Journal of Memory and Language 95 (2017) 146–158

Contents lists available at ScienceDirect

Journal of Memory and Language

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

The role of language proficiency in producing false memories

Jason Arndt [a,*], Maria Soledad Beato [b]

[a] Middlebury College, USA
[b] University of Salamanca, Spain

ARTICLE INFO

Article history:
Received 16 November 2016
revision received 9 February 2017

Keywords:
Language proficiency
False memory
DRM paradigm

ABSTRACT

We report three experiments examining the role that language proficiency plays in the production of false memory. We constructed Deese-Roediger-McDermott paradigm lists using both English and Spanish free association norms, which enabled us to control the associations between studied items and critical words. Experiment 1 showed that native English speakers who were learning Spanish produced more false memory when DRM critical words were studied and tested in English compared to Spanish. Experiment 2 showed that native Spanish speakers who were learning English produced more false memory when DRM critical words were studied and tested in Spanish compared to English. Experiment 3 showed that native Spanish speakers who were highly proficient in English produced more false memory for DRM critical words studied and tested in English compared to native Spanish speakers who were lower in English proficiency. Collectively, these results support the role that the automaticity of concept access plays in producing false memory.

Introduction

False memories have been extensively studied in monolingual speakers of many languages, including English (Deese, 1959; Roediger & McDermott, 1995), Spanish (Anastasi, Rhodes, Marquez, & Velino, 2005; Beato & Arndt, 2014; Beato & Díez, 2011), Portuguese (Carneiro & Fernandez, 2013; Carneiro et al., 2012), French (Corson & Verrier, 2007; Corson, Verrier, & Bucic, 2009; Dubuisson, Fiori, & Nicolas, 2012), Dutch (Van Damme & d'Ydewalle, 2009), Japanese (Kawasaki & Yama, 2006; Kawasaki-Miyaji, Inoue, & Yama, 2003), and Italian (Sergi, Senese, Pisani, & Nigro, 2014). The Deese–Roediger–McDermott (DRM) paradigm (Deese, 1959; Roediger & McDermott, 1995) had been the most commonly-used experimental paradigm to study false memories in the laboratory, and typically entails the presentation of a list of words to study (e.g., *thread*, *pin*, *eye*, *sewing*, *sharp*, *point*, *prick*, *thimble*, *haystack*, and *thorn*), all of them associated to a non-presented *critical word* (e.g., NEEDLE). On a subsequent memory test, participants often falsely recall or recognize critical words, even though the critical word was not studied.

One question that has been investigated in the literature is how language proficiency influences the incidence of false memories (for a review, see Graves & Altarriba, 2014). For example, studies have compared the rate of false memories produced when people study DRM lists in their native (and dominant) language and their second (non-dominant) language (Sahlin, Harding, & Seamon, 2005). In many of these studies, DRM lists were translated from the original English lists (Deese, 1959; Roediger & McDermott, 1995; Stadler, Roediger, & McDermott, 1999) to a second language, such as Spanish (Marmolejo, Diliberto-Macaluso, & Altarriba, 2009; Sahlin et al., 2005) or French (Cabeza & Lennartson, 2005; Howe, Gagnon, & Thouas, 2008). These studies have generally found that false memories are greater when DRM lists are studied and tested in participants' dominant language compared to their non-dominant language (Howe et al., 2008; Sahlin et al., 2005; but see Cabeza & Lennartson, 2005 for an exception). However, studying the rate of false memories across participants' dominant and non-dominant languages by directly translating DRM lists from English to another language may be problematic because the strength of the associations between study words and critical words can differ across languages (Marmolejo et al., 2009). Thus, studies that directly translate English DRM lists to other languages are comparing the level of false memory across languages by using lists that may have stronger associative relationships in the participants' dominant language (English) than in their non-dominant language (Spanish or French), which could account for the differences in false memory in bilinguals' dominant and non-dominant languages. As a result, it would be preferable to build, for example, Spanish lists from Spanish free-association norms in the same way

* Corresponding author at: Department of Psychology, Middlebury College, Middlebury, VT 05753, USA.
E-mail address: jarndt@middlebury.edu (J. Arndt).

http://dx.doi.org/10.1016/j.jml.2017.03.004
0749-596X/

that English lists are built from English free-association norms (Marmolejo et al., 2009).

There are only two studies that have investigated how false memory is affected by language proficiency and used DRM lists constructed from association norms in each language (Anastasi et al., 2005; Kawasaki-Miyaji et al., 2003). First, Anastasi et al. (2005, Exp. 2) examined false memory in participants' native language (Spanish) and their second language (English), and found that critical word false alarms were greater in the second language (English) than the first language (Spanish). However, these participants were highly fluent in English, and because they lived in the United States, used English regularly in their daily lives. Thus, it is possible that English was the participants' dominant language, which hinders the interpretation of this result as being due to the influence of language proficiency on false memory. In an effort to address this problem, Anastasi et al. (2005) tested monolingual Spanish (Exp. 3) and English (Exp. 4) speakers. In these last two experiments, the results favored the view that greater expertise in a language produced greater false memory. Specifically, participants falsely recognized more critical words in their native (and dominant) language (Spanish, Exp. 3, and English, Exp. 4) than in the language they did not speak. However, it is unclear if this result truly was due to greater proficiency with a language. In particular, studying monolingual participants may simply compare a condition where participants were able to access all of the concepts a word referred to (when they studied DRM lists in the language they spoke) and a condition where they were unable to access many or most of the concepts words referred to (when they studied DRM lists in the language they did not speak). Thus, while Experiments 3 and 4 of Anastasi et al. (2005) established that their DRM lists in both Spanish and English were able to evoke false memories when participants spoke those languages, those studies do not necessarily imply that greater language proficiency increases false memory.

Second, Kawasaki-Miyaji et al. (2003) studied Japanese-English bilinguals for whom Japanese was their native and dominant language, and constructed DRM lists using both Japanese (Miyaji & Yama, 2002) and English (Nelson, McEvoy, & Schreiber, 1998) free-association norms. Some DRM lists were tested in the same language as they were presented at study (i.e., Japanese-Japanese, English-English) while other DRM lists were tested in a different language than they were presented at study (i.e., Japanese-English, English-Japanese). The results generally showed that false recognition of critical words was greater when the test language was Japanese (participants' native and dominant language) than English (participants' second and non-dominant language). One complication with interpreting this result is that false alarm rates to unstudied, unrelated test items were not reported separately for items tested in Japanese and English. Thus, the finding that participants falsely recognized critical words more often when they were tested in Japanese is hard to interpret because it could simply reflect a bias to respond "studied" for items tested in Japanese. Further, when participants who had higher levels of English proficiency (balanced bilinguals) were compared with participants who had lower levels of English proficiency (unbalanced bilinguals), no differences in critical word false memory were observed. This result suggests that greater proficiency with a language does not produce greater levels of false recognition, in contrast to the conclusions of Anastasi et al. (2005).

In summary, studies of false memory in individuals with different language proficiency generally indicate that false recognition is lower in participants' non-dominant language compared to their dominant language. However, as noted above, there are two important limitations of existing studies. First, many studies have directly translated DRM lists from English to another language, instead of constructing DRM lists based upon association norms in each language (Howe et al., 2008; Marmolejo et al., 2009; Sahlin et al., 2005). This method of examining false memory in different languages leaves open the possibility that the DRM lists in participants' dominant language (English) were better able to elicit false memories than translated DRM lists in participants' non-dominant language. That is, it is possible, and perhaps likely, that translated lists had weaker overall associative relationships between studied items and critical words (Graves & Altarriba, 2014), which is known to produce lower levels of false memory for critical words (Roediger, Watson, McDermott, & Gallo, 2001). Thus, although the results of studies showing greater false memory in participants' dominant language compared to their non-dominant language can be interpreted to support the claim that greater language proficiency produces greater false memory, there are other, equally-plausible interpretations of this finding that do not claim that language proficiency differences affect false memory. Second, of the two studies that have constructed DRM lists using association norms specific to the two languages participants speak (Anastasi et al., 2005; Kawasaki-Miyaji et al., 2003), one study has favored the conclusion that greater language proficiency is associated with greater critical word false memory (Anastasi et al., 2005), while the other (Kawasaki-Miyaji et al., 2003) did not find differences in critical word false memory as a function of non-dominant language proficiency. Thus, there is uncertainty regarding whether variations in language proficiency are associated with differences in false memory.

An interesting parallel to studies examining the relationship between language proficiency in bilinguals and false memory is that there are also differences in false memory across development. Studies of false memory in children and adults using the DRM paradigm have consistently shown that false memory increased with age, such that children produced lower levels of false recognition than adults (Brainerd, Forrest, Karibian, & Reyna, 2006; Brainerd, Reyna, & Forrest, 2002; Carneiro, Albuquerque, Fernandez, & Esteves, 2007; Carneiro & Fernandez, 2010; Howe, 2006; Howe et al., 2008). This same developmental pattern has also been found when comparing younger and older children, such that older children produce more false memories that younger children (Carneiro & Fernandez, 2010).

According to Carneiro and Fernandez (2010), improvements in language proficiency across development explain the differential susceptibility to false memories that younger children have compared to older children and that older children have compared to adults. A second viewpoint that also explains developmental changes in false memory is offered by associative activation theory (Howe, Wimmer, Gagnon, & Plumpton, 2009). This theory claims that older children and adults are more susceptible to false memories than younger children because of increases in the strength and organization of associations in semantic memory, as well as the automaticity with which concepts are activated by words and activation spreads among associations. Thus, associative activation theory suggests that older children and adults are more likely to activate critical words' representations in semantic memory, resulting in higher levels of false memory.

Given the parallel between how false memory differs between individuals with different language proficiency and the developmental trajectory of false memory, it is possible to view the changes across cognitive development as comparable, at a theoretical level, to (1) the difference in dominant and non-dominant language proficiency in adults and (2) changes in non-dominant language proficiency in adults as second language learning progresses (Carneiro & Fernandez, 2010). This view would suggest that young children who study DRM lists in their native language and adults who study DRM lists in their non-dominant language will have lower levels of false memory than adults who study DRM lists in their dominant language because of differences in the strength of associations among concepts, as well as the automaticity of activating related concepts (e.g., critical words) in semantic memory (Howe et al., 2009). Further, this view would suggest that studying false recognition in adults with different levels of proficiency in their non-dominant language can provide a key test of theories of developmental changes in false memory. Specifically, it is difficult to interpret developmental changes in false memory as the product of changes in association strengths among concepts in semantic memory and/or improved automaticity of activating concepts related to study items because those changes co-occur with many other aspects of cognitive development. For example, source memory also improves across development (Cycowicz, Friedman, Snodgrass, & Duff, 2001), and erroneous beliefs that critical words were studied in a specific source is a key contributor to false memory (Arndt, 2012a). However, studying the effects of second language proficiency on false memory in adults is not subject to the same criticisms, since changes in second language proficiency can be examined within adulthood, when concomitant changes in cognitive and memory processes (e.g., source memory ability) are not occurring.

Although there are a number of models that seek to explain how second languages are learned and represented in the mind (e.g., Kroll & Stewart, 1994; Li, Farkas, & MacWhinney, 2004), they share the basic characteristic that second language lexical entries activate concepts less fully and quickly than first language lexical entries. To illustrate this idea, consider a prominent model of bilingualism, the Revised Hierarchical Model (RHM; Kroll & Stewart, 1994). The RHM argues that people develop and store lexical representations for their second (i.e., non-dominant) languages that are separate from the lexical representations of their first (i.e., dominant) language. Importantly, the RHM claims that non-dominant language lexical representations access the same conceptual representations in semantic memory as the lexical representations in a person's dominant language. Further, the RHM suggests that the lexical representations for a person's non-dominant language are less strongly associated with concepts compared to the lexical representations for a person's dominant language, which is how the model accounts for data suggesting that concepts in semantic memory are activated less strongly by a person's non-dominant language compared to their dominant language.[1]

If we adopt the representational viewpoint of the RHM, examining false memory across dominant and non-dominant languages, as well as across levels of non-dominant language proficiency, will allow us to examine how differences in automaticity of concept access and the resultant spreading activation from one concept to another contribute to false memory. Specifically, because participants should access concepts more automatically in their dominant language than their non-dominant language, we should find greater false memory when DRM lists are studied in their dominant language compared to their non-dominant language. Similarly, because participants with greater non-dominant language proficiency should access concepts more automatically in that language than participants with lower proficiency, we should find greater false memory when DRM lists are studied by participants with greater non-dominant language proficiency. Thus, by studying the effects of second language proficiency on false memory, we can not only understand false memory in participants' dominant and non-dominant languages, but we can also evaluate the role that automaticity of concept access, a key component of developmental theories of false memory, plays in producing false memory.

The present research seeks to address three primary limitations of prior research, while also systematically examining the influence of language proficiency on false memory. First, we developed DRM lists in both English and Spanish, and used free association norms specific to each language to do so (Fernández, Díez, & Alonso, 2009; Nelson et al., 1998). This enabled us to control the strength of the associative relationships between studied words and their associated critical words across the languages, alleviating concerns, cited above, that are raised by directly translating DRM lists from English to another language. Second, we examined false memory rates generated by these stimuli for native speakers of English who were learning Spanish (Experiment 1) and for native speakers of Spanish who were learning English (Experiment 2). Thus, we can ensure that differences in false memory across dominant and non-dominant languages were due to language proficiency and not stimulus differences across languages. Third, after establishing that our results from the first two experiments are due to studying DRM lists in participants' dominant vs. non-dominant language, we utilized the same stimuli to examine the influence of variations in second language proficiency on false memory. As noted above, the only two studies that have examined second language proficiency on false memory (Anastasi et al., 2005; Kawasaki-Miyaji et al., 2003) have reached different conclusions. Thus, the present work will contribute further data to understanding how second language proficiency influences false memory. In summary, this set of studies allowed us to test the influence of language proficiency on false memory both by comparing false memory in participants' dominant vs. non-dominant languages and by comparing false memory between participants with different levels of non-dominant language proficiency.

These three studies also allow us to evaluate the hypothesis, drawn from studies of the development of false memory, that language proficiency (Carneiro & Fernandez, 2010) and changes in the automaticity of access to concepts in semantic memory (Howe et al., 2009) underlie changes in false memory. Both of these hypotheses predict that (1) critical word false alarms should be greater in participants' dominant language than their non-dominant language and (2) as language proficiency in participants' non-dominant language increases, critical word false alarms should increase. Both of these predictions are based upon the idea that as participants' knowledge of a language increases (operationalized as a comparison between their dominant and non-dominant language in Experiments 1 and 2, and as a comparison between proficiency in their non-dominant language in Experiment 3), the automaticity with which concepts are accessed in semantic memory increases. The greater automaticity of concept access, in turn, makes it more likely that study words' conceptual representations will activate critical words' representations, producing increased false memory.

Experiment 1

In Experiment 1, we examined the influence of language dominance on false recognition by testing native English speakers who

[1] While we couch our explanation in terms of the RHM, it is important to note that this is for explanatory simplicity, because it offers the most intuitive explanation of concept access in bilinguals. The basic purpose of studying participants who are learning a second language, as well as the effects of second language proficiency, on false memory is the same even if one were to consider a different model of bilingual lexical/semantic representation. For example, although connectionist models of bilingual representation (Hernandez, Li, & MacWhinney, 2005) assume there is a single lexicon connected to semantics, the extent to which a given lexical entry activates concepts in semantic memory depends upon the person's experience with the lexical entity. Thus, when participants encounter a term in their second language, it will not activate semantic information as strongly or automatically as that term's equivalent in their first language. In this sense, the purpose of the present study, and the insights gained from studying false memory in participants' first and second language is based upon assumptions that models of bilingualism generally agree upon: how strongly concepts are activated by lexical entries in a person's first and second language.

were learning Spanish as a second language. We constructed DRM lists separately in English and Spanish using norms that were specific to each language. Further, we ensured that lists across languages were controlled for the associative strength between critical words and study items, as measured by both Forward Associative Strength (FAS) and Backward Associative Strength (BAS). As noted above, theories of how language proficiency (and cognitive development) impact false memory predict that participants will show more false memory in their dominant language (English) than in their non-dominant language (Spanish).

While it is not critical to our examination of language proficiency effects on false memory, it is important to note that we constructed DRM lists that differ in two ways from those used in most previous studies examining false recognition. First, the DRM lists we presented during encoding contained six associates per list, while many experiments using the DRM paradigm have employed lists that were eight to fifteen associates in length (e.g., Arndt, 2006; Roediger & McDermott, 1995). This is likely to make overall critical word error rates lower in the present studies than in studies using more associates for each DRM list. However, studying as few as two (Arndt, 2010) or four (Arndt & Gould, 2006) associates produces critical word false alarm rates that are well above false alarm rates to unstudied and unrelated test items. Thus, we expect our stimuli will produce critical word false alarm rates that are well above baseline false alarm rates, which is the key test of false memory induced by studying a series of associates. Second, each of the six associates in our DRM lists was simultaneously associated with three critical words, as indexed by both BAS and FAS, while most studies using the DRM paradigm employ a single critical word per list. This stimulus difference should not fundamentally change the nature of critical word false recognition, as our prior work has shown that such lists also show variation in false recognition as a function of both FAS (Beato & Arndt, 2014) and BAS (Beato, Arndt, & Arndt, in preparation), both of which are key drivers of false recognition in the DRM paradigm (Arndt, 2012b, 2015). Further, normative studies (Beato & Díez, 2011; Cadavid & Beato, 2017) document that lists constructed in this way produce robust false recognition. Thus, while our stimulus lists had slightly different features than most prior studies of false memory in the DRM paradigm, we expect they will nevertheless induce false recognition of critical words in the same way that lists have in previous work studying false memory with the DRM paradigm.

The primary reason we constructed stimuli in this way was in order to increase the number of observations for critical words while keeping the experiment to a reasonable duration. In particular, using typical DRM lists requires having participants study 10 or more associates for each critical word. Thus, in order to collect a large number of observations for critical words, it is typically necessary to have participants study a very large number of lists of 10 or more words, which in turn creates very lengthy experiments. In contrast, our current stimulus construction, where each list was associated with three critical words, enabled us to collect a large number of critical word observations while conducting a relatively short experiment.

Method

Participants

A total of 28 Middlebury College students, who were native English speakers and were in their third term of formally studying Spanish at Middlebury (46.4% women), participated as part of a course research appreciation requirement or in exchange for $10 payment. Participants' ages ranged from 18 to 22 years ($M = 19.75$, $SD = 1.38$). On average, participants rated their proficiency in Spanish as above average ($M = 6.32$, $SD = 2.50$) on a scale ranging from 1 (elementary knowledge) to 10 (native speaker proficiency). Students enrolled in third-term Spanish at Middlebury are in class five days a week for a total of six hours. However, outside of formal classroom instruction and formal practice related to their Spanish coursework, they will typically speak English. Our sample size for this experiment was determined by the number of participants we were able to recruit from third semester Spanish classes at Middlebury College in a single semester.

Materials

We constructed sixteen DRM lists (referred to as *themes* hereafter) composed of six words, each of which was related to three critical words. Half of the DRM lists were in English and half were in Spanish. The eight themes in English were constructed using the Nelson et al. (1998) free association norms, while the eight themes in Spanish were constructed using the Fernández et al. (2009) norms (see also, Beato & Arndt, 2014). In each theme, all critical words produced study items in free association (i.e., forward associative strength, or FAS). "*Critical word FAS*" was calculated as the sum the associative strengths between each critical word and its six associated words. "*FAS list strength*" was calculated as the sum of the FAS values for the three critical words (similar to Beato & Arndt, 2014; Beato & Díez, 2011). Critical word BAS values and BAS list strength values were similarly calculated. English themes' FAS list strength values ranged from 0.831 to 1.529 ($M = 1.244$, $SD = 0.25$), and Spanish themes' FAS list strength values ranged from 1.010 to 1.670 ($M = 1.314$, $SD = 0.23$). English themes' BAS list strength values ranged from 0.039 to 1.078 ($M = 0.483$, $SD = 0.34$), and Spanish themes' BAS list strength values ranged from 0.075 to 0.932 ($M = 0.482$, $SD = 0.29$). Neither FAS list strength nor BAS list strength differed between English and Spanish themes, $t(14) = -0.58$, $p = .570$, 95% CI $[-0.33, 0.19]$ for FAS; $t(14) = 0.01$, $p = .996$, 95% CI $[-0.34, 0.34]$ for BAS. A complete list of these stimuli are reported in the Appendix.

Procedure

Participants were tested individually, and provided informed consent prior to beginning the experimental session. An experimenter read study instructions aloud in English while participants followed along on a computer monitor. The instructions specified that participants would see words presented one at a time on a computer monitor, and that some of the words would be presented in English, while some of which would be presented in Spanish. They were instructed to do their best to remember the words in preparation for an unspecified future memory test. Participants first studied 12 themes, six in English and six in Spanish. Themes were presented visually on a computer monitor for 2000 ms, blocked by DRM list. The order of themes studied in English and Spanish were randomly intermixed. The words within each theme were always presented in decreasing order of FAS values. Following the presentation of the study list, participants were given old-new recognition memory instructions, which were read aloud by an experimenter while the participant followed along on a computer monitor. The recognition memory test consisted of 144 words, 72 in English and 72 in Spanish. The test list contained 72 studied words (half in English, half in Spanish), the 36 critical words related to studied themes (half in English, half in Spanish), 12 critical words related to lists that were not studied (*unrelated critical-distractors*; half in English, half in Spanish), and 24 words that were associates of unrelated critical-distractors (*unrelated-distractors*; half in English, half in Spanish). Thus, we included the same number of studied and non-studied words on the recognition memory test (72 studied and 72 non-studied words). The unrelated critical-distractors and unrelated-distractors both came from the two DRM themes in each language that were not presented during the study list. Studied words and their associated critical words were always presented in the same language that they were studied in. Themes were assigned to be studied and unstudied equally often, in order to control for item effects across themes' old-new test status.

Results and discussion

Mean percentages of true and false recognition for Experiment 1 are reported on the left-hand side of Table 1 as a function of Type of Word and Language.

False memory effect

A one-way repeated-measures analysis of variance (ANOVA) was conducted to test for the presence of false recognition by comparing the percentage of studied words, critical words, unrelated-distractors, and unrelated critical-distractors that were judged "old." This analysis revealed a significant difference among the types of words, $F(3,81) = 467.90$, $p < .001$, $\eta^2_p = .945$, and Bonferroni post hoc analyses showed that percentage of studied words correctly remembered (true recognition) (81.2%) was higher than false alarms to critical words (false recognition), false alarms to unrelated-distractors and false alarms to unrelated critical-distractors (22.1%, 7.4%, and 3.9%, respectively, all $ps < .001$). Further, false alarms to critical words were greater than false alarms to both unrelated-distractors and unrelated critical-distractors (both $ps < .001$), confirming that critical words produced the typical DRM false memory effect. There were no significant differences between false alarms to unrelated-distractors and unrelated critical-distractors ($p = .196$). Thus, as we anticipated, and as prior reports have established (Beato & Arndt, 2014; Beato & Díez, 2011; Cadavid & Beato, 2017), the stimulus lists we employed in this study (six associates related to each critical word; three critical words per list) produced robust levels of false recognition. We now turn to evaluate the impact that language proficiency had on false recognition.

Language dominance effects on true and false recognition

In order to determine the effect that language dominance had on true and false recognition, a 2 (Type of Word: Studied vs. Critical) × 2 (Language: English vs. Spanish) repeated-measures ANOVA was performed on the percentage of "old" responses given to each Type of Word.

This ANOVA yielded a significant main effect of Type of Word, $F(1,27) = 468.90$, $p < .001$, $\eta^2_p = .946$, and a significant Type of Word × Language interaction, $F(1,27) = 84.43$, $p < .001$, $\eta^2_p = .758$. The main effect of Type of Word indicates that true recognition ($M = 81.2\%$, $SD = 12.56$) was higher than false recognition ($M = 22.1\%$, $SD = 12.36$). Regarding the interaction between Type of Word and Language, Bonferroni post hoc analyses indicated that participants had higher true recognition in Spanish, their non-dominant language (87%), than in English, their dominant language (76%). In contrast, participants produced higher false recognition in their dominant language, English (29%) than in their non-dominant language, Spanish (15%). Regarding unrelated false alarms in English and Spanish, neither unrelated-distractors (7% vs. 8%) nor unrelated critical-distractors (2% vs. 5%) differed between English and Spanish, $t(27) = -0.23$, $p = .821$, 95% CI [−5.94, 4.75]; $t(27) = -1.54$, $p = .134$, 95% CI [−6.93, 0.98], respectively.

These results are consistent with prior reports showing that false recognition is greater in bilinguals' dominant language compared to their non-dominant language (e.g., Anastasi et al., 2005; Howe et al., 2009; Sahlin et al., 2005). Further, the lists that we employed in this experiment were constructed separately for each language based upon free association norms in English and Spanish in order to control the mean associative strengths (FAS and BAS) between study items and critical words. Thus, it seems unlikely that the present results can be explained by differences in the extent to which study items were related to critical words, which is a concern with many studies of false memory across bilinguals' dominant and non-dominant languages (Howe et al., 2009; Marmolejo et al., 2009; Sahlin et al., 2005). However, it is still possible to argue that controlling FAS and BAS across our English and Spanish lists did not truly equalize the extent to which they produce false memory. In the next experiment, we addressed this concern by replicating the effect of language dominance on false recognition by studying native Spanish speakers that were learning English as second language using exactly the same English and Spanish DRM lists used in Experiment 1. Thus, if the results of Experiment 1 were caused by our Spanish DRM lists eliciting lower false memory than our English DRM lists, we would replicate that finding when Spanish is participants' dominant language. On the other hand, if the results of Experiment 1 were due to differences in language dominance, we would find the opposite pattern of effects compared to Experiment 1 – greater false memory for DRM lists presented in Spanish than in English.

Experiment 2

Method

Participants

One hundred and fifty-six native Spanish speakers (82.7% women; M $age = 22.48$ years, $SD = 1.78$) who were undergraduates at the University of Salamanca received partial course credit for participating in this experiment. University students in Spain have typically studied English in primary and secondary school. However, it is typical that they would not speak English on a daily basis outside of formal instruction. Participants rated their proficiency with English as second language as moderate ($M = 5.25$, $SD = 1.63$) on a scale ranging from 1 (elementary knowledge) to 10 (native speaker proficiency). The sample size for this experiment was determined by the number of students we were able to recruit in a single semester.

Materials

In this experiment, we used the same materials employed in Experiment 1.

Procedure

The procedure was the same as Experiment 1, with the exception that participants were tested in groups of three to fifteen, and that experiment instructions were presented in Spanish. All participants completed the experiment individually on a computer, and were supervised by an experimenter for the duration of the experiment. Further, the instructions were read aloud by the experimenter while participants followed along on their individual monitors. All participants in a session began the study list at the same time (thus they ended the study list at the same time), and the experimenter then read the test instructions aloud to all participants while they followed along on their individual monitors. At the University of Salamanca, it is common practice to complete experiments in this way, such that participants are accustomed to completing individual tasks in a group setting, and they remain silent and focused on the task for the duration of the experiment.

Results and discussion

Mean percentages of true and false recognition for Experiment 2 are reported on the right-hand side of Table 1 as a function of Type of Word and Language.

Table 1
Mean percentage of true and false recognition in Experiment 1 and Experiment 2.

	Experiment 1		Experiment 2	
	English	Spanish	English	Spanish
True recognition	75.59 (15.31)	86.81 (11.20)	86.36 (10.75)	76.66 (13.95)
False recognition	29.17 (15.39)	15.08 (13.93)	13.75 (10.49)	30.20 (15.09)
Unrelated-distractors	7.14 (10.81)	7.74 (9.87)	7.96 (9.39)	5.82 (7.11)
Unrelated critical-distractors	2.38 (5.94)	5.36 (10.20)	4.81 (10.88)	4.38 (9.10)

Note: Standard deviations are reported in parenthesis.

False memory effect

A one-way repeated-measures ANOVA comparing the percentage of studied words, critical words, unrelated-distractors, and unrelated critical-distractors that were judged "old" on the recognition memory test revealed a significant difference among the types of words, $F(3,465) = 3053.32$, $p < .001$, $\eta^2_p = .952$. Bonferroni post hoc analyses showed that true recognition was higher than false alarms to critical words (false recognition), false alarms to unrelated-distractors and unrelated critical-distractors (22.0%, 6.9%, and 4.6%, respectively, all $ps < .001$). Further, false alarms to critical words were greater than false alarms to both unrelated-distractors and unrelated critical-distractors (both $ps < .001$), confirming the existence of false recognition. Finally, false alarms to unrelated-distractors were greater than false alarms to unrelated critical-distractors ($p < .05$).

Language dominance effects on true and false recognition

A 2 (Type of Word: Studied vs. Critical) × 2 (Language: Spanish vs. English) repeated-measures ANOVA was performed to examine the effects of language dominance on true and false recognition. The ANOVA revealed a significant main effect of Type of Word, $F(1,155) = 3403.63$, $p < .001$, $\eta^2_p = .956$, a significant main effect of Language, $F(1,155) = 14.78$, $p < .001$, $\eta^2_p = .087$, and a significant Type of Word × Language interaction, $F(1,155) = 402.28$, $p < .001$, $\eta^2_p = .722$. The main effect of Type of Word indicates that true recognition was higher than critical word false recognition (81.51% vs. 21.97%, respectively). The main effect of Language indicates that Spanish words were judged "old" more often than English words (53.43% vs. 50.05%, respectively). Finally, and critically, the interaction between Type of Word and Language indicated that true recognition was greater in English (86%) than in Spanish (77%) ($p < .001$), but false recognition was higher in Spanish (30%) than in English (14%) (both $ps < .001$ based on Bonferroni post-hoc analyses). Regarding unrelated false alarms in English and Spanish, unrelated critical-distractors did not differ between English and Spanish (5% vs. 4%), $t(155) = 0.40$, $p = .693$, 95% CI [−1.71, 2.56], while unrelated-distractor false alarms were higher in English (8%) than in Spanish (6%), $t(155) = 2.715$, $p < .01$, 95% CI [0.58, 3.69]. Note that this latter finding is in the opposite direction of the results for critical word false alarm rate differences across Language, and thus does not alter conclusions about the influence of language dominance on false memory.

One concern readers may have about this experiment is that we tested a considerably larger sample in Experiment 2 ($N = 156$) compared to Experiment 1 ($N = 28$). In order to evaluate the comparability of the results of this experiment to Experiment 1's, we statistically evaluated (1) the data from the first 28 participants in this experiment, and (2) we divided this experiment's sample into five groups based upon when they enrolled in the study. The first four groups were composed of 31 participants (the first, second, third, and fourth sets of 31 people who participated in the study), while the fifth group was composed of the final 32 people who participated in the study. To preview the outcome of these analyses, each of these sub-groups showed the same interaction between Item Type and Language that was found in analyses of the full sample of 156 participants. Further, the interaction took the same form, such that true recognition was greater in English than Spanish, but false recognition was greater in Spanish than English. Thus, the results of Experiment 2 were robust, even when sub-sets of the sample that were comparable in size to Experiment 1's were analyzed.

Recall that the analyses of the entire sample from this experiment showed a main effect of Type of Word, a main effect of Language, and an interaction between Type of Word and Language. The analysis of the first 28 participants showed similar results to the overall analyses, producing a main effect of Type of Word, $F(1,27) = 694.03$, $p < .001$, $\eta^2_p = .963$, and an interaction between Type of Word and Language, $F(1,27) = 43.61$, $p < .001$, $\eta^2_p = .618$. Bonferroni-adjusted comparisons demonstrated that the nature of the interaction was identical to the overall analyses. True recognition was higher in participants' non-dominant language (English; $M = 86.21\%$) than in their dominant language (Spanish; $M = 78.08\%$, $p < .001$), while false recognition was higher in participants' dominant language (Spanish; $M = 29.17\%$) than in their non-dominant language (English; $M = 16.47\%$, $p < .001$). Similarly, analyses of subsets of the sample throughout data collection (the first four sets of 31 participants and the final 32 participants) all showed main effects of Type of Word (smallest $F(1,31) = 571.51$, $p < .001$, $\eta^2_p = .949$, for the fifth set of participants, which contained 32 participants), documenting that true recognition was greater than false recognition. Further, all subsets showed an interaction between Type of Word and Language (smallest $F(1,30) = 44.57$, $p < .001$, $\eta^2_p = .598$). Bonferroni-adjusted post-hoc tests revealed that, in all subsets, true recognition was higher in participants' non-dominant language (English) than in their dominant language (Spanish; smallest mean difference = 7.08%; largest $p = .012$), while false recognition was higher in participants dominant language (Spanish) than in their non-dominant language (English; smallest mean difference = 12.19%, all $ps < .001$).[2] Thus, both the full sample from this experiment, as well as every smaller sub-set we analyzed (first 28 participants, sub-sets of 31 or 32 participants, collected throughout the course of running the study), showed the critical interaction between Type of Word and Language. Further, the interaction was due to the same pattern of results in each sub-sample – greater true recognition in participants' non-dominant language (English) and greater false recognition in participants' dominant language (Spanish).

The results of this experiment are precisely the opposite of those from Experiment 1 if viewed as a function of the language in which themes were studied and tested (English vs. Spanish). Specifically, participants who spoke Spanish as their dominant language and English as their non-dominant language had greater

[2] The main effect of Language, which was significant in the analysis of the full sample, only reached the significance criterion for the second set of 31 participants and the final set of 32 participants.

recognition of studied items and lower false recognition of critical words when themes were studied and tested in English compared to when themes were studied and tested in Spanish (Experiment 2). In contrast, participants who spoke English as their dominant language and Spanish as their non-dominant language had greater recognition of studied items and lower false recognition of critical words when themes were studied and tested in Spanish compared to when themes were studied and tested in English (Experiment 1). However, viewing these two experiments in terms of participants' dominant language reveals that the results of Experiments 1 and 2 lead to the same conceptual conclusion: Participants recognized more studied items and fewer critical words in their non-dominant language compared to their dominant language.

Comparison of false recognition in Experiment 1 and Experiment 2

In order to confirm that English word lists and Spanish word lists had similar associative power and produced similar levels of false recognition, we compared the levels of false recognition produced by native English speakers when they studied English DRM lists to the levels of false recognition produced by native Spanish speakers when they studied Spanish DRM lists, using data from Experiment 1 and Experiment 2. An independent-samples *t*-test confirmed that there were no differences in false recognition between English lists ($M = 29.17\%$, $SD = 15.39$) and Spanish lists ($M = 30.20\%$, $SD = 15.09$) in native English and Spanish speakers (i.e., in participants' dominant language), respectively, $t(182) = -0.33$, $p = .740$, 95% CI [−7.16, 5.19]. Similarly, there were no differences in false recognition between English lists ($M = 13.74\%$, $SD = 10.49$) and Spanish lists ($M = 15.08\%$, $SD = 13.93$) when they were studied by native Spanish and English speakers (i.e., participants' non-dominant language), respectively, $t(182) = 0.59$, $p = .558$, 95% CI [−3.15, 5.82]. These analyses document that our English and Spanish lists did not differ in their propensity to produce false recognition, as would be expected because they were constructed to have similar associative relationships across languages (see Experiment 1 Method).

The results of the first two experiments show that language proficiency, operationalized as a comparison between participants' dominant and non-dominant languages, produced differences in false recognition. Further, because the two experiments used the same stimuli, but tested participants with different dominant languages (English in Experiment 1, Spanish in Experiment 2), these results were not a consequence of stimulus differences across the DRM lists we constructed in English and Spanish, and instead are the result of language proficiency. To further test the role that language proficiency plays in producing false memory, we examined false memory differences in participants who are native Spanish speakers with high and low English language proficiency in Experiment 3. In both groups, English was their second, and non-dominant, language.

Experiment 3

Method

Participants

Fifty-two students at the University of Salamanca volunteered to participate in this study. They were native Spanish speakers and were learning English as a second language in an academic context in the University of Salamanca (University Language Centre; *M age* = 25.69, $SD = 7.63$; 58% women). Half of the participants were studying English at an elementary level (low English proficiency group; *M age* = 23.15, $SD = 3.02$), and the other half were studying English at advanced level (high English proficiency group; *M age* = 28.23, $SD = 9.81$), according to their scores on an English proficiency test.[3] In addition, participants indicated their proficiency in English as second language on a scale from 1 (elementary knowledge) to 10 (native speaker proficiency). There was a significant difference in self-assessed English proficiency between low ($M = 4.31$, $SD = 1.19$) and high ($M = 7.00$, $SD = 0.98$) English proficiency groups, $t(50) = 8.90$, $p < .001$, 95% CI [2.08, 3.30]. The sample size in this experiment was determined by recruiting as many high-proficiency English speakers as we could within a single semester (26) and matching that sample size with an equal number of low-proficiency English speakers.

Materials & procedure

The stimulus materials and procedure were the same as those used for Experiments 1 and 2. Similar to Experiment 2, participants were tested in groups ranging from 5 to 10 at a time, and completed the experiment individually on a computer. The same control procedures employed in Experiment 2 for group sessions were employed in Experiment 3.

Results and discussion

False memory effect

A one-way repeated-measures ANOVA comparing the percentage of studied words, critical words, unrelated-distractors, and unrelated critical-distractors that were judged "old" on the recognition memory test revealed a significant difference among the types of words, $F(3, 153) = 676.50$, $p < .001$, $\eta^2_p = .930$. Bonferroni post-hoc analyses showed a higher rate of true recognition (76.74%) than false alarms to critical words (26.28%), false alarms to unrelated-distractors (4.73%) and false alarms to unrelated critical-distractors (5.29%, $p < .001$ for all comparisons). This analysis also confirmed that false alarms to critical words were higher than false alarms to both unrelated-distractors and unrelated critical-distractors (both *ps* < .001). There was not a significant differences between false alarms to unrelated-distractors and unrelated critical-distractors ($p > .05$).

Language dominance effects on true and false recognition

A 2 (Type of Word: Studied vs. Critical) × 2 (Language: English vs. Spanish) repeated-measures ANOVA was performed to determine if the results obtained in Experiment 1 and Experiment 2 replicated in this study.

This ANOVA yielded a significant main effect of Type of Word, $F(1, 51) = 487.76$, $p < .001$, $\eta^2_p = .905$, a significant main effect of Language, $F(1, 51) = 13.97$, $p < .001$, $\eta^2_p = .215$, and a significant Type of Word × Language interaction, $F(1, 51) = 101.62$, $p < .001$, $\eta^2_p = .666$. The main effect of Type of Word indicates that true recognition was higher than false recognition (76.74% vs. 26.28%, respectively). The main effect of Language indicates that participants judged words "old" more often when they were studied and tested in Spanish than English (54.27% vs. 48.75%, respectively). Finally, and critically, the interaction indicated that there was greater true recognition in participants' non-dominant language (English, $M = 79.65\%$, $SD = 12.26$) than in their dominant language (Spanish, $M = 73.83\%$, $SD = 13.63$), while false recognition was greater in participants' dominant language (Spanish, $M = 34.72\%$, $SD = 16.72$) than in their non-dominant language (English, $M = 17.84\%$, $SD = 15.06$). Thus, as in Experiments 1 and 2, greater language proficiency produced higher false recognition. Regarding unrelated

[3] The English proficiency test was an online test with 130 multiple-choice questions, with four answer options for each question. The test assessed the students' knowledge of grammar and vocabulary, and tested a range of levels from A1.1 (low level) to C1.3 (high level). Furthermore, students who were determined to have high English proficiency had to pass a face-to-face test which consisted in a brief interview and role-play.

false alarms across English and Spanish lists, neither unrelated-distractors (5.77% vs. 3.69%) nor unrelated critical-distractors (6.09% vs. 4.49%) differed, $t(51) = 1.79$, $p = .079$, 95% CI [−0.25,4.42]; $t(51) = 0.962$, $p = .341$, 95% CI [−1.74,4.95], respectively.

False recognition and English language proficiency

In order to determine if participants' level of language proficiency in their non-dominant (English) language impacted false recognition rates, a 2 (Language: English vs. Spanish) × 2 (English Proficiency: High vs. Low) ANOVA was performed on critical word false alarms. Results, depicted in Fig. 1, indicated a main effect of Language, $F(1,50) = 66.02$, $p < .001$, $\eta^2_p = .569$, with higher false recognition in Spanish (34.72%) than English (17.84%), and a significant Language × English Proficiency interaction, $F(1,50) = 6.09$, $p < .05$, $\eta^2_p = .109$. Bonferroni-adjusted post-hoc analyses showed that there were differences in false recognition between the high and low English Proficiency groups only when themes were studied and tested in English (22.44% vs. 13.25%, $p < .05$, for the high and low English proficiency groups, respectively). Thus, as is evident in Fig. 1, English Proficiency was not related to false recognition when themes were studied and tested in Spanish, the participants' native language (34.19% vs. 35.26% for the high and low English proficiency groups, $p > .05$).

Finally, false alarms to unrelated-distractors did not differ between English and Spanish DRM lists in either the low English proficiency group (7.69% vs. 4.17%), $t(25) = 1.79$, $p = .086$, 95% CI [−0.53,7.58], or the high English proficiency group (3.85% vs. 3.20%), $t(25) = 0.53$, $p = .602$, 95% CI [−1.86,3.14]. Similarly, false alarms to unrelated critical-distractors did not differ between English and Spanish DRM lists in either the low English proficiency group (6.41% vs. 3.21%), $t(25) = 1.73$, $p = .096$, 95% CI [−0.61,7.02], or the high English proficiency group (5.769% vs. 5.770%), $t(25) = -0.0001$, $p = .999$, 95% CI [−5.71,5.71].

A second way to evaluate the influence of English Proficiency on false memory is to compare false recognition in English lists in native English speakers (29%, Experiment 1) with participants from the present experiment who had high and low English proficiency. These data are depicted in Fig. 2. This analysis found that there was a significant difference in false recognition between native English speakers (Experiment 1) and low proficiency English speakers (Experiment 3), $t(52) = 4.14$, $p < .001$, 95% CI [8.21,23.63], but that there was not a statistically-reliable difference between native English speakers and high proficiency English speakers, $t(52) = 1.57$, $p = .123$, 95% CI [−1.88,15.34]. Although this latter difference did not reach statistical significance, the ordinal pattern of the means is in the direction one would expect if language proficiency is systematically related to false recognition.

The results of this experiment built upon those from Experiments 1 and 2 by showing that language proficiency was related to false memory both across languages (i.e., comparing participants' dominant and non-dominant languages) as well as within a language (i.e., participants' non-dominant language). Further, cross-experiment comparisons between native English speakers in Experiment 1 and non-native English speakers in the present experiment documented that false memory for critical words whose associates were studied and tested in English was highest when English was a person's dominant language, lowest when a person had low proficiency in English, and in between when a person had moderate to high proficiency in English. Thus, the results of this experiment further underscore the relationship between language proficiency and false memory (Anastasi et al., 2005; Carneiro & Fernandez, 2010), and further support theories that suggest that one of the reasons that false memory increases is the automaticity with which lexical entities activate conceptual representations in a person's knowledge base (Howe et al., 2009).

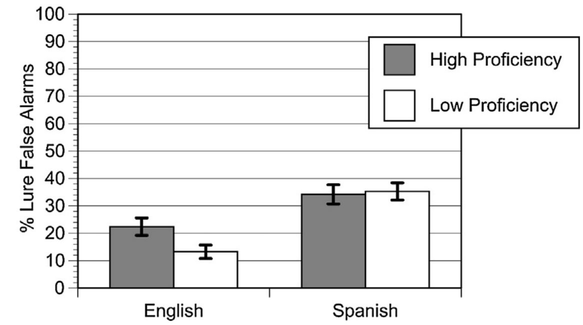

Fig. 1. Percentage of false recognition in English and Spanish as function of participants' English language proficiency in Experiment 3. Error bars represent the Standard Error of the Mean.

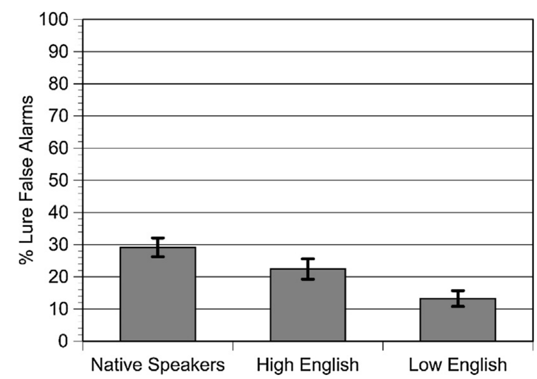

Fig. 2. Percentage of false recognition in English lists in Native speakers (Experiment 1) compared to Native Spanish speakers with High or Low English Proficiency in Experiment 3. Error bars represent the Standard Error of the Mean.

General discussion

The experiments reported in this paper produced two key results. First, false recognition was greater when participants studied and were tested in their dominant and native language compared to their non-dominant language (Experiment 1, 2, & 3). This finding occurred regardless of whether participants' dominant language was English (Experiment 1) or Spanish (Experiments 2 & 3). Second, participants who were more proficient in their second language produced higher rates of false recognition than participants who were less proficient in their second language (Experiment 3). Taken together, these results document that language proficiency plays a key role in producing false memory in the DRM paradigm.

These results replicate and extend prior research that has shown that false memory increases with language proficiency (Anastasi et al., 2005; Howe et al., 2008; Sahlin et al., 2005). Importantly, the studies reported in this paper were conducted using critical controls that strengthen the conclusion that prior findings were due to language proficiency differences and not confounding stimulus-based factors. Specifically, we rigorously controlled the associative strength between studied items and critical words across our English and Spanish lists, which has not been done in previous studies. Further, we verified that our results could not be explained by unknown stimulus-based factors by testing both participants for whom English was their dominant language and Spanish was their non-dominant language (Experiment 1), as well as participants for whom Spanish was their dominant language and English was their non-dominant language (Experiment 2). The results of these two experiments showed that language proficiency was the factor that was associated with higher rates of false memory, and not stimulus-based differences across languages. Finally, we tested participants who had varying levels of non-dominant language proficiency in English (Experiment 3), and found that the extent to which they falsely recognized critical words in English varied with their English proficiency. Thus, our results not only suggest that language proficiency increases false memory based upon comparisons of false memory in participants' dominant language relative to their non-dominant language, but also based upon comparisons within a single (non-dominant) language where participants varied in their proficiency with that language.

At a theoretical level, these results are consistent with views that suggest increased language proficiency enhances the automaticity with which lexical representations activate conceptual representations in semantic memory (Carneiro & Fernandez, 2010; Howe et al., 2009). Specifically, theories of developmental changes in false memory (Howe et al., 2009) argue there are three basic factors that change with development (and by association, with increased language proficiency). First, the strength of associations between concepts tends to increase as people gain more experience with those concepts. Second, the density of the interconnections among concepts in semantic memory tends to increase as people have greater experience with concepts and their inter-relationships. Third, as development progresses, the automaticity with which lexical representations activate conceptual representations increases, which in turn enables activation to spread to related concepts. As noted in the introduction, it is likely that the automaticity with which lexical representations activate conceptual representations plays a key role in producing the relationship between language proficiency and false memory in adult bilinguals. Thus, our finding that false memory was greater in participants' dominant language than their non-dominant language is likely a reflection of differences in the extent to which the lexical representations for each language activated their network of associations among concepts in semantic memory. Similarly, our finding that language proficiency differences in participants' non-dominant language was associated with false memory also likely reflects the extent to which their lexical representations of English words activated concepts in semantic memory. Thus, the present results support the specific proposition of developmental theories of false memory that experience with written language increases the extent to which conceptual representations are automatically activated, and in turn activate associated concepts in semantic memory.

One aspect of our results that merits further consideration is that true recognition of studied items was greater in participants' non-dominant language in all three of our experiments. There are two aspects of this finding that are important to consider. First, does it replicate prior findings in the literature? Second, this result can be seen as a potential challenge to activation-based explanations of why language proficiency impacts false recognition. Specifically, if false recognition is greater in participants' dominant language because they automatically activate a words' referent in semantic memory to a greater extent than they do in their non-dominant language, why does that same activation of semantic memory not also benefit true recognition? We address each of these points in turn.

Regarding whether this finding is anomalous relative to prior results, two studies examining how true and false memory vary across languages have reported similar results. First, while they did not directly test the comparison statistically, examination of the means reported in Fig. 1 of Kawasaki-Miyaji et al. (2003) shows that studying and testing words in participants' non-dominant language (English) produced higher true recognition, but lower false recognition, compared to studying and testing words in participants' dominant language (Japanese). Importantly, the means reported in Table 2 of Kawasaki-Miyaji et al. (2003) show that the true recognition result was driven by a sub-group of "unbalanced" bilinguals in their sample – participants who had lower English proficiency than Japanese proficiency (English-English TR = .79; Japanese-Japanese TR = .69). Similarly, Sahlin et al. (2005) studied English-dominant participants who were also highly proficient in Spanish. Focusing on the conditions that are most similar to the present study (the first study-test trial, English study & test vs. Spanish study & test conditions), the means reported in their Table 1 suggest that true recognition was greater when items were studied and tested in Spanish ($M = .69$) compared to when they were studied and tested in English ($M = .62$). In contrast, false recognition was greater when items were studied and tested in English ($M = .75$) compared to when items were studied and tested in Spanish ($M = .62$). Thus, both the results for unbalanced bilinguals studied by Kawasaki-Miyaji et al. (2003) and the first study-test trial reported in Sahlin et al. (2005) show similar patterns of true and false recognition to the present experiments.

Other findings in the literature, however, were not replicated by the present results. For example, Experiment 2 of Anastasi et al. (2005), Marmolejo et al. (2009), and Howe et al. (2008) all found slightly higher rates of true recognition for items studied and tested in participants' dominant language compared to items studied and tested in their non-dominant language.[4] It is best to view these outcomes as qualitative trends, because the mean differences tended to be small (4% or less), the true recognition difference was not statistically reliable in Anastasi et al. (2005) or Howe et al. (2008), and it was not tested in Marmolejo et al. (2009). Importantly, all of these studies examined participants who were highly proficient in their non-dominant languages, and who generally used both their dominant and non-dominant languages on a daily basis. Specifically, the participants in Experiment 2 of Anastasi et al. (2005) were native speakers of Spanish who lived in the United States, and the majority reported using both Spanish and English on a regular basis. Similarly, the participants in Marmolejo et al. (2009) all lived in South Florida, where both English and Spanish are used on a daily basis, and many were born in Spanish-speaking countries. Further, the participants in Howe et al. (2008) were either children in a French immersion program or were adults who were fluent in French and used French almost exclusively at work. Finally, and similar to the results of these three studies, Kawasaki-Miyaji et al.'s (2003) "balanced" bilingual group (reported in their Table 2) showed similar rates of true recognition across participants' dominant (Japanese) and non-dominant (English) languages (both conditions' $M = .79$). In contrast to the similar levels of true recognition across dominant and non-dominant languages, all of these studies reported higher rates of false recognition in participants' dominant language compared to their non-dominant language, with most mean differences being >10%.[5] This difference in false recognition was statistically significant in Anastasi et al. (2005) and Howe et al. (2008), while it was not directly tested in Marmolejo et al. (2009) or Kawasaki-Miyaji et al. (2003).

Taken together, the picture that emerges from the literature is that when participants are highly proficient in their non-dominant language and use it on a daily basis, true recognition

[4] We focused on Experiment 2 of Anastasi et al. (2005) because it was the study they reported that was most comparable to the present studies (bilingual participants, studying DRM lists in either English or Spanish). Thus, Anastasi et al's Experiments 3 & 4, which tested monolingual Spanish and English speakers, do not provide a test of how language proficiency per se impacts false recognition. Similarly, we focused on the recognition data reported by Marmolejo et al. (2009) because that is the task most comparable to the memory test we used in the present research.

[5] The single exception is Marmolejo et al. (2009), who found a mean difference of 7% when DRM lists were studied and tested in English vs. when they were studied and tested in Spanish for participants who were English dominant, but highly-proficient Spanish speakers. This difference was not statistically tested.

tends to be similar for words studied in their dominant and non-dominant languages, while false recognition is higher in their dominant language than in their non-dominant language. In contrast, when participants are less proficient in their non-dominant language (e.g., unbalanced bilinguals in Kawasaki-Miyaji et al., 2003; the present studies) and/or are less likely to use their non-dominant language on a daily basis (e.g., speaking Spanish while attending a University in Connecticut; Sahlin et al., 2005), true recognition is higher in their non-dominant language, while false recognition is higher in their dominant language. These tendencies may offer clues about the basis for the true recognition results in the present experiments as well as the bases of false recognition in general.

Regarding theoretical implications, the finding that studied items were recognized better in participants' non-dominant language can be seen as a challenge to activation-based interpretations of the false memory results in these studies. In particular, if it is the case that lexical representations in participants' dominant language activate conceptual representations more automatically and completely than lexical representations in their non-dominant language, it seems logical that true recognition should also be higher in participants' dominant language. Here, we outline two possible explanations for this result, while also evaluating the extent to which these explanations are consistent with the theory that differences in the automaticity of concept activation across dominant and non-dominant languages underlie our false memory findings.

One possible explanation these true recognition results can be derived from a proposal in the Revised Hierarchical Model (RHM; Kroll & Stewart, 1994) that second language learners sometimes translate terms from their non-dominant language directly into their dominant language. As a result, they do so without activating a non-dominant language term's meaning in semantic memory. The RHM proposes that this type of translation is more likely when a person is less proficient in their non-dominant language, and that its likelihood decreases with second language proficiency. Thus, one way that improved memory for items studied in participants' non-dominant language can be explained is that participants directly translated items from their non-dominant language to their dominant language, leading them to store a representation of those items in both languages. In turn, having multiple representations for words that were studied in their non-dominant language would tend to produce better memory on a later recognition memory test compared to words studied in their dominant language, which would only have produced storage of a single representation.

A second possible explanation is that reading and understanding words in participants' non-dominant language requires more processing effort than reading and understanding words in participants' dominant language. This explanation suggests the memory advantage for items studied in participants' non-dominant language stems from the greater difficulty they have in processing and understanding those studied items relative to processing and understanding words in their dominant language, which in turn benefits memory (i.e., it is a "desirable difficulty"; Bjork, 1994). For example, perhaps identifying and processing the orthography of words in a person's non-dominant language (i.e., accessing its lexical representation) requires more processing effort because that lexical representation has not been accessed as often as lexical representations in their dominant language. Similarly, perhaps accessing semantic representations using the non-dominant language's lexical representation is also more effortful (i.e., executed less automatically) than accessing that same semantic representation from the dominant language's lexical representation. Thus, this explanation flows naturally from bilingual representation models' proposal that lexical representations in peoples' non-dominant language are less strongly associated with their corresponding semantic memory representations than lexical entries in their dominant language (Hernandez et al., 2005; Kroll & Stewart, 1994).

Both of these possibilities can explain why true memory is greater when people study words in their non-dominant language, particularly when there is reason to believe they have somewhat lower proficiency in their non-dominant language. Specifically, lower levels of proficiency in peoples' non-dominant language should make it more likely that lexical entries in their non-dominant language are translated directly to their dominant language, producing multiple representations, and thus enhancing true memory for words studied in their non-dominant language. Similarly, people who have lower proficiency in their non-dominant language will activate lexical and/or semantic representations less automatically and strongly when they study a word in their non-dominant language, necessitating greater processing effort during encoding in order to comprehend words in their non-dominant language. Importantly, these two explanations can also account for why some studies have found similar levels of true recognition for items presented in participants' dominant and non-dominant languages. As highlighted above, there is reason to believe the studies that have observed this pattern have tested bilinguals who have very similar levels of proficiency in their dominant and non-dominant languages and use both languages daily (Anastasi et al., 2005, Exp. 2; Howe et al., 2008; Marmolejo et al., 2009), resulting in them being "balanced" bilinguals (Kawasaki-Miyaji et al., 2003). Thus, they would be less likely to have a tendency to translate words directly from their non-dominant language to their dominant language (Kroll & Stewart, 1994), and would be less likely to need to expend substantial extra processing effort to comprehend words in their non-dominant language, resulting in similar levels of true recognition across their dominant and non-dominant languages. As a result, each of these views seem capable of explaining both of the patterns of true recognition found in the extant literature – cases where people remembered studied items in their non-dominant language better than in their dominant language, as well as cases where people remembered studied items similarly in their dominant and non-dominant languages.

The final question that is important for evaluating the plausibility of these two explanations of our true recognition findings is if their propositions change the predictions made by an activation-based explanation of our false recognition findings, as well as the typical finding that false recognition is greater in participants' dominant language. Regarding the first explanation, that people sometimes translate terms directly from their non-dominant language to their dominant language, the primary concern one could raise is that if participants translate non-dominant language terms into their dominant language as part of the encoding process, it could lead them to process the word as if it was presented in their dominant language. This would have the effect of making activation spread to semantic memory from lexical representations in both their dominant and non-dominant languages, producing more, rather than fewer, critical word errors when items were studied in their non-dominant language. Thus, this view would require additional assumptions to specify how strongly lexical and semantic representations were activated in peoples' dominant language when they directly translate from their non-dominant language. Specifically, it would be necessary for the spreading activation inspired by translating study items from one's non-dominant to one's dominant language to be less than that produced by directly experiencing a study item in one's dominant language. For example, perhaps the processing resources consumed by the translation of a word from the non-dominant to the dominant language would reduce how fully the lexical representation is processed in the dominant language, thereby limiting the extent to which it produces automatic activation of the word's conceptual representation. Such an outcome would limit the amount of spreading activation translating a word into one's dominant language causes in semantic memory, and ultimately, the amount of activation that spreads to a critical word's representation.

The second proposal, that participants expend more processing resources in order to comprehend terms in their non-dominant language, can be seen as arguing that people expend more resources to activate both a word's lexical representation and the word's corresponding meaning in semantic memory when it is presented in their non-dominant language. This, in turn, could lead to the expectation that false recognition would be greater when items are studied and tested in peoples' non-dominant language, a pattern that has not been observed in the literature. However, as long as this viewpoint maintains the assumption, embedded in bilingual models of language representation, that lexical terms from one's non-dominant language activate semantic memory less well than dominant language lexical terms, this viewpoint is capable of explaining both the true and false recognition findings in the present experiments. Thus, as long as one assumes that any additional processing that is required to comprehend words in one's non-dominant language does not result in greater overall activation of a word's conceptual representation, it stands to reason that less overall activation will flow to critical words' representations in semantic memory when lists are studied in their non-dominant language. For example, it may be the case that extra processing resources are required to activate a word's lexical representation in one's non-dominant language, because those terms have not been experienced as often as terms in their dominant language. Alternatively, it is possible that semantic memory representations are activated differentially by words in a person's non-dominant and dominant languages, even though words in both languages activate semantic memory representations to a degree that is sufficient for conceptual comprehension. Indeed, such a claim is at the heart of models of bilingual language representation, as reviewed in the introduction, and those models were designed to explain data that imply such a differential in concept activation across dominant and non-dominant languages in bilinguals. The critical point for present purposes is that there exist mechanisms that can both improve true recognition of study items and produce lesser activation of critical words' representations in semantic memory, resulting in lower levels of critical word false recognition.

While both of the above explanations offer a plausible account of both true and false recognition observed in these studies, as well as broader trends in the literature, we favor the second explanation on the basis of parsimony. In particular, the second explanation adopts a central assumption of bilingual language representation, the degree to which lexical representations activate conceptual representations, and a well-validated notion regarding how effortful processing benefits memory to explain the present results (Hasher & Zacks, 1979). In contrast, the first explanation, while still adopting a key assumption of an influential bilingual representation model, requires considerably more complex assumptions regarding differences in spreading activation that result from translating words to one's dominant language in comparison to directly experiencing a word in one's dominant language. Thus, while both of these viewpoints have the potential to explain the present results, and therefore establish that true recognition data from these studies can be accounted for without compromising activation-based explanations of the false recognition results, the second offers a more compelling, and simpler, explanation in our view. Perhaps most important is that regardless of which viewpoint one adopts, it is possible to explain the true recognition data from these studies, as well as broader trends in the literature, while also accounting for the finding that critical word errors were greater in participants' dominant language compared to their non-dominant language.

In summary, the results of three experiments provide compelling evidence that language proficiency impacts false recognition. The evidence supporting this conclusion stems from both comparisons of false recognition when DRM lists were studied in participants' dominant and non-dominant languages (Experiments 1, 2, & 3), as well as when second language proficiency varied (Experiment 3). In contrast to prior studies, our results cannot be explained by potential stimulus-based artifacts stemming from translation of DRM word lists from English to a second language. Taken together, these data support activation-based explanations of language proficiency effects on false recognition (Carneiro & Fernandez, 2010), as well as developmental theories' claims that greater proficiency with a language increases the automaticity with which lexical representations activate concepts in semantic memory (Howe et al., 2009).

Appendix A

The DRM themes constructed as stimuli for these studies. FAS list strength and BAS list strength are reported for each theme.

Theme	Language	CRITICAL WORDS: associated words (approximated English translation in Spanish lists)	FAS list	BAS list
Alcohol	English	BOTTLE/DRINK/WHISKEY: beer, drunk, liquor, wine, glass, alcohol	0.98	0.72
Death	English	BURIAL/BURY/GRAVE: dead, death, funeral, cemetery, die, ground	1.53	0.32
Farm	English	CROPS/FARMER/HARVEST: corn, farm, food, wheat, field, vegetables	1.39	0.04
Buy	English	PRODUCT/PURCHASE/SALE: buy, item, store, sell, car, price	1.11	0.25
Crime	English	BURGLAR/THEFT/THIEF: steal, robber, crook, crime, criminal, bad	1.48	1.08
Planet	English	JUPITER/NEPTUNE/PLUTO: Mars, Saturn, Venus, Uranus, moon, space	0.83	0.32
School	English	ASSIGNMENT/LESSON/STUDY: homework, work, school, book, class, test	1.39	0.37
Normal	English	NORM/NORMAL/ROUTINE: abnormal, average, same, regular, usual, boring	1.24	0.78
Marriage	Spanish	ALIANZA/COMPROMISO/ENLACE: matrimonio, boda, unión, anillo, amor, casarse (ALLIANCE/COMMITMENT/BOND: marriage, wedding, union, ring, love, marry)	1.67	0.54

Appendix A (*continued*)

Theme	Language	CRITICAL WORDS: associated words (approximated English translation in Spanish lists)	FAS list	BAS list
Hell	Spanish	DEMONIO/INFIERNO/LUCIFER: diablo, fuego, rojo, Satán, maldad, mal (DEMON/HELL/LUCIFER: devil, fire, red, Satan, wickedness, evil)	1.35	0.53
Speed	Spanish	PRISA/VELOCIDAD/VELOZ: rápido, coche, rapidez, carrera, tiempo, lento (HURRY/SPEED/FAST: quick, car, rapidity, race, time, slow)	1.59	0.49
Dress	Spanish	COSER/COSTURA/DESCOSIDO: roto, hilo, aguja, pantalón, vestido, dedal (SEW/SEWING/UNSTITCHED: broken, yarn, needle, trousers, dress, thimble)	1.38	0.93
Sad	Spanish	DEPRIMIDO/INFELIZ/LLORAR: triste, tristeza, lágrima, pena, depresión, llanto (DEPRESSED/UNHAPPY/CRY: sad, sadness, tear, penalty, depression, crying)	1.21	0.38
God	Spanish	PLEGARIA/REZAR/RUEGO: Dios, orar, oración, iglesia, pedir, religión (APPEAL/PRAY/ENTREATY: God, implore, orison, church, to-request, religion)	1.21	0.78
Army	Spanish	CORONEL/GENERAL/OFICIAL: ejercito, soldado, militar, mando, sargento, jefe (COLONEL/GENERAL/OFFICIAL: army, soldier, military, command, sergeant, chief)	1.10	0.13
Injury	Spanish	LESIÓN/MULETA/TOBILLO: pierna, dolor, esguince, rodilla, rotura, codo (INJURY/CRUTCH/ANKLE: leg, pain, sprain, knee, break, elbow)	1.01	0.08

References

Anastasi, J., Rhodes, M., Marquez, S., & Velino, V. (2005). The incidence of false memories in native and non-native speakers. *Memory, 13*(8), 815–828. http://dx.doi.org/10.1080/09658210444000421.

Arndt, J. (2006). Distinctive information and false recognition: The contribution of encoding and retrieval factors. *Journal of Memory and Language, 54*(1), 113–130. http://dx.doi.org/10.1016/j.jml.2005.08.003.

Arndt, J. (2010). The role of memory activation in creating false memories of encoding context. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 36*(1), 66–79. http://dx.doi.org/10.1037/a0017394.

Arndt, J. (2012a). False recollection: Empirical findings and their theoretical implications. *The psychology of learning and motivation* (Vol. 56, pp. 81–124). San Diego, CA, US: Elsevier Academic Press.

Arndt, J. (2012b). The influence of forward and backward associative strength on false recognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 38*(3), 747–756. http://dx.doi.org/10.1037/a0026375.

Arndt, J. (2015). The influence of forward and backward associative strength on false memories for encoding context. *Memory, 23*(7), 1093–1111. http://dx.doi.org/10.1080/09658211.2014.959527.

Arndt, J., & Gould, C. (2006). An examination of two-process theories of false recognition. *Memory, 14*(7), 814–833. http://dx.doi.org/10.1080/09658210600680749.

Beato, M. S., & Arndt, J. (2017). The role of backward associative strength in false recognition of DRM lists with multiple critical words (in preparation).

Beato, M. S., & Arndt, J. (2014). False recognition production indexes in forward associative strength (FAS) lists with three critical words. *Psicothema, 26*(4), 457–463. http://dx.doi.org/10.7334/psicothema2014.79.

Beato, M. S., & Díez, E. (2011). False recognition production indexes in Spanish for 60 DRM lists with three critical words. *Behavior Research Methods, 43*(2), 499–507. http://dx.doi.org/10.3758/s13428-010-0045-9.

Bjork, R. A. (1994). Memory and metamemory considerations in the training of human beings. In J. Metcalfe & A. P. Shimamura (Eds.), *Metacognition: Knowing about knowing* (pp. 185–205). Cambridge, MA, US: The MIT Press.

Brainerd, C. J., Forrest, T. J., Karibian, D., & Reyna, V. F. (2006). Development of the false-memory illusion. *Developmental Psychology, 42*(5), 962–979. http://dx.doi.org/10.1037/0012-1649.42.5.962.

Brainerd, C. J., Reyna, V. F., & Forrest, T. J. (2002). Are young children susceptible to the false-memory illusion? *Child Development, 73*(5), 1363–1377. http://dx.doi.org/10.1111/1467-8624.00477.

Cabeza, R., & Lennartson, R. (2005). False memory across languages: Implicit associative response vs fuzzy trace views. *Memory, 13*(1), 1–5. http://dx.doi.org/10.1080/09658210344000161.

Cadavid, S., & Beato, M. S. (2017). False recognition in DRM lists with low association: A normative study. *Psicológica, 38*, 133–147.

Carneiro, P., Albuquerque, P., Fernandez, A., & Esteves, F. (2007). Analyzing false memories in children with associative lists specific for their age. *Child Development, 78*(4), 1171–1185. http://dx.doi.org/10.1111/j.1467-8624.2007.01059.x.

Carneiro, P., & Fernandez, A. (2010). Age differences in the rejection of false memories: The effects of giving warning instructions and slowing the presentation rate. *Journal of Experimental Child Psychology, 105*(1–2), 81–97. http://dx.doi.org/10.1016/j.jecp.2009.09.004.

Carneiro, P., & Fernandez, A. (2013). Retrieval dynamics in false recall: Revelations from identifiability manipulations. *Psychonomic Bulletin & Review, 20*(3), 488–495. http://dx.doi.org/10.3758/s13423-012-0361-4.

Carneiro, P., Fernandez, A., Diez, E., Garcia-Marques, L., Ramos, T., & Ferreira, M. B. (2012). "Identify-to-reject": A specific strategy to avoid false memories in the DRM paradigm. *Memory & Cognition, 40*(2), 252–265. http://dx.doi.org/10.3758/s13421-011-0152-6.

Corson, Y., & Verrier, N. (2007). Emotions and false memories: Valence or arousal? *Psychological Science, 18*(3), 208–211. http://dx.doi.org/10.1111/j.1467-9280.2007.01874.x.

Corson, Y., Verrier, N., & Bucic, A. (2009). False memories and individual variations: The role of field dependence-independence. *Personality and Individual Differences, 47*(1), 8–11. http://dx.doi.org/10.1016/j.paid.2009.01.036.

Cycowicz, Y. M., Friedman, D., Snodgrass, J. G., & Duff, M. (2001). Recognition and source memory for pictures in children and adults. *Neuropsychologia, 39*(3), 255–267. http://dx.doi.org/10.1016/S0028-3932(00)00108-1.

Deese, J. (1959). On the prediction of occurrence of particular verbal intrusions in immediate recall. *Journal of Experimental Psychology, 58*(1), 17–22. http://dx.doi.org/10.1037/h0046671.

Dubuisson, J.-B., Fiori, N., & Nicolas, S. (2012). Repetition and spacing effects on true and false recognition in the DRM paradigm. *Scandinavian Journal of Psychology, 53*(5), 382–389. http://dx.doi.org/10.1111/j.1467-9450.2012.00963.x.

Graves, D. F., & Altarriba, J. (2014). False memories in bilingual speakers. In *Foundations of bilingual memory* (pp. 205–221). New York, NY: Springer.

Fernández, A., Díez, E., & Alonso, M. A. (2009). Normas de Asociación libre en castellano, online database [Free Association norms in Spanish]. Retrieved from: <http://www.usal.es/gimc>.

Hasher, L., & Zacks, R. M. (1979). Automatic and effortful processes in memory. *Journal of Experimental Psychology: General, 108*, 356–388. http://dx.doi.org/10.1037/0096-3445.108.3.356.

Hernandez, A., Li, P., & MacWhinney, B. (2005). The emergence of competing modules in bilingualism. *Trends in Cognitive Sciences, 9*(5), 220–225. http://dx.doi.org/10.1016/j.tics.2005.03.003.

Howe, M. L. (2006). Developmentally invariant dissociations in children's true and false memories: Not all relatedness is created equal. *Child Development, 77*(4), 1112–1123.

Howe, M. L., Gagnon, N., & Thouas, L. (2008). Development of false memories in bilingual children and adults. *Journal of Memory and Language, 58*(3), 669–681. http://dx.doi.org/10.1016/j.jml.2007.09.001.

Howe, M. L., Wimmer, M. C., Gagnon, N., & Plumpton, S. (2009). An associative-activation theory of children's and adults' memory illusions. *Journal of Memory and Language, 60*(2), 229–251. http://dx.doi.org/10.1016/j.jml.2008.10.002.

Kawasaki, Y., & Yama, H. (2006). The difference between implicit and explicit associative processes at study in creating false memory in the DRM paradigm. *Memory, 14*(1), 68–78. http://dx.doi.org/10.1080/09658210444000520.

Kawasaki-Miyaji, Y., Inoue, T., & Yama, H. (2003). Cross-linguistic false recognition: How do Japanese-dominant bilinguals process two languages: Japanese and English? *Psychologia: An International Journal of Psychology in the Orient, 46*(4), 255–267. http://dx.doi.org/10.2117/psysoc.2003.255.

Kroll, J. F., & Stewart, E. (1994). Category interference in translation and picture naming: Evidence for asymmetric connection between bilingual memory representations. *Journal of Memory and Language, 33*(2), 149–174. http://dx.doi.org/10.1006/jmla.1994.1008.

Li, P., Farkas, I., & MacWhinney, B. (2004). Early lexical development in a self-organizing neural network. *Neural Networks, 17*(8–9), 1345–1362. http://dx.doi.org/10.1016/j.neunet.2004.07.004.

Marmolejo, G., Diliberto-Macaluso, K. A., & Altarriba, J. (2009). False memory in bilinguals: Does switching languages increase false memories? *The American Journal of Psychology, 122*(1), 1–16.

Miyaji, Y., & Yama, H. (2002). Making Japanese lists which induce false memory at high probability for the DRM paradigm. *The Japanese Journal of Psychonomic Science, 21*(1), 21–26.

Nelson, D. L., McEvoy, C. L., & Schreiber, T. A. (1998). The University of South Florida word association, rhyme, and word fragment norms. <http://www.usf.edu/FreeAssociation/>.

Roediger, H. L., & McDermott, K. B. (1995). Creating false memories: Remembering words not presented in lists. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 21*(4), 803–814. http://dx.doi.org/10.1037/0278-7393.21.4.803.

Roediger, H. L., Watson, J. M., McDermott, K. B., & Gallo, D. A. (2001). Factors that determine false recall: A multiple regression analysis. *Psychonomic Bulletin & Review, 8*(3), 385–407. http://dx.doi.org/10.3758/BF03196177.

Sahlin, B. H., Harding, M. G., & Seamon, J. G. (2005). When do false memories cross language boundaries in English-Spanish bilinguals? *Memory & Cognition, 33*(8), 1414–1421. http://dx.doi.org/10.3758/BF03193374.

Sergi, I., Senese, V. P., Pisani, M., & Nigro, G. (2014). Assessing activation of true and false memory traces: A study using the DRM paradigm. *Psychologica Belgica, 54*(1), 171–179. http://dx.doi.org/10.5334/pb.ak.

Stadler, M. A., Roediger, H. L., & McDermott, K. B. (1999). Norms for word lists that create false memories. *Memory & Cognition, 27*(3), 494–500. http://dx.doi.org/10.3758/BF03211543.

Van Damme, I., & d'Ydewalle, G. (2009). Memory loss versus memory distortion: The role of encoding and retrieval deficits in Korsakoff patients' false memories. *Memory, 17*(4), 349–366. http://dx.doi.org/10.1080/09658210802680349.