). Indeed, according to FTT, there should be an interaction between age and semantic cohesion/density such that semantic density effects should be larger with younger than older children given younger children's poorer spontaneous gist extraction abilities. Because adults' ability to extract gist regardless of cohesion is not contested in the literature, in Experiment 3 we concern ourselves solely with children.

Before turning to the experiment itself, it is worth noting that we are not aware of any existing research that has varied semantic cohesion/density with children. As close as we could find were studies in which thematic blocking of list materials was manipulated in DRM studies with adults (e.g., McDermott, 1996), resulting in increases in both true and false recall rates; blocking manipulations that have been used to examine children's true recall of categorized materials (e.g., see Lange, 1978), resulting in increases in correct recall (false recall was not examined); and several studies manipulating semantic density with adults that examined the role of episodic context in true recall (e.g., McEvoy et al., 1999; Nelson, Goodmon, & Akirmak, 2007; Nelson & Zhang, 2000). For example, McEvoy et al. (1999) found that when BAS was held constant, high levels of semantic density (as measured by the number of interitem connections) elevated true recall and low levels of semantic density decreased true recall. The reverse was found for false recall: low levels increased false recall whereas high levels of semantic density decreased false recall. It has been suggested that because high levels of semantic density increased true recall, false memories are less likely to occur. Furthermore, exactly the opposite pattern was found for false recognition. That is, high levels of semantic density increased false recognition. McEvoy et al. (1999) suggest that the high level of semantic density increased the familiarity of the critical lure on the recognition test and therefore false recognition. The latter finding could also be explained by FTT. However, because of the lower amount of false recollection that occurred for the same lures in the recall paradigm in the low-level semantic density condition, it seems unlikely that a pure gist-based explanation would suffice. Overall, the studies to date support the conclusion that both BAS and semantic density are more consistent with models emphasizing associative-activation.

In this experiment, we manipulated semantic cohesion/density by varying the degree to which items on the list were associated with one another. Items on DRM and category lists not only vary in the number of connections among items within the list (the metric used by McEvoy et al., 1999) but also in the degree to which they are interconnected (a metric akin to BAS). That is, in addition to measuring the degree to which items on a list give rise to a critical lure (as measured by BAS), one can measure the degree to which the list items themselves are linked to one another (semantic density or *SD*). Indeed, categorized lists that are semantically denser give rise to higher rates of true recall than lists whose members are not as highly interconnected (e.g., see Talmi & Moscovitch, 2004).

In Experiment 3, we operationalized *SD* using Nelson et al.'s (1999) norms for associative connectivity. Essentially, *SD* is a composite measure of the degree to which items within a list activate each other (both FAS and BAS between list items). Both *SD* and BAS are based on associative strength but the difference is that the latter is based on direct association between list items (e.g., *bed*) and the critical lure (e.g., SLEEP) whereas the former is based on associations within a list and these can be forward and backward (e.g., *bed* to *rest* and *res*t to *bed*). Because both BAS and *SD* are based on the same associative strength metric, direct comparisons can be made between the individual components' contributions to true and false recall. Research has shown that the higher a list's *SD*, the better recall is for materials on that list, partly because a remembered item is more likely to cue other, related items on a list (see McEvoy et al., 1999; Nelson et al., 2007; Talmi & Moscovitch, 2004). Under conditions such as these, most, if not all, memory models (including AAT and FTT) accurately predict that true recall should increase as *SD* increases. Theories differ, however, in whether *SD* should also affect false recall. Theories such as FTT predict that to the extent that higher levels of *SD* are correlated with a greater likelihood of extracting list gist, false recall rates should also increase (and as noted earlier, this may interact with age). However, for theories based on associative-activation (AAT), variation in *SD*, while affecting true recall, should have smaller or null effects on false recall, depending on how much *SD* contributes to item-specific distinctiveness (McEvoy et al., 1999). Moreover, according to AAT, the critical variable determining false recall of critical lures is BAS, not across-the-list connectivity.

In Experiment 3, then, we varied both BAS as well as *SD*. For half of the lists, BAS was higher than *SD* (higher BAS) and for the other half of the lists, *SD* was higher than BAS (higher *SD*)[3]. Both FTT and AAT predict that true recall rates should be higher for lists high on *SD* (i.e., true recall should be higher for higher *SD* lists than higher BAS lists). However, the theories differ for false recollection. According to the FTT, lists higher in *SD* should also exhibit greater false recall than those lower in *SD* [i.e., (higher *SD*) > (higher BAS)] and these effects should be stronger for younger than older children. According to AAT, because *SD* is not related to false recall but BAS is, *SD* should have no effect but lists higher in BAS should produce higher levels of false recall than those with lower levels of BAS [i.e., (higher BAS) > (higher *SD*)].

Method

Participants

A new sample of 77 children participated (50% male) in this experiment. There were 25 5-year-olds (*M* = 5.3, *SD* = 4 months), 25 7-year-olds (*M* = 7.4, *SD* = 4 months), and 27 11-year-olds (*M* = 11.3, *SD* = 4 months). The children were White, came from middle-class homes, and were tested following parental consent and their own assent on the testing day.

[3] Research has shown that recall probabilities can be well described by the strength of interitem connections in networks that include connections between, for example, cues (items) and targets (lures) and the organization between the cues (items) themselves. Nelson and Zhang (2000) demonstrated that it is possible to calculate an activation strength index that integrates these different forms of associations (e.g., the association between items and the critical target or the association between items themselves). Most importantly, this activation strength index can be used to predict recall. According to Nelson and Zhang (2000) this activation strength index (AS) can be seen as a parallel activation process involving all connections in an associative network. These connections contribute additively to strengthening the target word (in our case the critical lure) and its associates. In other words, AS includes the degree to which items are connected to a target (lure) plus the degree to which items are connected between themselves (AS = BAS + SD). For example, the activation of GREEN is determined by the degree to which the related concepts *yellow*, *blue* are associated with GREEN (backward associative strength) plus by the degree to which *yellow* and *blue* are associated. The degree to which list items are associated with each other gives an indication of how semantically dense a list is. In particular, the stronger items are associated with each other the denser is a list. In this context, the activation strength (AS) of a word or critical lure may not only be determined by the amount of backward associative strength (BAS) but also by the semantic density (SD) of the list items. Because AS may be an index of recall probabilities, it was necessary to hold AS constant across lists to obtain independent estimates of the roles of BAS and SD in children's true and false recall. That is, High BAS/High SD and Low BAS/Low SD conditions were not included because they confound overall activation strength with the manipulation of interest. Because our interest was focused squarely on the relative roles of SD and BAS, not overall AS, the only relevant test conditions were High BAS/Low SD and Low BAS/High SD, the only two conditions that can be equated for overall AS. In a full, 2 x 2 factorial design we would confound activation strength (e.g., it would be highest in the High BAS/High SD condition, moderate in the High BAS/Low SD and Low BAS/High SD conditions, and lowest in the Low BAS/Low SD condition) and it would not be possible to draw any conclusions from the individual components' contributions to true/false recall independent ofoverall activation strength. With AS held constant, we were able to assess how lists that were more dominant in SD than in BASor more dominant in BAS than SD affected true and false recall, independent of variation in AS.

Design, materials, and procedure

A 2(List condition: higher *SD* vs. higher BAS) × 3(Age: 5- vs. 7- vs. 11-year-olds) design was used where list condition was manipulated within subject and age was a between-subjects variable. Each child received 6, 11-item word lists (see Appendix C). The norms for BAS and *SD* were obtained from the n × n matrix in Nelson et al. (1999). As before, BAS values represent the associations between each list item (e.g., *yellow*) and the critical lure (e.g., GREEN). *SD* values represent associations between the list items within a list and were calculated following Nelson and Zhang (2000). For example, in the GREEN list the item *yellow* produces the list item *blue* with a probability of .07, the list item *blue* produces the list item *red* with a probability of .11, *red* produces *yellow* with a probability of .02, and so forth. *SD* comprises the sum of all interitem associations. Out of the 6 word lists, 3 lists were lower in BAS than in *SD* (*SD* > BAS: Mean difference = 1.48) and 3 lists were higher in BAS than in *SD* (BAS > *SD*: Mean difference = 1.47). There were no significant differences in the mean differences between these list conditions. Moreover, activation strength (see Footnote 3) did not differ across list conditions ($t(4) = 1.40$, $p = .24$). Similarly, there were no significant differences in mean word frequency for list items, $t(4) = .42$, $p = .70$, or in word frequencies of the critical lures between the two list conditions, $t(4) = -.71$, $p = .52$ (word frequencies were taken from Stuart et al. (1993-1996)).

Participants were tested individually in a quiet room. Prior to presentation of the first list, children were given standard memory instructions indicating that they would be presented with lists of words and that they should try to remember as many as possible. Following these instructions, participants were presented the first list, administered a 30-second distractor task (circling randomized pairs of letters), and then asked to recall as many words as possible from the list. This study-distractor-test cycle was continued until all six lists had been presented. Items were presented in the usual descending order of BAS. Items were audio recorded and presented to the children at a 3-second rate and list order was counterbalanced across children.

Results and discussion

Preliminary analyses showed that gender was not a significant source of variation and was eliminated from subsequent analyses. Like Experiments 1 and 2, true recall findings are presented first followed by false recall.

True recall

The proportion of correct responses was analyzed using a 2(List condition: higher *SD* vs. higher BAS) × 3(Age: 5-, 7-, and 11-year-olds) ANOVA. There was a main effect for age, $F(2, 74) = 111.39$, $p < .001$, $\eta^2 = .75$. Post-hoc Bonferroni comparisons (all *p*s < .001) showed that 11-year-olds ($M = .54$) correctly recalled more than 7-year-olds ($M = .33$) who correctly recalled more than 5-year-olds ($M = .20$). There was also a main effect for list condition, $F(1, 74) = 58.11$, $p < .001$, $\eta^2 = .44$, where more items were correctly recalled when semantic density was high (higher *SD*; $M = .40$) than when it was low (higher BAS; $M = .31$). As can be seen in the left panel of Table 7, the Age × List condition interaction was not significant (i.e., semantic density effects were similar at each age).

Table 7
Mean proportion of true and false recall as a function of age and list in Experiment 3 (standard errors in parentheses).

Age	True recall		False recall	
	Higher *SD*	Higher BAS	Higher *SD*	Higher BAS
5-year-olds	.24 (.022)	.16 (.018)	.03 (.028)	.05 (.038)
7-year-olds	.37 (.022)	.28 (.018)	.07 (.028)	.20 (.038)
11-year-olds	.60 (.021)	.49 (.017)	.10 (.027)	.17 (.037)

False recall

The proportion of critical lures falsely recalled was analyzed using a 2(List condition: higher *SD* vs. higher BAS) × 3(Age: 5-, 7-, and 11-year-olds) ANOVA. There was a main effect for age, $F(2, 74) = 5.50$, $p < .01$, $\eta^2 = .13$. Post-hoc Bonferroni comparisons (*p*s < .02) showed that 5-year-olds ($M = .04$) falsely recalled fewer critical lures than 7-year-olds ($M = .13$) and than 11-year-olds ($M = .14$) and the latter two age groups did not differ. There was also a main effect for list condition, $F(1, 74) = 8.10$, $p < .02$, $\eta^2 = .10$, where more critical lures were falsely recalled when BAS was high (higher BAS; $M = .14$) than when it was low (higher *SD*; $M = .06$). As can be seen in the right panel of Table 7, the Age × List condition interaction was not significant (i.e., BAS effects were similar at each age).

Overall, then, consistent with predictions of most theories (including AAT and FTT), semantic density (a measure of within-list cohesion) increased true recall rates for all children regardless of age. Like children's categorized recall, higher levels of intralist similarity produced higher rates of true recall. However, as predicted by AAT, these semantic density effects did not translate into higher rates of false recall. As seen in previous studies with children (e.g., Howe et al., 2004, 2008, in press), what controls false memory rates is BAS—when BAS is high, false recall of critical lures is high; when BAS is low, false recall of critical lures is low. Similar effects were obtained here—when BAS was high, false recall of critical lures was high; when BAS was low, false recall of critical lures was low. Thus, like the findings when across-the-list gist was explicitly cued (Experiments 1 and 2), when across-the-list gist was manipulated implicitly (as measured by *SD*), true recall is the beneficiary, consistent with most theories of memory and memory development. However, these effects do not translate into changes in false recall rates. Rather, false recall rates in all three experiments reported here are controlled by BAS, consistent with AAT.

General discussion and conclusions

That both true and false recall increased with age is consistent with previous developmental research and both FTT and AAT. However, across experiments, the pattern of findings was more consistent with AAT's predictions of developmental patterns in false memory illusions than FTT's. Moreover, the results provided clear evidence of developmental continuity in the emergence of false memory illusions, consistent with what is known about the development of accurate remembering. In the remainder of this article, we discuss both of these trends, separately.

Associative-activation and fuzzy-trace theories of false memory development

The results of these experiments are more consistent with AAT (see Howe, 2005, 2006, 2008a, 2008b; Howe et al., 2008, in press) than with FTT (e.g., Brainerd & Reyna, 2005; Brainerd et al., 2008). Specifically, regardless of age, false recall rates were contingent solely on variation in associative strength and not on manipulations related to gist-across-the-list. Indeed, regardless of age, more false recall was evidenced when lists were high in associative strength, particularly BAS (Experiments 1, 2, and 3). Although FAS (Experiment 1) did play a role in false recall rates (as shown previously by Brainerd & Wright, 2005, with adults using recognition measures), particularly when explicit category primes were given prior to list presentation in the first experiment, overall list strength was the single best predictor of false recall rates. Neither the use of categorized lists nor the provision of explicit cues that primed the lists' gist increased false memory rates (Experiments 1 and 2) except, as noted, when those primes were highly associated with the critical lure and not because they primed gist-across-the-list. Moreover, when the degree of semantic cohesion among list constituents was increased, true recall was enhanced but false recall levels did not change (Experiment 3).

That associative-activation, especially as indexed by BAS, plays a major role in false memory production is not new in the adult literature (e.g., Gallo, 2006). With these experiments as well as other recent research with children (Howe, 2005, 2006, 2008a; Howe et al., 2008, in press), it is becoming equally apparent in the developmental literature that associative-activation is the key variable driving children's false memories too. Moreover, it is also becoming clear that some of the best-fitting models for adult false memory production are ones based on associative-activation (e.g., Gallo, 2006; Hutchison & Balota, 2005; Kimball et al., 2007). What these experiments and other recent findings in the developmental literature suggest is that associative-activation theory may also provide an appropriate model for the development of false memories (e.g., Howe, 2005; Howe et al., 2008; Metzger et al., 2008). In the rest of this section, we focus on the findings relevant to this claim.

First, consider the manipulation of categorical versus associative lists across Experiments 1 and 2. FTT predicted that the additional vertical, or gist-across-the-list, connections available across categorized lists should enhance false recollection rates, particularly in younger children than adults. In addition, category lists should exceed DRM lists in false recall rates particularly for young children because they contain fewer potential gists (the category itself) than DRM lists that can contain multiple gists. By contrast, AAT predicted that associative-activation (BAS) is the critical determinant for false memories independent of list type (categorical vs. associative). In particular, these additional relations in categorical lists, while enhancing true recall, should not have an impact on false recall unless they too contributed to the overall associative-activation of the critical lure. Like prior research (Howe, 2006; Howe et al., in press), it is this latter outcome that was obtained in the first two experiments in this article. Specifically, considering only those comparisons between category and DRM lists without the addition of explicit gist-priming cues, false recall rates were no greater for category lists than DRM lists. Rather, across Experiments 1 and 2, false recall rates were determined by overall associative-activation (BAS and FAS).

Second, consider the effects of explicit priming (cuing) using theme-relevant labels. FTT predicted that such cues should increase false memory rates because they prime across-the-list gist (e.g., Holliday et al., 2008) and that these effects should be larger in younger children than older children and adults because it is the younger children who are particularly deficient in spontaneous gist processing. By contrast, AAT makes the same prediction here as it did with the category-structure manipulation, namely, that false recall rates should not be altered by explicit cues unless they alter the overall associative-activation of the critical lure. Again, the results of the first two experiments in this article are consistent with AAT. In Experiment 1, false recall for categorized lists was higher when explicit cues were presented, but the same was not true for DRM lists. As it turned out, this increase was an artifact of increased associative-activation not priming of gist-across-the-list. That is, increased false recall was a consequence of the category labels being more highly associated with the critical lure (the category prototype) on the category lists than the labels used for the DRM lists. When these associative-activation differences between the cues were eliminated in Experiment 2, there were no effects due to explicitly cuing the gist for category or DRM lists.

When a more implicit measure of gist-across-the-list was employed in Experiment 3 (semantic cohesion or density), and it was contrasted with variation in BAS, we once again obtained findings that were more consistent with AAT than FTT. That is, increases in interitem cohesion were associated with increases in true recall but not false recall. The only variable (besides age) that influenced children's rates of false recall was changes in associative-activation levels (as indexed by BAS).

Taken together, the predictions of FTT (more false recall for children when gist relations were present [category lists], when gist was primed [cues], and when interitem semantic cohesion was high [semantically dense lists]) were not confirmed in any of the three experiments in this article. Instead, what was found across all of the experiments was that manipulations designed to increase the probability of gist extraction led to increases in true recollection but had no effect on false recollection. Indeed, across all three experiments and regardless of which gist-enhancing manipulation was examined, associative-activation was the single best determinant of false memory rates regardless of age.

Overall, then, because true and false recall, but especially false recollection in children and adults, depended on the strength of associative relations, these findings are consistent with more general associative-activation models of memory illusions using the DRM paradigm (e.g., Gallo, 2006; Kimball et al., 2007). Indeed, this conclusion fits well with other recent findings concerning children's susceptibility to the DRM illusion (e.g., Howe, 2005, 2006, 2008a, 2008b; Howe et al., 2008, in press). More important, it is consistent with speculation on the mechanisms underlying adults' DRM illusions (Gallo, 2006; Hutchison & Balota, 2005; Roediger et al., 2001b). For example, Roediger et al. (2001b) found for adults that BAS was the critical factor predicting false memories when examining other variables (e.g., word length, word frequency, concreteness, FAS, connectivity, and true recall). This is exactly the pattern that has been obtained across the three experiments here. When controlling for FAS (Experiment 1), word frequency (Experiments 1, 2, and 3), and connectivity (*SD*) (Experiment 3), BAS was the key factor driving false memories, regardless of age. Most importantly, the role of BAS in the false memory illusion was prevalent despite using different methods and experimental manipulations. Again, although BAS may serve as a proxy variable for a host of different type of relationships (phonological, orthographic, temporal contiguity, spatial proximity, superordinate relations, property relations, physical or conceptual similarity, etc.), and it may be difficult to isolate associative-activation from semantic overlap (Brainerd et al., in press; Hutchison, 2003; Roediger et al., 2001b), it is still the single best predictor of children's and adults' false memory illusions. We now consider this developmental continuity in more detail.

Developmental continuity in false memory illusions

One of the novel findings to emerge from these experiments is that age-related increases in true and false recall were observed regardless of the list's associative strength. That is, both strong and weak lists showed similar developmental trends. Moreover, both BAS (Experiments 1, 2, and 3) and FAS (Experiment 1) played a key role in children's and adults' false memory rates. Although BAS was clearly the main determinant of false memory rates for children and adults, FAS also contributed. This is consistent with Brainerd and Wright's (2005) findings for adults using considerably shorter lists and a recognition procedure. The novel contribution of this research is that the results extend this trend to measures of recall and, more importantly, to children's false memory development. In doing so, these findings add to the growing literature on memory continuity effects across age (Howe, 2000; Howe et al., 2008, in press).

In fact, these findings add to a growing list of memory phenomena that exhibit strong developmental invariance. Evidence has been accumulating for some time that many of the variables that govern encoding, consolidation, storage, retention, and retrieval processes in young children (and even infants) are the same as those that regulate these memory processes in older children and adults (for reviews, see Bauer, 2004; Howe, 2000). Of course, these continuities notwithstanding, there do exist significant developmental advances in memory processes across childhood and into adulthood (e.g., faster and more efficient encoding, storage, and retrieval; increased retention of information over longer intervals) that are contingent on advances in related cognitive domains (e.g., knowledge base, attentional resources, strategies) (for reviews, see Bauer, 2005; Howe, 2000). What these experiments add to this ever-expanding list is that the development of both veridical and false memory can be viewed as developmentally invariant phenomena and that the development of false memory in childhood can be subsumed under the same model as that used to account for false recollection in adulthood. That is, these age-invariant effects suggest that a single theory should be able to account for false memory and its development, one that relies on associative-activation processes.

The model supported by these experiments, as well as a number of prior studies (Howe, 2005, 2006, 2008a; Howe et al., 2008, in press), is AAT. Here, developmental changes in veridical recall and false memory illusions are driven by increases in the number and strength of associative relations in children's knowledge base (Bjorklund, 1987, 2004) as well as the speed and automaticity with which these associative relations are activated (Bjorklund & Jacobs, 1985; Howe, 2005; Metzger et al., 2008). These changes in number, strength, speed, and automaticity of associations and their activation are linked to other well-known changes in children's knowledge base, resources, and speed of processing in related cognitive domains (for reviews, see Bjorklund, 2004; Courage & Howe, 2002; Dempster & Brainerd, 1995). Together, these developmental changes lead to increased activation of studied and nonstudied but related information in memory, which in turn increases both true and false recall rates across childhood and into adulthood. Because the associative-activation theory provides a necessary and sufficient account of true and false memory data from both children and adults, and because the developmental mechanisms identified in this theory are well-known across a variety of other, developmentally continuous phenomena in memory and cognition, it may be more parsimonious to adopt this theory of false memory development than to alter other extant models.

Acknowledgments

This research was supported by a grant from the Natural Sciences and Engineering Research Council of Canada and by a grant from the Economic and Social Research Council of the United Kingdom (RES-062-23-0452) to Mark L. Howe. The authors thank Charles J. Brainerd, Keith Hutchison, Jim Neely, and an anonymous reviewer for comments on a previous draft of this article. Portions of this research were presented at the 2007 biennial meeting of the Society for Research in Child Development.

Appendix A

DRM and category lists used in Experiment 1

1. DRM word lists and labels

(i) High BAS—Low FAS

Critical Lure: Smell
Label: Something to do with your senses
MBAS: .290 **MFAS**: .015

ITEM	BAS	FAS
Nose	.108	.116
Breathe	.000	.000
Sniff	.442	.043
Aroma	.678	.000
Hear	.000	.000
See	.000	.000
Nostril	.000	.000
Whiff	.577	.000
Scent	.625	.029
Reek	.510	.000
Stench	.562	.000
Fragrance	.389	.000
Perfume	.393	.036
Salts	.028	.000

Critical Lure: Mountain
Label: Things to do with high places
MBAS: .163 **MFAS**: .037

ITEM	BAS	FAS
Hill	.428	.265
Valley	.195	.020
Climb	.291	.092
Summit	.108	.000
Top	.000	.041
Molehill	.256	.031
Peak	.248	.020
Plain	.000	.000
Glacier	.020	.000
Goat	.028	.000
Bike	.033	.000
Climber	.603	.031
Range	.000	.051
Steep	.061	.000

(ii) High BAS—High FAS

Critical Lure: Needle
Label: Something to do with spiky objects
MBAS: .203 **MFAS**: .063

ITEM	BAS	FAS
Thread	.760	.420
Pin	.290	.210
Eye	.000	.000
Sewing	.181	.240
Sharp	.030	.020
Point	.020	.020
Prick	.110	.010
Thimble	.220	.000
Haystack	.420	.030
Thorn	.030	.000
Hurt	.000	.000
Injection	.330	.000
Syringe	.520	.000
Cloth	.000	.000

Critical Lure: Doctor
Label: Medical things
MBAS: .245 **MFAS**: .053

ITEM	BAS	FAS
Nurse	.550	.380
Sick	.030	.050
Lawyer	.150	.100
Medicine	.150	.070
Health	.050	.020
Hospital	.030	.010
Dentist	.210	.020
Physician	.800	.040
Ill	.000	.010
Patient	.370	.030
Office	.010	.010
Stethoscope	.520	.000
Surgeon	.480	.040
Clinic	.300	.000

(iii) Low BAS—High FAS

Critical Lure: Lamp
Label: Things that help you see
MBAS: .006 **MFAS**: .066

ITEM	BAS	FAS
Light	.020	.769
Shade	.028	.058
Bulb	.014	.045
Post	.000	.026
Black	.000	.000
Cord	.000	.000
Desk	.034	.019
Bright	.000	.000
Lighter	.000	.000
Read	.000	.000
On	.000	.000
Burn	.000	.013
Pole	.000	.000
Stand	.000	.000

Critical Lure: Cabbage
Label: Something that you eat
MBAS: .012 **MFAS**: .051

ITEM	BAS	FAS
Head	.000	.022
Lettuce	.021	.281
Vegetable	.000	.137
Food	.000	.022
Salad	.000	.022
Green	.000	.079
Garden	.000	.000
Leaf	.000	.029
Sauerkraut	.042	.000
Slaw	.041	.043
Patch	.066	.115
Plant	.000	.000
Carrots	.000	.000
Soup	.000	.014

2. Category word lists and labels

(i) High BAS—Low FAS

Critical Lure: Apple
Label: Fruit
MBAS: .051 **MFAS**: .018

ITEM	BAS	FAS
Orange	.080	.170
Banana	.150	.020
Grape	.030	.000
Pear	.250	.050
Peach	.060	.030
Strawberry	.000	.000
Kiwi	.000	.000
Pineapple	.000	.000
Watermelon	.000	.000
Tomato	.010	.000
Plum	.050	.000
Grapefruit	.010	.000
Mango	.000	.000
Cherry	.080	.000

Critical Lure: Car
Label: Transportation vehicle
MBAS: .117 **MFAS**: .020

ITEM	BAS	FAS
Bus	.250	.020
Truck	.260	.110
Airplane	.030	.010
Train	.060	.010
Bicycle	.040	.080
Van	.450	.000
Boat	.000	.060
Motorcycle	.090	.000
Skateboard	.000	.000
Subway	.030	.000
Taxi	.130	.000
Scooter	.000	.000
Jeep	.300	.010
Helicopter	.000	.000

(ii) High BAS—High FAS

Critical Lure: Instrument
Label: Musical things
MBAS: .125 **MFAS**: .039

ITEM	BAS	FAS
Drums	.010	.050
Guitar	.050	.150
Flute	.180	.080
Piano	.040	.160
Trumpet	.180	.070
Clarinet	.120	.030
Saxophone	.240	.020
Violin	.060	.020
Trombone	.170	.000
Tuba	.250	.010
Cello	.150	.000
Oboe	.180	.000
Bass	.030	.000
Harp	.080	.000

Critical Lure: Chair
Label: Pieces of Furniture
MBAS: .169 **MFAS**: .037

ITEM	BAS	FAS
Table	.760	.310
Couch	.290	.110
Bed	.000	.010
Desk	.290	.020
Sofa	.130	.080
Dresser	.000	.000
Loveseat	.000	.000
Nightstand	.000	.000
Ottoman	.000	.000
Recliner	.550	.000
Stool	.320	.030
Futon	.000	.000
Bookshelf	.000	.000
Cabinet	.020	.000

(iii) Low BAS—High FAS

Critical Lure: Hammer
Label: Carpenter's tools
MBAS: .089 **MFAS**: .056

ITEM	BAS	FAS
Nail	.620	.800
Saw	.100	.030
Screwdriver	.080	.000
Drill	.000	.000
Wrench	.090	.000
Screw(s)	.030	.000
Level	.000	.000
Ruler	.000	.000
Tape Measure	.000	.000
Wood	.000	.000
Sander	.000	.000
Knife	.000	.000
Chisel	.410	.050
Pencil	.000	.000

Critical Lure: Fly
Label: Insect
MBAS: .047 **MFAS**: .021

ITEM	BAS	FAS
Ant	.000	.000
Spider	.190	.020
Bee	.010	.010
Mosquito	.050	.040
Beetle	.000	.000
Ladybug	.000	.000
Grasshopper	.010	.000
Butterfly	.000	.010
Wasp	.020	.000
Roach	.000	.010
Moth	.110	.000
Gnat	.230	.000
Cockroach	.000	.000
Caterpillar	.030	.000

Appendix B

DRM and category lists used in Experiment 2

DRM lists and their backward associative strength

Items removed	List and label	CBAS	RBAS
	MAN: Things to do with people		
None	*Woman*, husband, uncle, **lady**, mouse, **male, father**, strong, friend, **beard**, person, handsome, muscle, suit, old	1.72	N/A
Strong	Husband, uncle, lady, mouse, male, father, strong, friend, beard, person, handsome, muscle, suit, old	1.12	.60
Weak	Woman, husband, uncle, mouse, strong, friend, person, handsome, muscle, suit, old	1.12	.60(.37,.13,.05,.05)
	COLD: Things to do with temperature		
None	*Hot*, snow, warm, winter, ice, **wet**, **frigid**, chilly, heat, weather, freeze, air, shiver, arctic, frost	5.30	N/A
Strong	Snow, warm, winter, ice, wet, frigid, chilly, heat, weather, freeze, air, shiver, arctic, frost	4.62	.68
Weak	Hot, snow, warm, winter, ice, chilly, heat, weather, freeze, air, shiver, arctic, frost	4.62	.68 (.11,.57)
	DOCTOR: Medical things		
None	Nurse, sick, lawyer, medicine, **health**, **hospital**, **dentist**, *physician*, ill, patient, office, **stethoscope**, surgeon, clinic, cure	3.68	N/A
Strong	Nurse, sick, lawyer, medicine, health, hospital, dentist, ill, patient, office, stethoscope, surgeon, clinic, cure	2.88	.80
Weak	Nurse, sick, lawyer, medicine, physician, ill, patient, office, surgeon, clinic, cure	2.88	.80(.05,.52,.03,.21)
	SHIRT: Things you wear		
None	*Blouse*, sleeves, pants, **tie**, button, shorts, iron, **polo**, **collar**, vest, **pocket**, jersey, belt, linen, cuffs	2.89	N/A
Strong	Sleeves, pants, tie, button, shorts, iron, polo, collar, vest, pocket, jersey, belt, linen, cuffs	2.24	.65
Weak	Blouse, sleeves, pants, button, shorts, iron, vest, jersey, belt, linen, cuffs	2.24	.65(.34,.18,.07,.06)
	SLEEP: Things you do lying down		
None	Bed, rest, awake, tired, dream, **wake**, snooze, **blanket**, doze, slumber, **snore**, *nap*, peace, yawn, dozy	6.47	N/A
Strong	Bed, rest, awake, tired, dream, wake, snooze, blanket, doze, slumber, snore, peace, yawn, dozy	5.74	.73
Weak	Bed, rest, awake, tired, dream, snooze, doze, slumber, nap, peace, yawn, dozy	5.74	.73(.02,.44,.28)
	LION: Animal things		
None	**Tiger**, circus, jungle, tamer, den, cub, Africa, **mane**, cage, feline, *roar*, **fierce**, bears, hunt, pride	2.02	N/A
Strong	Tiger, circus, jungle, tamer, den, cub, Africa, mane, cage, feline, fierce, bears, hunt, pride	1.41	.61
Weak	Circus, jungle, tamer, den, cub, Africa, cage, feline, roar, bears, hunt, pride	1.41	.61(.11,.20,.31)

Category lists and their backward associative strength

Items	Removed list and label	CBAS	RBAS
	ROBIN: Birds		
None	Eagle, bluejay, **cardinal**, hawk, crow, hummingbird, parrot, *sparrow*, pigeon, seagull, dove, **duck**, falcon, canary, **owl**	7.27	N/A
Strong	Eagle, bluejay, cardinal, hawk, crow, hummingbird, parrot, pigeon, seagull, dove, duck, falcon, canary, owl	6.52	.75
Weak	Eagle, bluejay, hawk, crow, hummingbird, parrot, sparrow, pigeon, seagull, dove, falcon, canary	6.52	.75(.47,.10,.18)
	APPLE: Fruits		
None	Orange, **banana**, **grape**, *pear*, peach, strawberry, kiwi, pineapple, watermelon, **tomato**, **plum**, grapefruit, mango, cherry, lemon	.73	N/A
Strong	Orange, banana, grape, peach, strawberry, kiwi, pineapple, watermelon, tomato, plum, grapefruit, mango, cherry, lemon	.48	.25
Weak	Orange, pear, peach, strawberry, kiwi, pineapple, watermelon, grapefruit, mango, cherry, lemon	.49	.24(.15,.05,.03,.01)
	INSTRUMENT: Musical things		
None	Drums, guitar, flute, **piano**, trumpet, clarinet, saxophone, violin, trombone, *tuba*, bass guitar, cello, **oboe**, **bass**, viola	1.79	N/A
Strong	Drums, guitar, flute, piano, trumpet, clarinet, saxophone, violin, trombone, bass guitar, cello, oboe, bass, viola	1.54	.25
Weak	Drums, guitar, flute, trumpet, clarinet, saxophone, violin, trombone, tuba, bass guitar, cello, viola	1.54	.25(.04,.03,.18)
	FLY: Insects		
None	Ant, **spider**, **bee**, mosquito, beetle, ladybug, **grasshopper**, butterfly, **wasp**, roach, moth, *gnat*, cockroach, caterpillar, bug	.71	N/A
Strong	Ant, spider, bee, mosquito, beetle, ladybug, grasshopper, butterfly, wasp, roach, moth, cockroach, caterpillar, bug	.48	.23
Weak	Ant, mosquito, beetle, ladybug, butterfly, roach, moth, gnat, cockroach, caterpillar, bug	.48	.23(.19,.01,.02,.01)
	HAMMER: Carpenter's tools		
None	*Nail*, **saw**, **screwdriver**, drill, wrench, **screw**, level, ruler, tape measure, wood, sander, knife, **chisel**, pencil, sandpaper	1.33	N/A
Strong	Saw, screwdriver, drill, wrench, screw, level, ruler, tape measure, wood, sander, knife, chisel, pencil, sandpaper	.71	.62
Weak	Nail, drill, wrench, level, ruler, tape measure, wood, sander, knife, pencil, sandpaper	.71	.62(.41,.10,.08,.03)
	CAR: Transportation vehicles		
None	Bus, truck, airplane, train, bicycle, *van*, boat, **motorcycle**, **Ford**, skateboard, subway, taxi, scooter, helicopter, jeep	2.00	N/A
Strong	Bus, truck, airplane, train, bicycle, boat, motorcycle, Ford, skateboard, subway, taxi, scooter, helicopter, jeep	1.55	.45
Weak	Bus, truck, airplane, train, bicycle, van, boat, skateboard, subway, taxi, scooter, helicopter, jeep	1.55	.45(.36,.09)

Note: CBAS: cumulative backward associative strength; RBAS: removed backward associative strength; italicized items were removed in strong lists; bolded items were removed in weak lists.

Appendix C

Semantic density dominant lists

Critical lure: green		Critical lure: crime		Critical lure: number	
Grass	.36	Punishment	.17	Math	.18
Yellow	.18	Criminal	.08	Nine	.14
Bean	.15	Murder	.05	Ten	.12
Blue	.14	Robbery	.04	Three	.10
Vegetables	.10	Steal	.04	Two	.07
Nature	.05	Prison	.02	One	.05
Tree	.05	Justice	.02	Phone	.05
Colour	.04	Jail	.02	Five	.05
Leaf	.04	Bad	.01	Eight	.03
Apple	.03	Thief	.01	Address	.02
Red	.02	Burglar	.01	Letter	.02
BAS = 2.16; *SD* = 3.17		BAS = 1.47; *SD* = 3.20		BAS = 1.83; *SD* = 3.52.	

Backward associative strength dominant lists

Critical lure: soft		Critical lure: hair		Critical lure: work	
Hard	.56	Brush	.44	Labor	.69
Pillow	.24	Lice	.37	Job	.57
Gentle	.20	Comb	.31	Earn	.20
Smooth	.18	Shampoo	.29	School	.13
Cotton	.17	Dryer	.20	Office	.13
Skin	.16	Head	.19	Sweat	.11
Silk	.09	Blonde	.16	Play	.07
Touch	.06	Style	.11	Money	.03
Feather	.05	Brown	.06	Fun	.02
Warm	.05	Cut	.06	Hard	.01
Comfortable	.02	Hat	.06	Lazy	.01
BAS = 2.78; *SD* = 1.15		BAS = 3.25; *SD* = 1.91		BAS = 2.97; *SD* = 1.53.	

Note: BAS = backward associative strength; *SD* = semantic density; the number associated with each word is that word's backward associative strength to the critical lure. Activation strength (AS) = BAS + *SD*.

References

Anderson, J. R. (1983). A spreading activation theory of memory. *Journal of Verbal Learning and Verbal Behavior, 22*, 261–295.

Anderson, J. R., & Bower, G. H. (1973). *Human associative memory*. Oxford, UK: Winston.

Bauer, P. J. (2004). Getting explicit memory off the ground: Steps toward construction of a neuro-developmental account of changes in the first two years of life. *Developmental Review, 24*, 347–373.

Bauer, P. J. (2005). Developments in declarative memory: Decreasing susceptibility to storage failure over the second year of life. *Psychological Science, 16*, 41–47.

Bjorklund, D. F. (1987). How changes in knowledge base contribute to the development of children's memory. *Developmental Review, 7*, 93–130.

Bjorklund, D. F. (2004). *Children's thinking: Developmental and individual differences* (4th edition). Belmont, CA: Wadsworth/Thompson.

Bjorklund, D. F., & Jacobs, J. W. (1985). Associative and categorical processes in children's memory: The role of automaticity in the development of organization in free recall. *Journal of Experimental Child Psychology, 39*, 599–617.

Brainerd, C. J., Yang, C., Reyna, V. F., Howe, M. L., & Mills, B. A. (in press). Semantic processing in "associative" false memory. *Psychonomic Bulletin & Review*.

Brainerd, C. J., Forrest, T. J., Karibian, D., & Reyna, V. F. (2006). Development of the false-memory illusion. *Developmental Psychology, 42*, 962–979.

Brainerd, C. J., & Reyna, V. F. (2005). *The science of false memory*. New York: Oxford University Press.

Brainerd, C. J., Reyna, V. F., & Ceci, S. J. (2008). Developmental reversals in false memory: A review of theory and data. *Psychological Bulletin, 134*, 343–382.

Brainerd, C. J., Reyna, V. F., & Forrest, T. J. (2002). Are young children more susceptible to the false-memory illusion? *Child Development, 73*, 1363–1377.

Brainerd, C. J., & Wright, R. (2005). Forward association, backward association, and the false memory illusion. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 31*, 554–567.

Chiarello, C. (1998). On codes of meaning and the meaning of codes: Semantic access and retrieval within and between hemispheres. In M. Beeman & C. Chiarello (Eds.), *Right hemisphere language communication: Perspectives from cognitive neuroscience* (pp. 141–160). Mahwah, NJ: Erlbaum.

Collins, A. M., & Loftus, E. F. (1975). A spreading-activation theory of semantic processing. *Psychological Review, 82*, 407–428.

Courage, M. L., & Howe, M. L. (2002). From infant to child: The dynamics of cognitive change in the second year of life. *Psychological Bulletin, 128*, 250–277.

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

Dempster, F. N., & Brainerd, C. J. (Eds.). (1995). *Interference and inhibition in cognition*. New York: Academic Press.

Dewhurst, S. A., Bould, E., Knott, L. M., & Thorley, C. (2009). The roles of encoding and retrieval processes in associative and categorical memory illusions. *Journal of Memory and Language, 60*, 154–164.

Frankel, M. T., & Rollins, H. A. (1985). Associative and categorical hypotheses of organization in the free recall of adults and children. *Journal of Experimental Child Psychology, 40*, 304–318.

Gallo, D. A. (2006). *Associative illusions of memory: False memory research in DRM and related tasks*. New York: Psychology Press.

Gallo, D. A., & Roediger, H. L. III. (2002). Variability among word lists in eliciting memory illusions: Evidence for associative activation and monitoring. *Journal of Memory and Language, 47*, 469–497.

Hamilton, W. (1859). *Lecture on metaphysics and logic*. New York: Sheldon and Company.

Hege, A. C. G., & Dodson, C. S. (2004). Why distinctive information reduces false memories: Evidence for both impoverished relational-encoding and distinctiveness heuristic accounts. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 30*, 787–795.

Holliday, R. E., Reyna, V. F., & Brainerd, C. J. (2008). Recall of details never experienced: Effects of age, repetition, and semantic cues. *Cognitive Development, 23*, 67–78.

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

Howe, M. L. (2005). Children (but not adults) can inhibit false memories. *Psychological Science, 16*, 927–931.

Howe, M. L. (2006). Developmentally invariant dissociations in children's true and false memories: Not all relatedness is created equal. *Child Development, 77*, 1112–1123.

Howe, M. L. (2008a). Visual distinctiveness and the development of children's false memories. *Child Development, 79*, 65–79.

Howe, M. L. (2008b). What is false memory development the development of? Comment on Brainerd, Reyna, and Ceci (2008). *Psychological Bulletin, 134*, 768–772.

Howe, M. L., Brainerd, C. J., & Kingma, J. (1985). Development of organization in recall: A stages-of-learning analysis. *Journal of Experimental Child Psychology, 39*, 230–251.

Howe, M. L., Cichetti, D., Toth, S. L., & Cerrito, B. M. (2004). True and false memories in maltreated children. *Child Development, 75*, 1402–1417.

Howe, M. L., Wimmer, M. C., & Blease, K. (in press). The role of associative strength in children's false memory illusions. *Memory*.

Howe, M. L., Gagnon, N., & Thouas, L. (2008). Development of false memories in bilingual children and adults. *Journal of Memory and Language, 58*, 669–681.

Hutchison, K. A. (2003). Is semantic priming due to association strength or feature overlap? A microanalytic review. *Psychonomic Bulletin & Review, 10*, 785–813.

Hutchison, K. A., & Balota, D. A. (2005). Decoupling semantic and associative information in false memories: Explorations with semantically ambiguous and unambiguous critical lures. *Journal of Memory and Language, 52*, 1–28.

Kimball, D. R., & Bjork, R. A. (2002). Influences of intentional and unintentional forgetting on false memories. *Journal of Experimental Psychology: General, 131*, 116–130.

Kimball, D. R., Smith, T. A., & Kahana, M. J. (2007). The fSAM model of false recall. *Psychological Review, 114*, 954–993.

Kiss, G., Armstrong, C., Milroy, R., & Piper, J. (1973). An associative thesaurus of English and its computer analysis. Available from http://www.eat.rl.ac.uk/.

Landauer, T. K., & Dumais, S. T. (1997). A solution to Plato's problem: The latent semantic analysis theory of acquisition, induction, and the representation of knowledge. *Psychological Review, 104*, 211–240.

Lange, G. W. (1978). Organization-related processes in children's recall. In P. A. Ornstein (Ed.), *Memory development in children* (pp. 101–128). Hillsdale, NJ: Erlbaum.

McDermott, K. B. (1996). The persistence of false memories in list recall. *Journal of Memory and Language, 35*, 212–230.

McEvoy, C. L., Nelson, D. L., & Komatsu, T. (1999). What is the connection between true and false memories? The differential roles of interitem associations in recall and recognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 25*, 1177–1194.

Metzger, R. L., Warren, A. R., Shelton, J. T., Price, J., Reed, A. W., & Williams, D. (2008). Do children "DRM" like adults? False memory production in children. *Developmental Psychology, 44*, 169–181.

Nelson, D. L., McEvoy, C. L., & Schreiber, T. A. (1999). The University of South Florida Word Association, Rhyme, and Word Fragment Norms. Available from http://w3.usf.edu/FreeAssociation.

Nelson, D. L., Fisher, S. L., & Akirmak, U. (2007a). How implicitly activated and explicitly acquired knowledge contribute to the effectiveness of retrieval cues. *Memory & Cognition, 35*, 1892–1904.

Nelson, D. L., Goodmon, L. B., & Akirmak, U. (2007b). Implicitly activated memories are associated to general context cues. *Memory & Cognition, 35*, 1878–1891.

Nelson, D. L., & Zhang, N. (2000). The ties that bind what is known to the recall of what is new. *Psychonomic Bulletin & Review, 7*, 604–617.

Odegard, T. N., Holliday, R. E., Brainerd, C. J., & Reyna, V. F. (2008). Attention to global gist processing eliminates age differences in false memories. *Journal of Experimental Child Psychology, 99*, 96–113.

Park, L., Shobe, K. K., & Kihlstrom, J. F. (2005). Associative and categorical relations in the associative memory illusion. *Psychological Science, 16*, 792–797.

Reyna, V. F., & Lloyd, F. (1997). Theories of false memory in children and adults. *Learning and individual differences, 9*, 95–123.

Robinson, K. J., & Roediger, H. L. III, (1997). Associative processes in false recall and false recognition. *Psychological Science, 8*, 231–237.

Roediger, H. L., III, Balota, D. A., & Watson, J. M. (2001a). Spreading activation and the arousal of false memories. In H. L. Roediger, III, J. S. Nairne, I. Neath, & A. M. Surprenant (Eds.), *The nature of remembering: Essays in honor of Robert G. Crowder* (pp. 95–115). Washington, DC: American Psychological Association.

Roediger, H. L., III, & McDermott, K. B. (1995). Creating false memories: Remembering words not presented on lists. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 21*, 803–814.

Roediger, H. L., III, Watson, J. M., McDermott, K. B., & Gallo, D. A. (2001b). Factors that determine false recall: A multiple regression analysis. *Psychonomic Bulletin & Review, 8*, 385–407.

Schneider, W., & Pressley, M. P. (1997). *Memory development between two and twenty*. Mahwah, NJ: Erlbaum.

Seamon, J. G., Luo, C. R., Schlegel, S. E., Greene, S. E., & Goldenberg, A. B. (2000). False memory for categorized pictures and words: The category associates procedure for studying memory errors in children and adults. *Journal of Memory and Language, 42*, 120–146.

Smith, S. M., Ward, T. B., Tindell, D. R., Sifonis, C. M., & Wilkenfeld, M. J. (2000). Category structure and created memories. *Memory & Cognition, 28*, 386–395.

Stuart, M., Masterson, J., Dixon, M., & Quilan, P. (1993-1996). Children's printed word database. Available from http://www.essex.ac.uk/psychology/cpwd/.

Sugrue, K., & Hayne, H. (2006). False memories produced by children and adults in the DRM paradigm. *Applied Cognitive Psychology, 20*, 625–631.

Talmi, D., & Moscovitch, M. (2004). Can semantic relatedness explain the enhancement of memory for emotional events? *Memory & Cognition, 32*, 742–751.

Tse, C.-S., & Neely, J. H. (2007). Semantic and repetition priming effects for Deese/Roediger–McDermott (DRM) critical items and associates produced by DRM and unrelated study lists. *Memory & Cognition, 35*, 1047–1066.

Underwood, B. J. (1965). False recognition produced by implicit verbal responses. *Journal of Experimental Psychology, 70*, 122–129.

Van Overschelde, J. P., Rawson, K. A., & Dunlosky, J. (2004). Category norms: An updated and expanded version of the Battig and Montague (1969) norms. *Journal of Memory and Language, 50*, 289–335.

Wu, L. L., & Barsalou, L. W. (2007). Grounding concepts in perceptual simulation: I. Evidence from feature generation, submitted for publication.