Development of False Memories in Bilingual Children and Adults

Mark L. Howe[1], Nadine Gagnon[2], and Lisa Thouas[3]

[1]Lancaster University, Lancaster, United Kingdom

[2]University of British Columbia, Vancouver, British Columbia, Canada

[3]University of Manitoba, Winnipeg, Manitoba, Canada

Send correspondence to: Mark L. Howe
Centre for Research in Human Development
Department of Psychology
Lancaster University
Lancaster, U.K. LA1 4YF
e-mail: mark.howe@lancaster.ac.uk

IN PRESS: *Journal of Memory and Language*.

Abstract

The effects of within- versus between-languages (English-French) study and test on rates of bilingual children's and adults' true and false memories were examined. Children aged 6 through 12 and university-aged adults participated in a standard Deese-Roediger-McDermott false memory task using free recall and recognition. Recall results showed that: (1) both true and false memories increased with age, (2) true recall was higher in within- than between-languages conditions for all ages, and (3) there were fewer false memories in between-languages conditions than within-language conditions for the youngest children, no differences for the 8- and 12-year-olds, and by adulthood, there were more false memories in between-languages than within-language conditions. Recognition results showed that regardless of age, false recognition rates tended to be higher in between-languages than within-language conditions. These findings are discussed in the context of models of false memory development.

Keywords: DRM paradigm, false memories, bilingual memory, memory development, children's false memory.

Development of False Memories in Bilingual Children and Adults

Past studies using the Deese-Roediger-McDermott (DRM) paradigm (Deese, 1959; Roediger & McDermott, 1995) have demonstrated that young children may be less susceptible to false memories than adults (Brainerd, Reyna, & Forrest, 2002; Howe, 2005, 2006, in press; Howe, Cicchetti, Toth, & Cerrito, 2004; but see Ghetti, Qin, & Goodman, 2002). Theoretical explanations such as fuzzy-trace theory (FTT) have attempted to account for these findings by suggesting that both verbatim and gist traces are encoded during list presentation (Brainerd & Reyna, 2005). These traces are qualitatively different from each other where the former is concerned with item-specific surface information and the latter with meaning-based information. It is this gist trace that is thought to be responsible for false recall in the DRM paradigm. Although young children are capable of extracting these types of traces, the ability to do so improves with age and cognitive development into later childhood and early adulthood (Brainerd & Reyna, 2005).

An alternative, associative-activation model (e.g., Howe, 2005, 2006, in press) argues that developmental trends in false memories occurs not just because of correlated changes in children's meaning extraction skills (i.e., growth in knowledge base) but also because of increased automaticity in the activation and accessibility of those concepts and associations in the child's knowledge base. These increases in automaticity are brought about by additional exposure and proficiency using these concepts as well as the associations between related concepts. As exposure and proficiency increases, so too does the automaticity of activation of concepts and their associative links, making false memory production more and more adult-like, occurring without conscious effort of awareness (Howe, 2005).

Although these models have been contrasted in other forums (Howe, 2005, 2006, in press), a key test of these positions can be found by considering the development of bilingual memory. Bilingual children may be able to use meaning more effectively if they can access semantic representations through two different lexical forms. However, this may come at a cost as increased processing of meaning in young children may increase their levels of false recollection. This is exactly what has been found in studies with bilingual adults. Specifically, when adults study a list in one language (e.g., English or Spanish) and are asked to recall or recognize those items in the other language (i.e., English (Spanish or Spanish (English), true memory performance is reduced and false memory performance is enhanced relative to within-language study-test conditions (i.e., English (English or Spanish (Spanish). For example, Marmolejo, Dilberto-Macaluso, and Altarriba (2003) found that for Spanish-English bilinguals, there was lower true recall and higher false recall in between-languages study and test conditions than in within-language conditions. Similar findings were reported by Wakeford, Carlin, and Toglia (2005) with English-Spanish bilinguals using both recall and recognition measures. Using a 2(Acquisition: English, Spanish) x 2(Test: English, Spanish) design, Wakeford et al. (2005) found that veridical recall and recognition rates were lower in between-languages conditions than within-language conditions. Using a similar design, Sahlin, Harding, and Seamon (2005) found false recognition increased in between-languages as opposed to within-language conditions for English-Spanish bilingual adults. However, as additional study-test trials were provided, participants' false recognition rates decreased as they came to rely more on language-specific lexical representations rather than on the conceptual representations used initially. Finally, Cabeza and Lennartson (2005), who used a paradigm similar to the previous ones but testing English-French bilingual adults, found that false recognition was robust in both within-language and between-languages conditions, but more correct recognition of "old" items occurred in the within-language than between-languages conditions.

All of these findings with bilingual adults have been attributed to increased meaning processing in the between-languages study-test conditions than in the within-language conditions (e.g., Cabeza & Lennartson, 2005). Consistent with this is the finding that as the number of study-test trials increases, participants' reliance on meaning decreases because stronger cue- or language-specific representations are available, and both of these events are associated with lowered false recollection (Sahlin et al., 2005). Taken together, these findings have been seen by some as support for FTT (Brainerd, Forrest, Karibian, & Reyna, 2006). The argument is that presentation of a list of words in one language (e.g., English) and requiring output (recall, recognition) in a different language (e.g., French, Spanish) makes it (a) less likely that bilingual adults will be able to access verbatim traces to support true memories, (b) less likely that bilingual adults will be able to access verbatim traces to suppress false memories, and (c) more likely that bilingual adults will access gist traces that support false memories. Because of the dual effect of decreases in verbatim memory (a and b), coupled with increases in gist processing (c), false memories are more likely in between-languages than within-language study-test conditions.

However, these outcomes with adult participants do not provide necessary and sufficient support for FTT. This is because these results are also consistent with predictions from associative-activation models (e.g., Howe, 2005; Hutchison & Balota, 2005). Like FTT, these latter models predict that increased meaning processing, along with decreased discriminability of the original stimulus, both circumstances that pertain to between-languages conditions but not within-language conditions, will lead to fewer true memories and more false memories. These theories do, however, make different predictions when we consider the *development* of false memories in bilingual children.

To see how these theories can be distinguished, first consider what FTT predicts about bilingual children's false memories. If FTT is correct, relative increases and decreases in children's true and false memories depend on the extent to which between-languages study-test conditions prompt bilingual children to increase gist processing at the expense of verbatim memory. Specifically, according to FTT, the development of children's false memories is related to their increased ability to extract gist. To the extent that between-languages memory tasks increase gist processing and decrease reliance on verbatim memory, FTT predicts that bilingual children should produce more false memories in between-languages study-test conditions than within-language ones. Of course, whether this task increases children's gist processing depends on the development and organization of bilinguals' mental lexicon. According to Silverberg and Samuel (2004), the age of acquisition of an additional language can play an important role in the organization of the mental lexicon such that only early learners (acquisition prior to the age of 7) possess a shared conceptual/semantic store for both languages (L1 and L2) whereas those of proficient late learners remain separate, making conceptual mediation in the second learned language (L2) impossible without translating it into the first learned language (L1). Although their conceptual stores remain unconnected, proficient late learners possess a shared lexical store for both languages where similar lexical forms compete with each other independent of language (Silverberg & Samuel, 2004). Thus regardless of high proficiency in L2, the organization of the late learners' bilingual memory can never resemble that of an early bilingual.

Other models of bilingual development assume either that there is a single conceptual store or if separate stores do exist, they are overlapping. In Kroll and Stewart's (1994) revised hierarchical model, beginning L2 learners' access conceptual representations in their nondominant language by initially translating to L1. However, conceptual mediation in L2 is feasible for highly proficient learners but reliance on L2-L1 lexical links to access conceptual meaning remain nonetheless (French & Jacquet, 2004). Kroll and Stewart (1994) demonstrated that even highly fluent bilinguals possess asymmetrical strength between L1-L2 lexical and conceptual links, where conceptual links are stronger for L1 than for L2 and L2-L1 lexical links are stronger than L1-L2 links. Thus, in this model both strength and direction are important factors in bilinguals' memory performance, where L2 learners' performance may be affected by the direction of the translation or the strength of the association between L2 and the conceptual store.

Although the revised hierarchical model has been supported in bilingual research (McElree, Jia, & Litvak, 2000; Talamas, Kroll, & Dufour, 1999), others have reported that even novice L2 learners are capable of conceptual mediation in L2 suggesting that semantic information is stored and accessed even by late, low proficiency L2 learners (Altarriba & Mathis, 1997). These findings may suggest an alternate path in the development and organization of bilingual memory, one in which both lexical and conceptual stores, as well as the connections between them, continue to grow as proficiency and exposure increase. Overall, then, the literature on bilingual development supports the idea that even young second-language learners are capable of increased gist processing given a between-languages task. Thus, to the extent that between-languages memory tasks reduce reliance on verbatim traces and increase gist processing, FTT predicts increases in children's false memory rates in between-languages than within-language study-test conditions. Such increases may be more likely in older than younger children given older children's better gist processing more generally. However, to the extent that younger children have greater difficulty spontaneously extracting gist than older children, between-languages study-test conditions might just boost younger children's gist processing too, leading to larger false memory "gains" for younger than older children.

For FTT, then, the predictions are that for false recall: (a) older children and adults will evince more false memories than younger children regardless of language condition and (b) all participants should show higher rates of false memories in between-languages than within-language conditions. For true recall, because within-language conditions preserve verbatim information, true recall rates should be higher in these conditions than the between-languages conditions for all participants regardless of age.

As mentioned earlier, Howe (2005, 2006, in press) has pointed out that increases in children's false memory with age may not be due simply to increases in processing meaning (gist) but rather to increases in the automaticity with which children access or activate associations in their knowledge base, associations that mediate false remembering (for a similar model of adults false recollection, see Hutchison & Balota, 2005). Automaticity, particularly in children's recall, may play a role in the rate of between-languages false memories where highly proficient bilinguals may be more skilled in accessing concepts, as well as the associative and categorical relations among them, in both languages (Kotz & Elston-Guttler, 2004). Dufour and Kroll (1995) demonstrated that more fluent bilinguals are equally rapid when categorizing words within and between languages as compared to less fluent bilinguals who demonstrated different categorization speeds depending on the target's language. This has also been confirmed by McElree et al. (2000) who examined the speed of processing in balanced and unbalanced bilinguals and determined that access to conceptual information in both languages was *highly dependent on exposure and proficiency*. This increase in speed of translation may affect rates of true and false memories in the DRM task, where distinctive information may be lost at the cost of L2 automaticity, resulting in more false memories in conditions that require accessing the concept through two distinct forms (dual processing) for bilinguals equally proficient in both languages.

Although FTT and the associative-activation accounts frequently make similar predictions, they do so for different reasons. For example, both models agree that true recall should be better in within-language than between-languages conditions and that both true and false recall rates should increase with age. FTT makes these predictions because of changes in the relative involvement of gist and verbatim memories and the associative-activation account predicts this because of increases in the automaticity of access to memory traces and their interitem links. However, these account do diverge when making predictions about the development of false memories in bilingual processing tasks. Unlike FTT, the associative-activation account requires not only that the conceptual structures associated with both languages be present in a participant's knowledge base, but also that the processes supporting the activation of these concepts and their interitem associations be automatic. As Howe (2005) showed, despite the fact that concepts used in DRM tasks are present in young children's knowledge bases, their access to them (as well as interitem links) is not as automatic as that found in adults. Because higher rates of false memories are associated with increased automaticity of activation (e.g., Gallo & Seamon, 2004; Raaijmakers & Zeelenberg, 2004; Roediger, Balota, & Watson, 2002), even though bilingual concepts exist in children's knowledge base their mere existence does not guarantee increases in false memories. It is only when access is automatic that increases in false memories will occur. Indeed, as already seen, this automaticity in access to bilingual memory comes about with additional exposure and proficiency in both languages (McElree et al., 2000). Because this automaticity is more likely in adults than young children, the associative-activation account, in contrast to FTT, predicts more false memories in between-languages than within-language conditions for adults but not young children. That is, while FTT predicts higher false recall for between-languages than within-language conditions regardless of age, the associative-activation account predicts a cross-over interaction with age.

In the current study, we tested these different predictions by examining true and false memories in bilingual individuals varying in age (and, hence, bilingual exposure and proficiency) using a bilingual DRM task. As noted, both FTT and the associative-activation accounts predict true recall should be better in the within-language than the between-languages conditions and that true and false recall rates should increase with age. Where they differ is with false recall. According to FTT, bilingual individuals should have more false memories in the between-languages conditions (when translation is required) than in the within-language conditions. FTT also predicts increases in children's false recollection rates when gist processing is made easier and these increases should be more evident with children whose gist processing is particularly compromised (i.e., the youngest children). Alternatively, associative-activation models predict that increases in false recollection occur depending on the relative automaticity of the associative connections underlying false memories and given that automaticity is correlated with exposure and proficiency, between-languages false memories should exceed within-language false memories more frequently in older, more proficient bilinguals. That is, age-related differences in automaticity may mean that only older, adult participants should exhibit higher rates of false memories than younger participants even though conceptual-level representations for both languages overlap very early in bilingual development.

In addition to recall, we also examined bilingual participants' recognition performance. These recognition tests provide a validity check inasmuch as they assess (a) whether children and adults have access to the concepts used in the DRM task in both languages, (b) that this access is similar across languages both within age groups as well as across age groups, and (c) whether children and adults have access to the originally presented items in the language in which they were originally presented. Of course, because recognition performance depends more on matching the probe with the same representation in memory, all theories would predict that within-language true recognition rates should exceed between-languages rates as the item was experienced twice in the same language in the former but not the latter condition.

More importantly, both FTT and the associative-activation accounts have different predictions when it comes to recognition. According to FTT, because recognition tests provide explicit retrieval cues for verbatim traces that are not found in recall tests, they are more likely to tap verbatim traces of targets than free recall tasks where cues, if any, are generated internally and emphasize gist processing (see Brainerd & Reyna, 2005). What this means is that true recognition should be higher and false recognition lower in within-language conditions. For between-languages conditions, because participants have experienced concepts in both languages (one at study and the other at test), recognition cues, regardless of language, should tap at least one of the verbatim traces. This could mean either that true recognition should again exceed false recognition performance or the true and false recognition should be approximately equal depending on the extent to which verbatim processing has been compromised given the greater likelihood of gist processing in between-language conditions. Finally, true recognition should be higher and false recognition rates lower in the within-language conditions than in the between-languages conditions because the same verbatim trace has been presented twice in the former but not the latter conditions. This item type (true vs. false) x condition (within-language vs. between-languages) interaction is similar to the findings mentioned earlier in which additional within-language repetition for English-Spanish bilinguals decreased false recognition rates (Sahlin et al., 2005).

In contrast, for the associative-activation model, although the cues for memory performance are more explicit in recognition tests, performance should resemble that found at recall. That is, if bilingual children and adults are accessing overlapping conceptual representations (Altarriba & Mathis, 1997) in between-languages conditions but language cues are no longer diagnostic (i.e., they do not provide distinctive information that can be used to discriminate presented from unpresented items in memory), then there should be more confusion as to what was and what was not presented in the between-languages than within-language conditions. What this means is that false recognition should be higher and true recognition lower in between-languages conditions than within-language conditions. Moreover, although true and false recognition rates may not differ in within-language conditions, false recognition rates should be higher than true recognition rates in between-languages conditions.

Method

Participants

A total of 408 children [229 girls and 179 boys; 160 Grade 1-2 students (6-year-olds); 128 Grade 3-4 students (8-year-olds); 120 Grade 5-6 students (12-year-olds)] were recruited from local French-Immersion schools in a Canadian city. All children (predominantly White and middle class) were registered in the Early French-Immersion program where complete submersion in a French speaking environment begins in kindergarten (age 5 years) and slowly tapers off to an even split between French and English instructions in grade 6 (11-12 years old). The majority of the children came from English speaking households where the classroom was the sole French-speaking environment to which children were exposed. Due to the formal nature of classroom instruction and the frequency of exposure to a French environment it is unlikely that the participants were balanced bilinguals. Eighty undergraduate students (40 males and 40 females; 20-year-olds) were recruited through undergraduate French courses at the local university. Individuals were selected on the basis of fluency in both English and French. Specifically, all participants were living in an English community and had passed a French proficiency test. In addition, many were working in an almost entirely French environment.

Materials

Eight DRM lists were selected from Stadler, Roediger, and McDermott (1999). Selection of DRM lists was based on both appropriateness for young children and ease of translation. Lists from popular categorical themes (e.g., *weather*, *fruit*, *clothing*) were sought to ensure that children were familiar with the items. Each list was translated into French resulting in a total of 16 14-item DRM lists (see Appendix). The DRM lists were read aloud by a female speaker in English and French and recorded onto audiotape. It was felt by the investigators that the suprasegmental features were respected for both English and French lists in the audio recording.

Procedure

This design yielded a total of 4 conditions: 2 within-language conditions (English study-English recall, French study-French recall) and 2 between-languages conditions (English study-French recall, French study-English recall). Like other DRM studies with children (e.g., Howe, in press), all items were presented at a 3-second rate using an audio recorder. After the presentation of the first list, the audio recording was interrupted and the participants were asked to verbally report the previously studied items in the condition-appropriate language. All of the items reported by the participant was recorded by the experimenter. Following the last recalled item, the experimenter confirmed with the participant that no further items could be recalled. Once this was established, the audiotape was restarted for presentation of the next study list. This procedure was repeated for all lists. The elapsed time between list presentations was dependent on the number of items reported by the participant.

After all of the lists had been presented and recalled a simple yes-no recognition task was administered. Test words were presented one at a time on an audiotape and participants were instructed to say "yes" to items that had previously been presented on the study lists regardless of which language they had appeared in previously (study or recall), and to say "no" to items that had not appeared. The test list was composed of 32 targets (4 randomly selected from each studied list), all 8 critical lures, 8 semantically related but unpresented items from the studied lists (the 15th item from each list), and 8 unrelated items from unpresented lists. For half of the participants in each of the 4 study-recall combinations, the recognition items were presented in English. For the remaining half of the participants, items were presented in French. This resulted in two within-language study-recall-recognition conditions (English-English-English, French-French-French) and six between-languages conditions, two of which were within-language until the time of recognition (English-English-French, French-French-English) and four that were varied beforehand (English-French-English, English-French-French, French-English-English, French-English-French).

Results

We begin by reporting the findings for recall and then turn to the recognition results. It should be noted that although the recognition results must be interpreted in the context of prior recall, there was no evidence to suggest that recognition was contaminated by recall. That is, items recalled and recognized did not occur more frequently than recognized items that were not recalled.

Recall

To make contact with previous studies (e.g., Cabeza & Lennartson, 2005), we report the general trends for true and false recall within each language (English-English, French-French) first. This is followed by a comparison of trends in between-languages conditions (English-French, French-English) and within-language conditions (English-English, French-French) conditions.

Within-language Recall. To evaluate levels of true and false recall in the standard within-language conditions, separate 2(Language: English, French) x 4(Age: 6, 8, 12, and 20 years old) analyses of variance (ANOVAs) were conducted for true and false recall. For true recall, the results showed a main effect for language, $F(1, 236) = 58.26$, $p < .001$, (^{2} = .20, where the mean proportion of true recall in English ($M = .41$) was greater than that for French ($M = .33$), as well as a main effect for age, $F(3, 236) = 84.31$, $p < .001$, (^{2} = .52, where post-hoc Newman-Keuls tests ($p < .05$) revealed that 6-year-olds ($M = .25$) recalled less than 8-year-olds ($M = .34$) who recalled less than 12-year-olds ($M = .41$) who recalled less than 20-year-olds ($M = .50$). Finally, there was a Language x Age interaction, $F(3, 236) = 14.02$, $p < .001$, (^{2} = .15, which is shown in Figure 1a. As can be seen, and was confirmed by post-hoc tests, the source of this interaction was focused squarely on adult recall. That is, although the advantage of English over French was the same (qualitatively and quantitatively) for children, there were no differences in true recall rates for adults. Thus, for true recall, the amount recalled increased with age and children's recall was better in English than French, perhaps confirming that they were not balanced bilinguals (see Method). For adults, because there were no English-French differences, it is likely that they were balanced bilinguals.

Turning to false recall, the results indicated that regardless of whether one is a balanced or unbalanced bilingual, more false memories were generated within English than within French (see Figure 1b). The results of the 2(Language: English, French) x 4(Age: 6, 8, 12, and 20 year olds) ANOVA for false recall revealed a main effect for language, $F(1, 236) = 48.81$, $p < .001$, (^{2} = .17, where the mean proportion of false recall was greater in English ($M = .28$) than French ($M = .13$), as well as a main effect for age, $F(3, 236) = 5,82$, $p < .001$, (^{2} = .07, where post-hoc tests revealed the following pattern of age differences: 6-year-olds ($M = .13$) < 8-year-olds ($M = .19$) = 12-year-olds ($M = .21$) < 20-year-olds ($M = .28$). As can be seen in Figure 1b, there was no Language x Age interaction. Thus, age increments in false recall were similar within language regardless of whether that language was English or French, although there were more false memories produced at each age for English than French unilingual conditions. Like other recent findings concerning conceptual-level processing in balanced and unbalanced bilinguals (Duyck & Brysbaert, 2004), it would seem that false recall is more likely within English than within French regardless of age or whether one is a balanced or unbalanced bilingual. This may be because of cross-linguistic differences in the associative or conceptual attributes activated by DRM lists normed in English-speaking populations. This will be addressed later in the Discussion and Conclusions.

Within-Language versus Between-Languages Recall. To examine our main hypotheses concerning the effects of within-language versus between-languages study and test on true and false recall, we conducted two separate ANOVAs, one for true recall and one for false recall. Each analysis consisted of a 2(Language switch: within-language, between-languages) x 4(Age: 6, 8, 12, and 20 years old) ANOVA where the within-language condition was comprised of the English-English and French-French conditions and the between-languages condition was comprised of the English-French and French-English conditions. The results for true recall showed a main effect for language switch, $F(1, 480) = 265.76$, $p < .001$, (^{2} = .36, where the mean proportion of true recall was greater in within-language ($M = .37$) than between-languages ($M = .24$) conditions. There was also a main effect for age, $F(3, 480) = 154.85$, $p < .001$, (^{2} = .49, where post-hoc tests revealed the following pattern of age differences: 6-year-olds ($M = .20$) < 8-year-olds ($M = .27$) < 12-year-olds ($M = .32$) < 20-year-olds ($M = .43$). Finally, as can be seen in Figure 2a, there was no Language Switch x Age interaction. That is, within-language true recall was better than between-languages true recall, age increases in true recall occurred regardless of whether study and test was within-language or between-languages, and the magnitude of the within- versus between-languages recall difference did not vary reliably across age.

Turning to false recall, the results of the ANOVA revealed no main effect for language switch but there was one for age, $F(3, 480) = 24.52$, $p < .001$, (^{2} = .13, where post-hoc tests revealed the following pattern of age differences: 6-year-olds ($M = .12$) < 8-year-olds ($M = .18$) = 12-year-olds ($M = .21$) < 20-year-olds ($M = .31$). More importantly, there was a Language switch x Age interaction, $F(3, 480) = 2.67$, $p < .05$, (^{2} = .02, shown in Figure 2b. As can be seen in this figure, and was confirmed using post-hoc tests, the interaction is a cross-over one in which the youngest children have more false memories in within-language than between-languages conditions, 8- and 12-year-olds have equivalent rates of false recall across these conditions, and 20-year-olds have more false memories in between-languages than within-language conditions.

This pattern of results for true and false recall for adults is the same as that obtained in previous studies with bilingual adults (e.g., Cabeza & Lennartson, 2005; Sahlin et al., 2005). What has not been seen before is the different development trends for true and false recall. Specifically, these results show that developmentally, requiring study and test in two different languages does not affect increases in true recall but does alter the trajectory of false memory production across age. Indeed, the cross-over interaction predicted by the associative-activation model is exactly what can be seen in Figure 2b. Thus, the developmental pattern of false recall obtained here is consistent with the associative-activation model, not FTT. We will return to this point later in the Discussion and Conclusions.

Recognition

Because there can be age differences in yea-saying rates (see Brainerd et al., 2002; Howe et al., 2004), it is usual to deploy the nonparametric signal detection statistic *A'* when analyzing developmental recognition data rather than simply analyze raw recognition scores. Although the validity of *A'* has been questioned and its utility in "correcting" or adjusting raw recognition responses in populations whose yea-saying propensities vary, doubted (e.g., Pastore, Crawley, Berens, & Skelly, 2003), the need to adjust scores in developmental studies exists nonetheless. To make contact with prior studies, we report the findings associated with the raw recognition scores first. To adjust for age trends in yea-saying, we follow each of the raw score analyses with analyses based on the corrected, *A'* scores.[1]

Within-Language Recognition. As before, we begin with an analysis of recognition scores for the within-language conditions. Like Cabeza and Lennartson (2005), we were interested in comparisons between the different recognition items, specifically targets, critical lures, and new (unrelated) items. We conducted a 2(Language: English, French) x 3(Item: target, critical lure, new) x 4(Age: 6, 8, 12, and 20 years old) ANOVA on the raw recognition scores. The results revealed a main effect for item, *F*(2, 234) = 452.82, *p* < .001, (^{2} = .80, where post-hoc Newman-Keuls tests (*p* < .05) showed that recognition rates for targets (*M* = .68) were the same as those for critical lures (*M* = .67), both of which were higher than for new items (*M* = .19). There was an Item x Age interaction, *F*(6, 234) = 8.19, *p* < .001, (^{2} = .17, shown in Figure 3. It can be seen in this figure, and was confirmed by post-hoc testing, that this interaction was due to different developmental trends associated with the different items. Specifically, recognition rates increased with age for targets and critical lures whereas recognition of new items decreased with age. Finally, there was an Item x Language interaction, *F*(2, 234) = 14.04, *p* < .001, (^{2} = .11, shown in Figure 4. It is clear from this figure, and confirmed by post-hoc tests, that target recognition did not differ as a function of language, recognition of critical lures was higher for English than for French, and new items were more often recognized incorrectly in French than in English. Thus, unlike recall, target recognition did not differ between languages or with age. These findings are particularly important as they validate claims that both children and adults share similar lexicons for the DRM lists used in this study regardless of language. Thus, any developmental or between-language differences in true and false memories obtained here cannot be simply attributed to age or language confounds in the availability of DRM concepts in memory.

In contrast to true recognition, false recognition of critical lures, like false recall, was higher in English than in French. Interestingly, incorrect recognition of new, unrelated items was higher in French than in English, a trend opposite to that for critical lures. We will return to this point in the Discussion and Conclusions.

A' scores for targets and critical lures were analyzed using a 2(Language: English, French) x 2(Item: target, critical lure) x 4(Age: 6, 8, 12, and 20 years old) ANOVA. The results were similar to the raw score analyses. That is, there was no difference between correct recognition of targets (*A'* = .80) and false recognition of critical lures (*A'* = .80). There was a main effect for age, *F*(3, 226) = 14.44, *p* < .001, (^{2} = .16, where post-hoc tests revealed the following pattern of reliable differences: 6-year-olds (*A'* = .73) < 8-year-olds (*A'* = .79) < 12-year-olds (*A'* = .82) < 20-year-olds (*A'* = .85). Like the raw score analyses, there were no differences in *A'* as a function of age for correct versus false recognition rates. There was also a main effect for language, *F*(1,

226) = 4.83, *p* < .05, (^{2} = .02, where recognition rates were modestly higher for English (*A'* = .82) than French (*A'* = . 79). Unlike the raw score analysis, there was no Item x Language interaction.

Within-Language versus Between-Languages Recognition. Our next series of analyses parallel those for the recall data as well and focus on our hypotheses concerning true and false recognition within versus between languages. To do this, we conducted a 2(Language shift: within language – English-English-English and French-French-French; versus between languages – English-English-French, English-French-English, English-French-French, French-French-English, French-English-French, French-English-English)[2] x 3(Item: target, critical lure, new) x 4(Age: 6, 8, 12, 20 years old) ANOVA using raw recognition scores. The results revealed a main effect for item, *F*(1, 472) = 723.41, *p* < .001, (^{2} = .75, where post-hoc tests showed that critical lures (*M* = .71) were more frequently (falsely) recognized than targets were (correctly) recognized (*M* = .62) and both were more frequently recognized than new, unrelated items (*M* = .20). There was also an Item x Age interaction, *F*(6, 472) = 13.96, *p* < .001, (^{2} = .15, shown in Figure 5. What this figure shows, and what was confirmed in post-hoc tests, is that recognition rates for critical lures and targets both increased with age whereas they decreased for new, unrelated items. Finally, there was a Language shift x Item interaction, *F*(2, 472) = 24.76, *p* < .001, (^{2} = .10, shown in Figure 6. There were two key outcomes in this interaction. First, consistent with the associative-activation view, but not FTT, critical lures were more frequently (falsely) recognized than targets in the between-languages conditions, but there were no differences between these recognition rates in the within-language conditions. Second, targets were better recognized in the within-language than between-languages conditions, something that would be predicted given the additional repetition of cues associated with same-language recognition tests. However, the opposite effect was observed with critical lures. Here, and again consistent with the predictions of the associative-activation view but not FTT, (false) recognition of critical lures was more frequent in the between-languages than within-language conditions. There were no differences in (incorrect) recognition rates for new, unrelated items.

A' scores for targets and critical lures were also analyzed using a 2(Language shift: within-language versus between-languages) x 2(Item: target, critical lure) x 4(Age: 6, 8, 12, and 20 years old) ANOVA. The *A'* analyses revealed a main effect for item, *F*(1, 459) = 4.67, *p* < .05, (^{2} = .01, where, not surprisingly, targets (*A'* = .82) were more easily discriminated than critical lures (*A'* = .80). There was also a main effect for age, *F*(3, 459) = 22.84, *p* < .001, (^{2} = .13, where post-hoc analyses revealed the following pattern of reliable age differences: 6-year-olds (*A'* = .74) < 8-year-olds (*A'* = .81) < 12-year-olds (*A'* = .83) < 20-year-olds (*A'* = .86). Like the raw score analyses, developmental increases in recognition rates were similar for targets and critical lures. Unlike the raw score analysis, there was no Language shift x Item interaction.

Discussion and Conclusions

That both true and false recall increased with age and that these increases were greater for true than false memories, is consistent with previous developmental research (e.g., Brainerd & Reyna, 2005; Howe, 2005). What is novel about the current findings is that (a) these age-related increases in true and false recall were observed in within-language study-test conditions (in both French and English) as well as between-languages study-test conditions (English-French, French-English) (Figures 1 and 2) and (b) the developmental pattern for false recall in the between-languages conditions was consistent with the associative-activation view of false memory development, not FTT (Figure 2b). Although both FTT and the associative-activation accounts made similar predictions concerning true recall, FTT predicted that false recall rates should be higher in between-languages conditions than within-language conditions regardless of age due to the increased gist processing afforded by processing information across languages. However, the pattern of results was consistent with the associative-activation view that false memory rates should vary across age such that older (adult) participants with more exposure and proficiency (hence greater automaticity in activation of concepts and their interitem associations) should exhibit higher false recall rates in between-languages conditions than within-language conditions but the reverse should be true for younger participants. The data are clearly in line with this latter prediction and the cross-over interaction that was obtained suggests that the development of false memories is not simply a matter acquiring concepts and processing gist across the list. Indeed, the recognition findings clearly showed that the concepts studied here were available in memory regardless of the age of our participants or the language in which the concepts were presented. Thus, the key to false memory development is not simply the existence of concepts and their associations in children's and adults' knowledge base, but the automaticity with which these concepts and links are activated.

These results are consistent with the McElree et al. (2000) results for unbalanced bilinguals as the between-languages manipulation did not increase false memories in young children. This suggests that early learners may have been unable to automatically activate items and their associations in the second language. This is not due to the fact that the youngest children failed to access conceptual representations in L2 as indicated by their recognition performance and by the finding that false memories were obtained in both within-language conditions (in English and French). Rather, it is due to younger children's inability to access these concepts and their relations automatically.

In addition, although false recall was more frequent in the within-language than the between-languages conditions, this lower false memory rate for young children in the between-languages conditions may be related to item-distinctiveness, making it easier to reject items for which translation was required. One problem with this interpretation is that it may apply solely to recall outcomes. As already noted, when recognition measures were used instead, all participants, regardless of age, exhibited more false recognition in between-languages than within-language conditions.

Eight- and 12-year-old children's performance indicated that they were equally proficient in both languages. That is, like the other child participants in this study, true recall was better in within-language English than French conditions and in within-language than between-languages conditions. For false recall, although there were more items falsely recalled in English than French in within-language conditions (like all other participants in this study), there were no differences in rates of false recall between within-language and between-languages study-test conditions. Consistent with the associative-activation view but not FTT, it would seem that as exposure and proficiency in a second language increases, so too does automaticity of semantic processing (Frenck-Mestre & Prince, 1997), leading to similar levels of false recollection in both within-language and between-languages tasks. Because there were no differences in false recollection between within-language and between-languages conditions for these children it would seem that cross-language priming does not increase automatic processing beyond that of within-language priming. It could be that each language still possesses separate but partially overlapping conceptual stores at these ages, making cross-language priming less automatic and thus more difficult, but not impossible.

Adults' performance was consistent with increased automaticity of conceptual processing in both languages and the pattern of findings obtained here is the same as that found in other studies with bilingual adults (e.g., Cabeza & Lennartson, 2005; Sahlin et al., 2005). What these findings indicate is that as proficiency and exposure to a second language increases in bilinguals, conceptual stores for both languages overlap more and more, rendering cross-language conceptual processing more automatic and more likely, increasing the rate of false memories for the between-languages conditions beyond that for within-language conditions. However, even for adults, false memories were more prevalent in English than in French, at least in the within-language conditions. Thus, despite the increased likelihood of being balanced bilinguals, adults still experience more false recollection in English than French. Although the adults in this study had passed a French proficiency test, it is not known whether they were early or late L2 learners. Because late learners may exhibit less proficiency in L2 than early learners (e.g., Kotz & Elston-Guttler, 2004; Silverberg & Samuels, 2004), this might account for the finding that there were more false memories in English than in French even for adults. However, as noted earlier, others have found similar language asymmetries in both balanced and unbalanced bilinguals (Duyck & Brysbaert, 2004). An alternative and perhaps simpler explanation for these asymmetries might be that because the DRM lists used here were normed in English, not French, studies like this (as well as others that do not have separate norming statistics in the second language, e.g., Cabeza & Lennartson, 2005) may underestimate false memory production in the second language for bilingual adults. Nevertheless, that more false memories occurred in between-languages than within-language conditions for individuals likely to be balanced bilinguals shows that there is considerable semantic "cross-talk" that occurs relatively automatically among languages.

The recognition findings also favored the associative-activation view over FTT, although the pattern of recognition results depended on whether raw scores or corrected (*A'*) measures were analyzed. Relative to the findings for recall, developmental and language differences were somewhat attenuated (something that is not unexpected given the relative ease of recognition versus recall). Unlike recall, analysis of raw recognition scores showed that there were no differences in true recognition rates as a function of language in within-language conditions, but like recall, false recognition rates were higher in English than French. This finding was not statistically reliable when *A'* scores were analyzed. For within-language versus between-languages comparisons, targets were better recognized in within-language than between-languages conditions, but false recognition rates were higher in bilingual than unilingual conditions, regardless of age. Although this same effect did not achieve statistical significance in the *A'* analysis, it is important because it shows that even for unbalanced bilinguals (including the youngest of the children tested here), false recognition rates can be influenced by processing information both languages. Although recall measures failed to show overlapping conceptual representations across languages in young children, our raw-score recognition findings indicated that our early learners did exhibit higher false memory rates when processing involved both languages.

Overall, then, the findings from this study are more consistent with the associative-activation explanation of false memory development than with FTT. Specifically, these results support the idea that true and false recall and recognition, but especially false recollection in children (e.g., Howe, 2005, 2006, in press) and adults (e.g., Hutchison & Balota, 2005; Roediger, Watson, McDermott, & Gallo, 2001; Underwood, 1965), depends on the automaticity of associative activation. Indeed, consistent with the literature on the development of bilingual memory, the growth of two (or more) lexicons that point to similar (or the same) concepts begins first and foremost with associative connections. Activation in one language can and does prime lexical entries from the other language by pointing to (being associated with) the same concept. Although there may be variation between languages in their associative structures, there tends to be greater similarity than dissimilarity. As bilingual memory develops, lexical access from both languages to increasingly overlapping conceptual representations becomes more and more automatic (also see French & Jacquet, 2004). The lexicons' (associative) relation to conceptual representations depends on the proficiency in the second language (French & Jacquet, 2004). This associative model posits that when second language proficiency is low, concepts are initially accessed through the first language requiring translation before activating the conceptual-level representation. As development proceeds and lexical access to conceptual representations from the second language become more readily available, translation processes are no longer necessary and now both languages can access conceptual information more rapidly and automatically. In the final analysis, whether using one or more languages, it is the automaticity of associative activation that determines memory illusions in the DRM paradigm. Regardless of whether words are presented within or between languages, increases in children's ability to automatically access conceptual representations drive increases in false recollection with development.

This conclusion and the data provided here that supports it fits well with other recent findings concerning children's susceptibility to the DRM illusion (e.g., Howe, 2005, 2006, in press). Moreover, it is consistent with speculation on the mechanisms underlying adults' DRM illusions using one (Hutchison & Balota, 2005) or more (Cabeza & Lennartson, 2005; Sahlin et al., 2005) languages. There is no reason to suppose that the same processes should not apply to the development of the DRM illusion in bilingual children and the results of this study do not give us any reason to modify these claims.

References

Altarriba, J., & Mathis, K. M. (1997). Conceptual and lexical development in second language acquisition. *Journal of Memory and Language, 36,* 550-568.

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

Cabeza, R., & Lennartson, E. R. (2005). False memory across languages: Implicit associative response vs fuzzy trace views. *Memory, 13*, 1-5.

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

Dufour, R., & Kroll, J. F. (1995). Matching words to concepts in two languages: A test of the concept mediation model of bilingual representation. *Memory & Cognition,* 23, 166-180.

Duyck, W., & Brysbaert, M. (2004). Forward and backward number translation requires conceptual mediation in both balanced and unbalanced bilinguals. *Journal of Experimental Psychology: Human Perception and Performance, 30,* 889-906.

French, R. M. & Jacquet, M. (2004). Understanding bilingual memory: Models and data. *Trends in Cognitive Sciences*, *8,* 87-93.

Frenck-Mestre, C., & Prince, P. (1997). Second language autonomy. *Journal of Memory and Language, 37,* 481-501.

Gallo, D. A., & Seamon, J. G. (2004). Are nonconscious processes sufficient to produce false memories? *Consciousness and Cognition*, *13,* 158-168.

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

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. (in press). Visual distinctiveness and the development of children's false memories. *Child Development*.

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

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.

Kotz, S. A., & Elston-Guttler, K. (2004). The role of proficiency on processing categorical and associative information in the L2 as revealed by reaction time and event-related brain potentials. *Journal of Neurolinguistics*, *17,* 215-235.

Kroll, J. F. & Stewart, E. (1994). Category interference in translation and picture naming: Evidence for asymmetric connections between bilingual memory representations. *Journal of Memory and Language*, *33*, 149-174.

Marmolejo, G., Diliberto-Macaluso, K., & Altarriba, J. (2003, November). *Correct and false* *recall in Spanish-English bilinguals*. Poster presented at the annual meeting of the Psychonomic Society, Vancouver, Canada.

McElree, B., Jia, G., & Litvak, A. (2000). The time course of conceptual processing in three bilingual populations. *Journal of Memory and Language*, *42*, 229-254.

Pastore, R. E., Crawley, E. J., Berens, M. S., & Skelly, M. A. (2003). “Nonparametric” A’ and other modern misconceptions about signal detection theory. *Psychonomic Bulletin & Review*, *10*, 556-569.

Raaijmakers, J. G. W., & Zeelenberg, R. (2004). Evaluating the evidence for nonconscious processes in false memories. *Consciousness and Cognition*, *13*, 169-172.

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, Balota, D. A., & Watson, J. M. (2002). Spreading activation and arousal of false memories. In H. L. Roediger, III, J. S. Nairne, I. Neath, & A. M. Suprenant (Eds.), *The nature of remembering: Essays in honor of Robert G. Crowder* (pp. 95-115). Washington, DC: American Psychological Association.

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

Sahlin, B. H., Harding, M. G., & Seamon, J. G. (2005). When do false memories cross language boundaries in English-Spanish bilinguals? *Memory & Cognition*, *33*, 1414-1421.

Silverberg, S. & Samuel, A. G. (2004). The effect of age of second language acquisition on the representation and processing of second language words. *Journal of Memory and Language*, *51*, 381-398.

Snodgrass, J. G., & Corwin, J. (1988). Pragmatics of measuring recognition memory: Applications to dementia and amnesia. *Journal of Experimental Psychology: General*, *117*, 34-50.

Talamas, A., Kroll, J. F., & Dufour, R. (1999). From form to meaning: Stages in the acquisition of second language vocabulary. *Bilingualism: Language and Cognition*, *2*, 45-58.

Wakeford, Y., Carlin, M. T., & Toglia, M. P. (2005, November). *Effects of bilingual processing on veridical and false memory*. Poster presented at the annual meeting of the Psychonomic Society, Toronto, Canada.

Authors' Note

This research was supported by a grant from the Natural Sciences and Engineering Research Council of Canada to Mark L. Howe. The authors thank Chuck Brainerd, Judith Kroll, and an anonymous reviewer for their extremely constructive comments on a previous version of this article. Correspondence concerning this research should be sent to Professor Mark L. Howe, Department of Psychology, Lancaster University, Lancaster, LA1 4YF, United Kingdom, e-mail: mark.howe@lancaster.ac.uk.

Footnotes

[1]The calculation of A', the nonparametric counterpart to the more familiar signal detection statistic d' (Snodgrass & Corwin, 1988), is straightforward. For targets, $A' = .5 + [(H - FA_T)(1 + H - FA_T)] / [4H(1 - FA_T)]$, where H is the proportion of hits for targets from presented lists and FA_T is the proportion of false alarms from unpresented lists. This equation applies when H (FA_T. When $H < FA_T$, the relevant equation is, $A' = .5 - [(FA_T - H)(1 - H + FA_T)] / [4FA_T(1 - H)]$. For critical lures, $A' = .5 + [(FA_{CLP} - FA_{CLN})(1 + FA_{CLP} - FA_{CLN})] / [4FA_{CLP}(1 - FA_{CLN})]$, where FA_{CLP} is the proportion of false alarms for critical lures from presented lists and FA_{CLN} is the proportion of false alarms from unpresented lists. This equation applies when FA_{CLP} (FA_{CLN}. When $FA_{CLP} < FA_{CLN}$, the relevant equation is, $A' = .5 - [(FA_{CLN} - FA_{CLP})(1 - FA_{CLP} + FA_{CLN})] / [4FA_{CLN}(1 - FA_{CLP})]$. Once calculated, a value of .5 indicates an absence of true recognition or an absence of false recognition and a value of 1 indicates very high (perfect) levels of true or false recognition.

[2]These latter conditions were collapsed as there were no important differences among them.

Appendix

DRM lists used in this study: Critical target (in caps) with list items 1-14 (From Stadler, Roediger, & McDermott, 1999)

English Item	French equivalent	English Item	French equivalent
MAN	HOMME	COLD	FROID
woman	femme	hot	très chaud
husband	mari	snow	neige
uncle	oncle	warm	chaud
lady	dame	winter	hiver
mouse	souris	ice	glace

male	male	wet	mouillé
father	père	frigid	glacial
strong	fort	chilly	frilleux
friend	ami	heat	chaleur
beard	barbe	weather	temperature
person	personne	freeze	geler
handsome	beau	air	air
muscle	muscle	shiver	frissoner
suit	habit	arctic	arctique

English Item	*French equivalent*	*English Item*	*French equivalent*
DOCTOR	DOCTEUR	SLEEP	DORMIR
nurse	infirmière	bed	lit
sick	souffrant	rest	reposer
lawyer	avocat	awake	éveillé
medicine	médicament	tired	fatigué
health	santé	dream	rêve
hospital	hôpital	wake	réveille
dentist	dentiste	snooze	petit somme
physician	médecin	blanket	couverture
ill	malade	doze	sommeiller
patient	patient	slumber	sommeil
office	bureau	snore	ronfler
stethoscope	stéthoscope	nap	sieste
surgeon	chirurgien	peace	paix
clinic	clinique	yawn	bailler

English Item	*French equivalent*	*English Item*	*French equivalent*
FRUIT	FRUIT	LION	LION
apple	pomme	tiger	tigre
vegetable	légume	circus	cirque
orange	orange	jungle	jungle
kiwi	kiwi	tamer	dompteur
citrus	agrume	den	**tanière**
ripe	mûr	cub	petit
pear	poire	Africa	Afrique
banana	banana	mane	**crinière**
berry	baies	cage	cage
cherry	cerise	feline	félin
basket	panier	roar	grondement
juice	jus	fierce	féroce
salad	salade	bears	ours
bowl	bol	hunt	chasse

English Item	*French equivalent*	*English Item*	*French equivalent*
SWEET	**SUCRÉ**	SHIRT	CHEMISE
sour	aigre	blouse	chemisier
candy	bonbon	sleeves	manches
sugar	sucre	pants	pantalon

bitter	amer	tie	cravate
good	savoureux	button	bouton
taste	goût	shorts	short
tooth	dent	iron	fer **à** repasser
nice	bon	polo	polo
honey	miel	collar	collet
soda	soda	vest	gilet
chocolate	chocolat	pocket	poche
heart	coeur	jersey	chandail
cake	gâteau	belt	ceinture
tart	tartelette	linen	lin

Figure Captions

Figure 1. Within-language true (1a) and false (1b) recall trends as a function of age.

Figure 2. Within-language versus between-languages true (2a) and false (2b) recall trends by as a function of age.

Figure 3. Within-language recognition trends for targets, critical lures, and new (unrelated) items by age.

Figure 4. Within-language recognition trends for targets, critical lures, and new (unrelated) items by language.

Figure 5. Within-language and between-languages recognition trends for targets, critical lures, and new (unrelated) items by age.

Figure 6. Within-language and between-languages recognition trends for targets, critical lures, and new (unrelated) items by language.