Psychological Research (2021) 85:3134–3151
https://doi.org/10.1007/s00426-020-01449-3

ORIGINAL ARTICLE

The effect of language proficiency and associative strength on false memory

Maria Soledad Beato[1] · Jason Arndt[2]

Received: 10 June 2020 / Accepted: 9 November 2020 / Published online: 2 January 2021

Abstract

In two experiments we examined the role of language proficiency and associative strength in the production of false memory. We constructed Deese–Roediger–McDermott lists using both Spanish and English free association norms. Lists were constructed to vary in backward associative strength (BAS). Experiment 1 participants were native Spanish speakers with some proficiency in English while Experiment 2 participants were native Spanish participants that had either high, intermediate, or low English proficiency. Results showed that, in both Experiment 1 and Experiment 2, false recognition was greater in participants' dominant language (L1 or Spanish) than in their non-dominant language (L2 or English), and false recognition in L2 increased with L2 proficiency when low-BAS lists were studied (Experiment 2). Further, false recognition was higher in high-BAS lists than in low-BAS lists in both L1 and L2. Finally, we collected a measure of participants' knowledge of our stimulus words in L2. These data showed that participants had far from perfect knowledge of all L2 stimuli. Analyses that factored out the effects of L2 word knowledge failed to alter the effects of L1 vs. L2, L2 proficiency and BAS on false recognition.

Introduction

False memory has been the subject of intensive research focus, both for applied purposes, such as understanding how reliable people's memories of witnessed crimes will be (e.g., Calado, Otgaar, & Muris, 2018; Otgaar, de Ruiter, Howe, Hoetmer, & van Reekum, 2017; Otgaar, Howe, Muris, & Merckelbach, 2019), as well as for theoretical purposes, such as understanding the memory processes that underlie accurate and inaccurate remembering (Beato & Cadavid, 2016; Cadavid & Beato, 2016; Huff & Bodner, 2013; Lim & Goh, 2019; Wang, Otgaar, Howe, Lippe, & Smeets, 2018). A common paradigm that has been used to examine true and false memory is the Deese–Roediger–McDermott (DRM) paradigm (Deese, 1959; Roediger & McDermott, 1995). In this paradigm, participants are presented with a series of words to study (e.g., crib, infant, cradle, diaper) that are all associated with a single word that is not shown during study (e.g., baby), known as the critical lure. People generally falsely recall and recognize critical lures with a high probability (Arndt, 2012a, 2015; Beato & Arndt, 2014; Roediger & McDermott, 1995) and believe they can recollect critical lures (Roediger & McDermott, 1995; see Arndt, 2012b for a review).

A key question in understanding false memory is how experience-based changes in semantic and/or associative memory structures impact false memory generation. Two of the ways that this question has been investigated are to examine how false memory changes across early cognitive development (e.g., Brainerd, Reyna, & Forrest, 2002; Howe, Gagnon, & Thouas, 2008) and how false memory differs as a function of language proficiency within adults (e.g., Anastasi, Rhodes, Marquez, & Velino, 2005; Arndt & Beato, 2017). Regarding development, false memory generally increases as cognitive development progresses such that younger children show lower levels of false memory than older children, and older children show lower levels of false memory than young adults (e.g., Brainerd, Forrest, Karibian, & Reyna, 2006; Dewhurst & Robinson, 2004; Howe, Wimmer, & Blease, 2009). Regarding language proficiency, false memory is generally greater in participants' dominant language (L1 hereafter) than their non-dominant language (L2 hereafter; Anastasi et al., 2005; Arndt & Beato, 2017;

✉ Maria Soledad Beato
msol@usal.es

1 Faculty of Psychology, University of Salamanca, 37005 Salamanca, Spain

2 Department of Psychology, Middlebury College, 276 Bicentennial Hall, Middlebury, VT 05753, USA

Howe et al., 2008; Sahlin, Harding, & Seamon, 2005; but see Cabeza & Lennartson, 2005 for an exception), and false memory increases as people become more proficient in their L2 (Arndt & Beato, 2017).

Two general viewpoints have been advanced to explain why false memory increases with development, and these ideas also have been applied to explaining why false memory increases with language proficiency. One explanation is that improved knowledge of words in a language is responsible for increased false memory in adults compared to children (Carneiro & Fernandez, 2010), as well as L1 vs. L2 differences and L2 proficiency differences in false memory. This explanation suggests that as one's vocabulary increases (e.g., with development or increased learning of a second language), so does the likelihood a studied word's referent will be understood. In turn, this viewpoint proposes that knowing a greater number of study words' referents means that more concepts related to critical lures will be activated during encoding, leading to increased potential for false memory. Thus, this view argues that the primary basis for false memory differences between adults and children is that children will understand fewer words than adults, and thus critical lures will have fewer concepts to activate them in semantic memory. Similarly, this explanation suggests that L1 vs. L2 differences in false memory occur because participants will tend to know the meaning of fewer words in L2, which leads to fewer concepts activating critical lures' representations in semantic memory. In effect, this viewpoint proposes that developmental and language proficiency effects are akin to manipulations of the number of associates that are related to a critical lure that are encoded, which reliably increases false memory (Arndt, 2010; Arndt & Gould, 2006; Robinson & Roediger, 1997).

A second explanation, derived from associative activation theory, is that adults are more susceptible to false recognition than children because their associations in semantic memory are stronger and better organized, which in turn leads to greater automatic activation of concepts associated with study items during encoding (e.g., Howe, 2006; Howe, Wimmer, Gagnon, et al., 2009). As a result, related information, such as critical lures in the DRM paradigm, are activated more strongly and thus are more likely to be falsely remembered by adults than children. Similarly, this account suggests that the more proficient a person is in a language, the more strongly and automatically studied concepts, and thus related but unstudied concepts such as critical lures, will be activated during encoding. Thus, this theory can also explain why false memory is greater in participants' L1 than their L2 (Anastasi et al., 2005; Arndt & Beato, 2017; Howe et al., 2008; Sahlin et al., 2005), as well as why false memory increases with participants' proficiency in L2 (Arndt & Beato, 2017). In particular, study items' lexical representations will activate their associated representations in semantic memory more fully and automatically in L1 than L2, and the more proficient a person is in L2, the more strongly and automatically lexical representations in that language will activate representations in semantic memory. In turn, increased activation of studied items' representations in semantic memory will enable greater activation of related concepts, such as those of critical lures.

While much of the literature favors the view that increases in false memory with development and language proficiency are best explained by associative activation theories (Arndt & Beato, 2017; Howe, 2006; Howe, Wimmer, Gagnon, et al., 2009), it remains plausible that some or all of the increase in false memory across development and language proficiency can be explained by participants' word knowledge in the language they are learning (L1 for children, L2 for adults). In particular, studies that have examined developmental changes in false memory (e.g., Anastasi & Rhodes, 2008; Carneiro & Fernandez, 2010; Howe, Wimmer, Gagnon, et al., 2009; Knott, Howe, Wimmer, & Dewhurst, 2011; Metzger, Warren, Shelton, Price, Reed, & Williams, 2008; Sugrue & Hayne, 2006), false memory differences between L1 and L2 (Anastasi et al., 2005; Arndt & Beato, 2017; Howe et al., 2008; Sahlin et al., 2005), and language proficiency differences in false memory (Arndt & Beato, 2017) generally have not assessed participants' word knowledge.[1] Thus, this alternative explanation for developmental and language proficiency effects has not been directly evaluated in prior studies.

Implicit in the fact that prior studies have not assessed word knowledge is the belief that participants with different levels of language proficiency have comparable knowledge of the words shown as stimuli, at least in the sense that they would be able to access all words' conceptual representations in semantic memory. If it were the case that word knowledge of stimuli used in prior studies is similar across groups (e.g., L1 vs. L2; proficiency differences within L2), logically that would leave differences in spreading activation between studied items and lure items as the most likely basis for differences in false memory. However, if there are differences in the extent to which people know the meaning of words across groups, it would leave open the possibility that associative activation is not the sole basis for language proficiency differences in false memory. Thus, it is important to ensure that word knowledge for the stimuli used

[1] Knott et al. (2011) addressed this issue in their research by ensuring that the age-related frequency for words in their stimuli were equated across age groups. Thus, word knowledge differences may be less likely to complicate the age-related false memory differences they found. Nevertheless, this concern applies to other work that did not construct stimuli as purposefully, including all studies that have examined language dominance and second language proficiency to date.

in experiments does not differ across L1 and L2 or across groups of participants that differ in L2 proficiency in order to infer that associative activation explains language proficiency effects. This consideration was a primary motivating factor for the present research.

A second goal of the present research was to examine the generality of existing findings that show language dominance and proficiency effects on false memory. In particular, prior research examining L1 vs. L2 differences and L2 proficiency differences in false recognition presented participants with study words that were strongly associated with critical lures (Anastasi et al., 2005; Arndt & Beato, 2017; Howe et al., 2008; Sahlin et al., 2005). In the present research, we examined if these results generalized to stimuli that are more weakly associated with critical lures by manipulating the backward associative strength (BAS; Roediger, Watson, McDermott, & Gallo, 2001) between studied items and critical lures. One of the primary bases for the claim that associative activation is a key factor in producing false memory is that manipulations of BAS produce strong, reliable effects on critical lure false memory (Arndt & Gould, 2006; Gallo & Roediger, 2003; McEvoy, Nelson, & Komatsu, 1999; Roediger et al., 2001). From the perspective of associative activation theory, this result stems from strong associates activating critical lures to a greater extent than weak associates. While this explanation has been evaluated in L1, parallel examinations of the effects of BAS on false memory in L2 and false memory across levels of language proficiency within L2 have not been conducted to date. Thus, manipulating BAS will evaluate if the same basic associative processes that have been used to explain BAS's effect on false memory in L1 also explain false memory differences between L1 and L2 and changes in false memory with L2 proficiency.

The importance of examining whether manipulating BAS produces the same effects in L2 as in L1 can be found in contemporary models of bilingual language representation. While models of bilingualism differ in the exact nature of the representations that underlie L1 and L2, they agree that L1 and L2 lexicons access a common semantic memory store (Dijkstra & Van Heuven, 2002; Green, 1998; Kroll & Stewart, 1994; Li, Farkas, & MacWhinney, 2004; see Francis, 1999 for a review of evidence suggesting that L1 and L2 lexicons access common semantic memory representations). Thus, the knowledge representations activated by lexicons in L1 and L2 share the same semantic and associative properties. In turn, this implies that variables that are thought to impact the extent to which unstudied concepts, such as critical lures, are activated during encoding will be similar across lists that are studied in L1 and L2. As a result, the primary difference between encoding in L1 and L2 should be (1) the extent to which study items' meaning is known to participants (Carneiro & Fernandez, 2010) and/or (2) the extent to which study items' lexical representations activate their conceptual representations in semantic memory (Arndt & Beato, 2017; Howe, Wimmer, Gagnon, et al., 2009).

In order to accomplish these two goals, we examined false memory in the DRM paradigm using lists that were strongly (high-BAS lists) or weakly (low-BAS lists) related to critical lures and presented those lists in participants' L1 or L2 (Experiment 1). Following the completion of a memory test, participants were tested on their L2 word knowledge by asking them to translate words from L2 to L1. Experiment 2 replicated this basic design and evaluated the impact that language proficiency in L2 had on false recognition for high- and low-BAS lists.

The expectation from associative activation theories of false memory is that critical lures will be falsely recognized more often following the study of high-BAS than low-BAS lists and that this outcome should occur when lists are studied and tested in L1 as well as L2. Similarly, associative activation theories expect the effect of BAS on false memory to occur regardless of participants' language proficiency in L2. As reviewed above, these predictions stem from the view that L1 and L2 access common semantic memory representations, and thus activate critical lures in a similar way once study items activate their conceptual representations in semantic memory. In addition, associative activation theories expect false memory to be greater in L1 than L2, and to increase with L2 proficiency. Further, the effects of studying stimuli in L1 vs. L2 and L2 proficiency on false recognition should occur both when strong BAS lists are studied and when weak BAS lists are studied. These predicted findings stem from the view that study items' conceptual representations will be activated more strongly and automatically when lists are studied in L1 than L2, and when participants are more proficient in L2. In turn, these differences in study item activation between L1 and L2, as well as across L2 proficiency, should produce greater activation of critical lures' representations in semantic memory, resulting in greater false memory.

After participants completed the memory task, we assessed their word knowledge for low- and high-BAS lists in L2. This allowed us to determine the extent to which knowledge of words' meaning in L2 explained some, or all, of the false memory effects observed in these studies (Carneiro & Fernandez, 2010). To illustrate the utility of examining word knowledge, consider the likely outcome that participants produce greater false recognition in L1 than L2 (Anastasi et al., 2005; Arndt & Beato, 2017; Howe et al., 2008; Sahlin et al., 2005). Associative activation theory explains this finding by suggesting that studying lists in L1 activates critical lures' representations in semantic memory more than studying lists in L2. However, it is also possible that study item word knowledge is lower in L2 than in L1 (Carneiro & Fernandez, 2010). If this were the case, L1 vs. L2 differences could occur because participants do not have a connection between lexical and semantic representations for some L2 words, and thus studying words in L2 fails to activate a representation in semantic memory. Similar considerations apply to interpreting false memory differences that are associated with language proficiency in L2, such that participants with greater proficiency in L2 may simply know more study words in L2 than participants with lower L2 proficiency. Thus, the examination of L2 word knowledge in the present studies will allow us to evaluate the extent to which word knowledge explains the false memory effects found in these studies. In turn, examining word knowledge will allow us to determine how fully associative activation theories explain false memory differences between participants' L1 and L2, as well as the effect of L2 proficiency on false memory.

Experiment 1

The goals of Experiment 1 were to (1) examine if the effect of BAS occurred in both L1 and L2, (2) examine whether L1 vs. L2 differences in false memory occurred both when study items were strongly associated with critical lures and when study items were weakly associated with critical lures, and (3) evaluate the extent to which differences in false memory across BAS, as well as L1 vs. L2 differences in false memory, were explained by word knowledge. Thus, we generated DRM word lists in both Spanish and English, using norms appropriate for each language (Fernández, Díez, & Alonso, 2009; Nelson, McEvoy & Schreiber, 1998, respectively). These lists were constructed such that each critical lure was related to a set of high-BAS associates and a set of low-BAS associates. Further, we controlled BAS within each set of associates across the Spanish and English language lists to ensure that L1 and L2 associative strength was comparable (Arndt & Beato, 2017). This method of constructing stimuli allowed us to infer that any differences between lists studied in L1 and L2 were due to language dominance and not associative strength differences that can result from translating lists directly from one language (e.g., Spanish) to another (e.g., English; see Arndt & Beato, 2017 for a discussion of this issue in bilingual false memory research).

Method

Participants

Ninety undergraduate students participated voluntarily in exchange for course credit and signed a consent form.[2] They were native Spanish speakers (89% women) with ages ranging from 21 to 38 years (M age = 21.76, SD = 1.97). On average, participants rated their proficiency in English as above average (M = 6.02, SD = 1.78, range = 2–10) on a scale ranging from 1 (elementary knowledge) to 10 (native speaker proficiency).

Materials

We constructed 32 DRM word lists, 16 in Spanish and 16 in English. The lists were constructed using Spanish and English free association norms, respectively (Fernández et al., 2009; Nelson et al., 1998). Lists were composed of a critical lure and eight associates with nonzero backward associative strength (BAS) to a critical lure. Four of the associates of each critical lure had relatively high BAS and the other four associates of each critical lure had relatively low BAS. Lists were constructed such that they maximized the probability that studied items produced the critical lure in free association for the high-BAS studied associates, and ensured that the probability low-BAS associates produced the critical lure in free association was lower than any of the high-BAS associates for a given lure item. For example, for the critical lure BUG, the high BAS associates were beetle, insect, termite, and pest (mean BAS = 0.442), while the low BAS associates were irritate, ant, tick, and annoy (mean BAS = 0.077). Finally, lists were constructed to ensure that (1) forward associative strength (FAS), the probability that critical lures produced study items in free association, was controlled across high- and low-BAS study lists and (2) both BAS and FAS were equated across lists constructed from Spanish free association norms and English free association norms. This method of list construction ensured that the effects of BAS on critical lure false memory were not due to item differences that could occur if high- and low-BAS lists were associated with distinct critical lures.

The mean backward associative strength values of the Spanish high-BAS list ranged from 0.420 to 0.533 (M = 0.466), and the mean backward associative strength values of the Spanish low-BAS list ranged from 0.045 to 0.079 (M = 0.062). The mean backward associative strength values of the English high-BAS lists ranged from 0.435 to 0.492 (M = 0.465), and the mean backward associative strength values of the English low-BAS list ranged from 0.050 to 0.082 (M = 0.066). The full set of stimuli used in these studies are reported in the "Appendix".

We confirmed that lists had comparable BAS values in Spanish and English by analyzing BAS with a 2 (associative strength: high-BAS vs. low-BAS) × 2 (language: Spanish vs. English) repeated-measures ANOVA. This analysis revealed a significant main effect of associative strength, $F(1, 252) = 680.26$, $p < 0.001$, $\eta^2_p = 0.730$, no significant main effect of language, $F(1, 252) = 0.01$, $p = 0.932$, $\eta^2_p =$

[2] One participant was excluded from all analyses due to a computer error in the recording of their data.

Table 1 Mean associative strength (BAS and FAS) as a function of language (Spanish vs. English) in lists used in Experiment 1 and Experiment 2

	BAS		FAS	
	High	Low	High-BAS lists	Low-BAS lists
Spanish	0.466 [0.44, 0.50]	0.062 [0.03, 0.09]	0.037 [0.02, 0.05]	0.027 [0.01, 0.04]
English	0.465 [0.43, 0.50]	0.066 [0.04, 0.10]	0.037 [0.02, 0.05]	0.033 [0.02, 0.05]

95% CIs are reported in square brackets

0.001, and no associative strength × language interaction, $F(1, 252) = 0.02$, $p = 0.880$, $\eta^2_p = 0.001$. The main effect of associative strength indicated that high-BAS lists had higher backward associative strength than low-BAS lists (0.47 vs. 0.06, respectively), 95% CI [0.37, 0.43]. The absence of a main effect of language or an interaction between associative strength and language indicated that Spanish and English DRM lists were similar in associative strength between study items and critical lures, as is evident from inspection of the mean BAS values in Table 1.

Further, we verified the similarity in FAS across conditions using a 2 (associative strength: high-BAS vs. low-BAS) × 2 (language: Spanish vs. English) repeated-measures ANOVA. This analysis failed to produce a main effect of associative strength, $F(1, 252) = 1.09$, $p = 0.297$, $\eta^2_p = 0.004$, a main effect of language, $F(1, 252) = 0.23$, $p = 0.634$, $\eta^2_p =$ 0.001, or a BAS × language interaction, $F(1, 252) = 0.26$, $p = 0.611$, $\eta^2_p = 0.001$. Therefore, FAS was similar in high-BAS lists and low-BAS lists, both in Spanish and English.

Procedure

Participants were tested in groups of 23–25, and completed the experiment individually on a computer with the supervision of an experimenter for the duration of the experiment. Prior to beginning the experiment, all participants signed a consent form for this experiment. The research protocol was approved by the Bioethics Committee at the University of Salamanca. The experimenter read the instructions aloud while participants followed along on their individual computer screens. Then, participants began the study phase at the same time. Participants studied 96 words, half in Spanish and half in English (24 lists of four words: 6 Spanish high-BAS, 6 Spanish low-BAS, 6 English high-BAS, 6 English low-BAS), with words presented one at a time on a computer screen for 2 s. The order of study lists was randomly determined. The words within each list were always presented in decreasing order of BAS values. Participants were told to read the words and do their best to remember the words that were presented to them in preparation for an unspecified memory test. Lists were assigned to be studied and unstudied, and to the low or high BAS conditions equally often.

After the study phase, participants completed an old-new recognition test. The 96-item recognition test was composed of 48 studied words (two per study list, first and third word in each study list), and 48 unstudied words. The unstudied words were the 24 critical lures related to studied lists, 8 unrelated critical distractors, which were the critical lures from the non-studied lists, and 16 unrelated distractors, which were the first and third word in 8 non-studied lists, taken from the low and high BAS conditions equally often. Half of the items within each type of word were tested in Spanish, with the other half being tested in English. Thus, studied words were tested in the language in which they were studied, critical lures were tested in the language in which their associates were studied, and half of the critical distractors and unrelated distractors were tested in each language. Recognition test items were presented in random order. Finally, after the recognition memory test, participants completed a translation test where all the studied words and critical lures in English were included. For this task, we gave participants a sheet with the words in English and their task was to write the Spanish translation for each word. Stimulus lists served equally often in the high-BAS, low-BAS and unstudied conditions across participants.

Results and discussion

Translation test: Did participants know the words included in the lists?

A key aim of this experiment was to evaluate if word knowledge explained L1 vs. L2 differences in false memory or high- vs. low-BAS effects, if they occurred, in L2. To examine these questions, we analyzed the English to Spanish translation data to determine (1) the extent to which participants knew the stimulus words in L2 generally and (2) how well participants knew the stimulus words included in high-BAS lists and low-BAS lists in L2.

The analysis of English to Spanish translation data indicated that L2 knowledge of stimulus words was far from perfect, with participants knowing the translation of 63% of study words on average. This finding leaves open the possibility that some or all of the false memory differences we expect to find between L1 and L2 (Anastasi et al., 2005; Arndt & Beato, 2017; Howe et al., 2008; Sahlin et al., 2005) could be explained by word knowledge differences instead of associative activation differences. Further, participants knew a smaller proportion of high-BAS ($M = 0.53$) than low-BAS ($M = 0.73$) study list words in L2, $t(95) = 14.79$, $p < 0.001$,

Table 2 Mean proportion of false recognition and false alarms to unrelated critical distractors in Experiment 1 as a function of language and BAS

	L1 (Spanish)		L2 (English)	
	High BAS	Low BAS	High BAS	Low BAS
False recognition	0.42 [0.37, 0.47]	0.23 [0.19, 0.27]	0.23 [0.19, 0.28]	0.19 [0.15, 0.22]
Unrelated critical distractors	0.15 [0.09, 0.20]	0.10 [0.06, 0.14]	0.12 [0.07, 0.18]	0.16 [0.10, 0.22]

95% CIs are reported in square brackets

Cohen's $d = 1.29$, 95% CI [0.18, 0.23]. Thus, the translation data suggest that word knowledge can only explain differences in false recognition between high- and low-BAS lists studied in L2 if low-BAS items produce higher rates of false recognition, because participants knew *more* low-BAS than high-BAS study words. That is, in order for word knowledge to explain the effects of BAS on false recognition in L2, its effects would have to be in the opposite direction of the robust and reliable effects of BAS that are found in L1 (Arndt, 2012a, 2015; Arndt & Gould, 2006; Gallo & Roediger, 2003; Roediger et al., 2001).

False memory effect

The mean false alarm rates to critical lures are presented in Table 2 as a function of language (L1 vs. L2) and BAS (high vs. low). In order to evaluate the effects of these factors on false memory, we analyzed the proportion of false alarms to critical lures using a 2 (language: L1 [Spanish] vs. L2 [English]) × 2 (BAS: high vs. low) within-subjects ANOVA. Furthermore, in order to investigate the extent to which word knowledge in L2 explains L1 vs. L2 false recognition differences, we used analysis of covariance (ANCOVA) to control for the variance in false memory due to L2 word knowledge that occurred in this experiment. To generate the covariates used in our ANCOVA analyses, we computed a composite translation score based upon the proportion of all studied items participants knew, a composite translation score based upon the proportion of all critical lures that participants knew, and separate translation scores for studied items in the high- and low-BAS conditions.[3] Translation scores were centered (Delaney & Maxwell, 1981) by subtracting each participant's score from the appropriate mean translation score (i.e., the mean translation score for all study items or for all lure items for the composite translation scores and the mean high- or low-BAS translation score for separate translation scores). We then entered these covariates in a series of 2 (language) × 2 (BAS) within-subjects ANCOVAs to examine the impact each covariate had on the statistical effects found in the ANOVA without covariates. Finally, we ran ANCOVAs examining the simple effect of language (L1 vs. L2) on false memory following the study of high-BAS lists (using high-BAS study item word knowledge as a covariate), and the effect of language on false memory following the study of low-BAS lists (using low-BAS study item word knowledge as a covariate).

This analysis strategy provides two ways to assess the impact that L2 word knowledge had on false memory differences in this experiment. First, we evaluated if any of the covariates explained a signifcant amount of the variance in false memory. Second, we evaluated whether including any of the covariates in the language × BAS analyses changed the pattern of statistical effects by comparing the ANCOVA analysis results with the ANOVA results (i.e., without covariates included). Regarding the first assessment, none of the measures of L2 knowledge were significant as covariates, although critical lure knowledge approached significance, $F(1, 87) = 3.55$, MSE = 0.059, $p = 0.063$, $\eta^2_p = 0.039$. Second, and as expected, given the covariates did not explain a significant amount of variance in the ANCOVAs, all of the statistically-significant effects observed in our ANOVA analyses remained significant in our ANCOVA analyses. For simplicity, we report the ANOVA-based analyses of critical lure false alarms.

The ANOVA on false recognition revealed a significant main effect of language, showing there was greater false recognition in L1 ($M = 0.33$) than L2 ($M = 0.21$), $F(1, 88) = 37.90$, MSE = 0.032, $p < 0.001$, $\eta^2_p = 0.301$, a significant main effect of BAS, showing that false recognition was greater following study of high-BAS ($M = 0.33$) than low-BAS lists ($M = 0.21$), $F(1, 88) = 45.26$, MSE = 0.028, $p < 0.001$, $\eta^2_p = 0.340$, and a significant language × BAS interaction, $F(1, 88) = 16.94$, MSE = 0.027, $p < 0.001$, $\eta^2_p = 0.161$. The interaction between language and BAS indicated that the effect of BAS on false recogniton was greater in L1 than in L2. However, and relevant for evaluating the first aim of this experiment, false recognition was greater in high-BAS lists than in low-BAS lists in both L1, $p < 0.001$, 95%

[3] We did not use separate translation scores for critical lures assigned to the low- and high-BAS conditions, because study lists were constructed such that each critical lure had low- and high-BAS associates, and each participant studied either the low- or high-BAS associates for a given lure. Thus, the same lure items were shown in the low- and high-BAS conditions across participants, making a composite measure of critical lure translation the most appropriate way of measuring critical lure knowledge.

Table 3 Mean proportion of true recognition and false alarms to unrelated distractors in Experiment 1 as a function of language and BAS

	L1 (Spanish)		L2 (English)	
	High BAS	Low BAS	High BAS	Low BAS
True recognition	0.79 [0.75, 0.83]	0.77 [0.73, 0.80]	0.86 [0.83, 0.89]	0.81 [0.78, 0.84]
Unrelated distractors	0.06 [0.03, 0.08]	0.10 [0.06, 0.13]	0.13 [0.09, 0.17]	0.10 [0.07, 0.13]

95% CIs are reported in square brackets

CI [0.14, 0.25], and L2, $p = 0.033$, 95% CI [0.004, 0.09]. In relation to our second aim for this study, to examine the generality of language proficiency effect on false memory, false recognition was higher in L1 than in L2, both when high-BAS lists were studied, $p < 0.001$, 95% CI [0.13, 0.24], and when low-BAS lists were studied, $p = 0.043$, 95% CI [0.001, 0.09], although the difference in false recognition between L1 and L2 was greater following the study of high-BAS lists (see Table 2).

We also examined the simple effect of language on false recognition in the high-BAS condition using high-BAS study item word knowledge as a covariate, and in the low-BAS condition using low-BAS study item word knowledge as a covariate. These analyses were conducted to evaluate if word knowledge was able to explain the L1 vs. L2 differences in false recognition for either the high-BAS condition or the low-BAS condition that were found in the full factorial analyses. Specifically, for false recognition in the high-BAS condition, the effect of language remained significant, $F(1, 87) = 41.84$, $\text{MSE} = 0.002$, $p < 0.001$, $\eta^2_p = 0.325$, and high-BAS L2 study item knowledge was not a significant covariate, $F(1, 87) < 1$. Similarly, for false recognition in the low-BAS condition, the effect of Language remained significant, $F(1, 87) = 4.14$, $\text{MSE} = 0.020$, $p = 0.045$, $\eta^2_p = 0.045$, and low-BAS L2 study item knowledge was not a significant covariate, $F(1, 87) = 2.87$, $p = 0.094$.

In addition to the finding that L2 word knowledge was not a significant covariate in any of the analyses of false memory data, and that including L2 word knowledge as a covariate failed to alter the statistical outcomes of ANOVA-based analyses of our false memory data, it is also important to note that the magnitude of the statistical effects found in ANOVA-based analyses was largely unchanged when L2 word knowledge was included in our ANCOVA analyses. Thus, L2 word knowledge not only failed to alter the statistical main effects found in Experiment 1, it tended to not alter the magnitude of those effects either.

Finally, we conducted an exploratory analysis to examine the correlation between the four covariates and false recognition in L2. These analyses showed that none of the covariates were positively correlated with false recognition in L2. Indeed, the only significant correlations between the covariates and false recognition were negative: All four covariates were negatively correlated with low-BAS false recognition in L2, $r(87)$ ranged from -0.281 ($p = 0.008$) to -0.312 ($p = 0.003$). Importantly, the lack of positive correlations between the covariates and L2 false recognition was not because false recognition scores failed to correlate positively with any variable, since high- and low-BAS false recognition were positively correlated in L2, $r(87) = 0.313$, $p = 0.003$. Similarly, the four measures of L2 word knowledge that were used as covariates were positively correlated with one another [smallest $r(87) = 0.523$, $p < 0.001$ for the correlation between critical lure word knowledge and low-BAS study item knowledge].

True recognition

Mean proportions of true recognition and false alarms to unrelated distractors are presented in Table 3 as a function of language (L1 vs. L2) and BAS (high vs. low). We analyzed hit rates for studied items (i.e., true recognition) using a 2 (language: L1 vs. L2) × 2 (BAS: high vs. low) ANOVA. This analysis revealed reliable main effects of language, $F(1, 88) = 14.16$, $\text{MSE} = 0.022$, $p < 0.001$, $\eta^2_p = 0.139$ and BAS, $F(1, 88) = 8.05$, $\text{MSE} = 0.013$, $p = 0.006$, $\eta^2_p = 0.084$. The main effect of language showed that hit rates were greater when items were studied in L2 ($M = 0.84$) than in L1 ($M = 0.78$). The main effect of BAS showed that hit rates were greater for high-BAS study items ($M = 0.82$) than low-BAS study items ($M = 0.79$). The interaction was not reliable, $F(1, 88) = 1.90$, $\text{MSE} = 0.011$, $p = 0.172$, $\eta^2_p = 0.021$.

Baseline false alarm rates

Finally, we analyzed the baseline false alarm rates for unrelated critical distractors (Table 2) and for unrelated distractors (Table 3) using 2 (language) × 2 (BAS) ANOVAs. The analysis of unrelated critical distractor false alarms failed to produce any significant main effects or an interaction (all $p > 0.135$). The analysis of unrelated distractor false alarm rates produced a main effect of language, $F(1, 88) = 14.16$, $\text{MSE} = 0.022$, $p < 0.001$, $\eta^2_p = 0.139$, and an interaction between language and BAS, $F(1, 88) = 8.13$, $\text{MSE} = 0.013$, $p = 0.005$, $\eta^2_p = 0.085$. The main effect of language indicated that unrelated distractor false alarms were lower in L1 ($M = 0.08$) than in L2 ($M = 0.12$). The interaction occurred because low-BAS unrelated distractors produced higher false alarm rates than high-BAS unrelated distractors in L1, $t(88) = 2.32$, $p = 0.023$, while the difference between high and low-BAS unrelated distractor false alarms was not significant in L2, $t(88) = 1.56$, $p = 0.122$.

In summary, we found higher rates of false recognition in L1 than in L2 in high-BAS lists, replicating previous research (Anastasi et al., 2005; Arndt & Beato, 2017; Howe et al., 2008; Sahlin et al., 2005). Further, Experiment 1 documented that this effect extended to low-BAS lists and that false recognition was greater in high-BAS lists than low-BAS lists in participants' L1 as well as in their L2. Both of the latter two results are new contributions to the literature. To the extent that second language learners can be considered analogous to children learning their first language (see Arndt & Beato, 2017), these data are consistent with studies showing false memory is greater for high- than low-BAS lists across development (e.g., Howe, Wimmer, Gagnon, et al., 2009). At a theoretical level, these results support the predictions of associative activation theories, both in general and in terms of their explanation for how false memory changes in an experience-dependent manner.

A third unique contribution of this study to the literature on false memory in L1 vs. L2 is that we measured L2 word knowledge. These data showed that participants did not know the translation of all L2 words that were used as stimuli. Nevertheless, covarying out the influence of word knowledge in L2 did not eliminate the effects of L1 vs. L2 on false memory, and indeed, largely did not alter the statistical magnitude of those effects, as judged by effect sizes. Thus, while L2 word knowledge is a plausible explanation for some or all of the effects of language proficiency on false memory, the data from Experiment 1 did not support this conclusion, further strengthening the support Experiment 1 lends to associative activation theories' explanation of false memory and how false memory changes with linguistic experience.

Experiment 2

In Experiment 2, we sought to build on the results from Experiment 1 by exploring how L2 proficiency affects false recognition in both high- and low-BAS lists. Thus, we studied false recognition in three groups of native Spanish speakers that were learning English as second language: those that were highly proficient in English, those that had an intermediate level of proficiency, and those that had a low level of English proficiency. One previous study has examined the effects of language proficiency on false recognition in L2 and found that increased language proficiency was associated with increased false memory (Arndt & Beato, 2017). As with previous research examining L1 vs. L2 differences in false memory, our prior work utilized study lists primarily composed of critical lures' high-BAS associates. Experiment 2 will examine if our previous findings generalize to the encoding of critical lures' low-BAS associates. In addition, we again measured participants' knowledge of studied words and critical lures in L2, and used those data to evaluate the extent to which false recognition differences across language proficiency levels were explicable by L2 word knowledge, as well as the extent to which high- vs. low-BAS differences in L2 false recognition can be explained by L2 word knowledge.

Although no previous studies have analyzed the effect of langauge proficiency on false memory with high- and low-BAS DRM lists, developmental studies of the effects of BAS on false memory may offer clues about how L2 proficiency will moderate the effects of BAS on false memory since both development and L2 proficiency involve changes in knowledge of words and word meaning in a language. Some studies of development have found that false recall increased across early life development (i.e., language proficiency) and that high-BAS lists produced greater false recall than low-BAS lists for all age groups (Howe, Wimmer, & Blease, 2009; Howe, Wimmer, Gagnon, et al., 2009). However, other studies have shown that associative strength differences in false memory emerge later in development, such that they are not evident in young children (Brainerd et al., 2002, 2006). Thus, the developmental literature does not offer a clear expectation regarding how BAS should impact false memory as proficiency with a language improves.

In addition, it is an open question how much prior results showing increases in false memory with language proficiency in L2 (Arndt & Beato, 2017) are due to changes in the extent to which study items activate semantic memory, as is expected by associative activation theories, or are due to differences in L2 word knowledge. The results from Experiment 1 suggest that it is relatively common for participants to have incomplete knowledge of words in L2, which leaves open the possibility that word knowledge may, in part or in whole, mediate false memory differences across language proficiency in L2. Although the results of Experiment 1 did not support the conclusion that L1 vs. L2 differences in false recognition are explicable based upon word knowledge, it is important to ensure that the results of Experiment 1 replicate, in addition to evaluating if L2 word knowledge differences mediate some or all of the L2 proficiency differences in false recognition found in prior work (Arndt & Beato, 2017). We investigated these questions by replicating Experiment 1 in the aforementioned groups: high-, medium-, and low-proficiency L2 speakers. In summary, the purpose of Experiment 2 was to (1) replicate the results of Experiment 1, (2) examine whether L2 proficiency affected false recognition following the study of both high- and low-BAS lists, and (3) evaluate the extent to which differences in false memory across L1 vs. L2, BAS effects in L2, and L2 proficiency were explained by L2 word knowledge.

Table 4 Mean proportion of L2 knowledge for high-BAS and low-BAS studied items and for critical lures as a function of L2 proficiency group in Experiment 2

L2 proficiency group	Studied		Critical lures
	High-BAS	Low-BAS	
Low proficiency	0.47 [0.44, 0.51]	0.69 [0.66, 0.71]	0.88 [0.85, 0.90]
Intermediate proficiency	0.57 [0.54, 0.60]	0.77 [0.74, 0.79]	0.93 [0.91, 0.95]
High proficiency	0.69 [0.66, 0.73]	0.84 [0.82, 0.87]	0.95 [0.92, 0.97]

95% CIs are reported in square brackets

Method

Participants

One hundred and sixty-four students who were native Spanish speakers (60% female) participated voluntarily and signed a consent form. All participants were learning English as a second language in the Official School of Languages in Salamanca (Spain). Forty-seven participants were studying English at an advanced level (high L2 proficiency group, M age = 28.72, SD = 11.36), 59 participants were studying English at an intermediate English level (intermediate L2 proficiency group, M age = 28.98, SD = 10.45), and 58 participants were studying English at a low English level (low L2 proficiency group, M age = 31.19, SD = 11.60).[4] Participants' ages ranged from 18 to 55 years (M age = 29.69, SD = 11.12) and there were no significant differences among the mean ages of the three L2 proficiency groups, $F(2, 161) = 0.82$, $p = 0.441$, $\eta^2_p = 0.010$. In addition to their objective L2 proficiency classification based upon proficiency exam performance, participants' self-assessed English proficiency (1 = elementary knowledge; 10 = native speaker proficiency) significantly differed between the groups, $F(2, 161) = 39.07$, $p < 0.001$, $\eta^2_p = 0.327$. Bonferroni post-hoc analyses indicated that low L2 proficiency group judged their proficiency ($M = 5.36$, SD = 1.15) lower than intermediate L2 proficiency group ($M = 6.54$, SD = 0.77), 95% CI [− 1.60, − 0.76], and high L2 proficiency group ($M = 6.89$, SD = 0.87), 95% CI [− 1.98, − 1.08] (all ps < 0.001). The mean difference in self-assessed proficiency between the intermediate L2 proficiency group and the high L2 proficiency group was not significant, $p = 0.179$, 95% CI [− 0.80, 0.10], although the qualitative pattern in the means is in the direction of the high L2 proficiency group having higher self-assessed L2 proficiency.

[4] The participants were in a proficiency group because they have passed an English test. In this test, both reading and listening comprehension, and written and oral expression were evaluated. Thus, high proficiency participants had passed both the low and intermediate level exams, while intermediate proficiency participants had passed the low-level exam. Low proficiency participants had yet to pass the low-level proficiency exam.

Materials and procedure

The materials and procedure were identical to those used in the Experiment 1 with the exception that individual computers were not used to present the study lists for this experiment. Instead, study lists were presented visually on a projection screen. The order of study lists was randomly determined and there were six different versions to ensure that stimulus lists were rotated through the high-BAS, low-BAS, and unstudied conditions across participants. Furthermore, participants responded to the recognition test on a sheet of paper. Test items were randomly ordered, with fifteen different versions of the recognition test.

Results and discussion

Translation test: Did the L2 proficiency groups know the English words used as stimuli?

The mean translation scores as a function of BAS and L2 proficiency for studied items, and for critical lures collapsed across BAS (see Footnote 3) are presented in Table 4. We used a 2 (BAS: high vs. low) × 3 (L2 proficiency: high, intermediate, low) repeated-measures ANOVA to examine the proportion of studied words translated correctly. There was a significant main effect of BAS, $F(1, 161) = 542.64$, MSE = 0.005, $p < 0.001$, $\eta^2_p = 0.771$, a significant main effect of L2 proficiency, $F(2, 161) = 51.60$, MSE = 0.017, $p < 0.001$, $\eta^2_p = 0.391$, and a BAS × L2 proficiency level interaction, $F(2, 161) = 4.62$, MSE = 0.005, $p = 0.011$, $\eta^2_p = 0.054$. The main effect of BAS indicated that participants knew more studied words in low-BAS lists ($M = 0.77$) than in high-BAS lists ($M = 0.58$), $p < 0.001$, 95% CI [0.17, 0.20]. The main effect of L2 proficiency indicated that the high L2 proficiency group ($M = 0.77$) knew more studied words than the intermediate L2 proficiency group ($M = 0.67$), $p < 0.001$, 95% CI [0.05, 0.14], and the low L2 proficiency group ($M = 0.58$), $p < 0.001$, 95% CI [0.14, 0.23]. Furthermore, the intermediate L2 proficiency group knew more studied words than the low L2 proficiency group, $p < 0.001$, 95% CI [0.05, 0.13]. Finally, the interaction between BAS and L2 proficiency indicated that, although L2 proficiency groups differed in word knowledge for both high- and low-BAS study

Table 5 Mean proportion of false recognition as a function of language, BAS, and L2 proficiency group in Experiment 2

L2 proficiency group	L1 words (Spanish)		L2 words (English)	
	High-BAS	Low-BAS	High-BAS	Low-BAS
Low proficiency	0.39 [0.33, 0.44]	0.21 [0.16, 0.26]	0.19 [0.14, 0.23]	0.12 [0.08, 0.15]
Intermediate proficiency	0.40 [0.34, 0.45]	0.16 [0.11, 0.20]	0.17 [0.13, 0.22]	0.12 [0.08, 0.16]
High proficiency	0.33 [0.27, 0.40]	0.19 [0.14, 0.24]	0.21 [0.16, 0.26]	0.19 [0.14, 0.23]

95% CIs are reported in square brackets

lists (all ps < 0.001), the differences across L2 proficiency groups was greater for high-BAS study lists.

We also evaluated if participants' knowledge of critical lures in L2 varied across the L2 proficiency groups. This analysis revealed a significant effect of L2 proficiency, $F(2, 161) = 11.64$, $p < 0.001$, $\eta^2_p = 0.126$. The low L2 proficiency group ($M = 0.88$, SD = 0.09) knew fewer critical words than the intermediate L2 proficiency group ($M = 0.93$, SD = 0.07), $p = 0.002$, 95% CI [− 0.09, − 0.02], and the high L2 proficiency group ($M = 0.95$, SD = 0.07), $p < 0.001$, 95% CI [− 0.11, − 0.03]. However, there was no difference between intermediate and high L2 proficiency groups in critical lure word knowledge, $p = 0.586$, 95% CI [− 0.06, 0.02].

False memory effect

The mean proportions of false recognition as a function of language (L1 vs. L2), BAS (high vs. low), and L2 proficiency (high vs. intermediate vs. low) are presented in Table 5. We evaluated the extent to which L2 word knowledge explained critical lure false memory with the same strategy used to analyze Experiment 1's data. Specifically, we computed composite measures of L2 study word knowledge and critical lure knowledge, as well as seaparate measures of L2 study word knowledge for high- and low-BAS items. Further, L2 knowledge scores were centered by subtracting each participants' translation score from the overall mean of the same type of translation score (e.g., composite study word knowledge scores were subtracted from the overall mean study word knowledge score). Finally, 2 (language: L1 vs. L2) × 2 (BAS: high vs. low) × 3 (L2 proficiency: high vs. intermediate vs. low) ANCOVAs were run using each covariate to examine how covarying out the effects of L2 knowledge affected (1) the false recognition effects found in the full design of this experiment, (2) the effect of language on false recognition for low- and high-BAS lists, and (3) the effects of L2 proficiency on false recognition for low- and high-BAS lists.

In the first set of analyses, none of the covariates explained significant variance in false memory [largest $F(2, 160) = 0.498$, MSE = 0.064, $p = 0.609$, $\eta^2_p = 0.006$ for low-BAS study item L2 knowledge]. Further, the results found in a 2 (language) × 2 (BAS) × 3 (language proficiency) ANOVA were all replicated in the ANCOVA analysis, with one exception.[5] Specifically, the ANOVA and ANCOVAs produced a significant main effect of language, $F(1, 161) = 74.90$, MSE = 0.028, $p < 0.001$, $\eta^2_p = 0.318$ and a significant main effect of BAS, $F(1, 161) = 95.315$, MSE = 0.023, $p < 0.001$, $\eta^2_p = 0.372$, but no significant main effect of L2 proficiency, $F(2, 161) = 0.31$, MSE = 0.064, $p = 0.738$, $\eta^2_p = 0.004$. The main effect of language indicated that false recognition was higher in L1 ($M = 0.28$) than in L2 ($M = 0.16$), and the main effect of BAS indicated that false recognition was higher following the study of high-BAS lists ($M = 0.28$) than low-BAS lists ($M = 0.16$).

Both the ANOVA and ANCOVAs produced a significant language × BAS interaction, $F(1, 161) = 33.67$, MSE = 0.023, $p < 0.001$, $\eta^2_p = 0.173$, but no significant BAS × L2 proficiency level interaction, $F(2, 161) = 2.43$, MSE = 0.023, $p = 0.091$, $\eta^2_p = 0.029$, nor a language × BAS × L2 proficiency interaction, $F(2, 161) = 0.99$, MSE = 0.023, $p = 0.374$, $\eta^2_p = 0.012$. The language × BAS interaction occurred because BAS had larger effects in L1 than in L2, the same pattern found in Experiment 1. Further, and again replicating Experiment 1, the effects of BAS were significant in both L1, $t(163) = 10.41$, $p < 0.001$, and L2, $t(163) = 3.31$, $p = 0.001$, and the effects of language on false recognition were significant for both high-BAS lists, $t(163) = 9.30$, $p < 0.001$, and low-BAS lists, $t(163) = 3.23$, $p = 0.002$. Finally, while the ANOVA produced a language × L2 proficiency level interaction, $F(2, 161) = 3.50$, MSE = 0.028, $p = 0.033$, $\eta^2_p = 0.042$, this interaction was not significant in any of the ANCOVA analyses.

In the second set of analyses, we examined the simple effect of language on false recognition in the high-BAS condition using high-BAS study item word knowledge as a covariate, and in the low-BAS condition using low-BAS study item word knowledge as a covariate. These analyses tested if word knowledge was able to explain the L1 vs. L2 differences in false recognition in either the high- or low-BAS conditions that we reported above. For false recognition in the high-BAS condition, the effect of language remained significant, $F(1, 162) = 11.16$, MSE = 0.033, $p = 0.001$, $\eta^2_p = 0.064$, and high-BAS study item knowledge was not

[5] F ratios, mean squared error, p values and η^2_p are reported from the ANOVA-based analysis of the false recognition data.

Table 6 Mean proportion of true recognition as a function of language, BAS, and L2 proficiency group in Experiment 2

L2 proficiency group	L1 words (Spanish)		L2 words (English)	
	High-BAS	Low-BAS	High-BAS	Low-BAS
Low proficiency	0.73 [0.68, 0.78]	0.69 [0.65, 0.74]	0.77 [0.73, 0.81]	0.75 [0.71, 0.79]
Intermediate proficiency	0.75 [0.70, 0.79]	0.65 [0.60, 0.70]	0.79 [0.75, 0.83]	0.75 [0.71, 0.79]
High proficiency	0.77 [0.71, 0.82]	0.67 [0.62, 0.73]	0.82 [0.77, 0.86]	0.75 [0.70, 0.79]

95% CIs are reported in square brackets

a significant covariate, $F(1, 162) < 1$. Similarly, in the low-BAS condition, the effect of language remained significant, $F(1, 162) = 5.48$, MSE = 0.020, $p = 0.020$, $\eta^2_p = 0.033$, and low-BAS study item knowledge was not a significant covariate, $F(1, 162) < 1$. These results are further supported by the fact that high-BAS L2 word knowledge did not correlate with high-BAS L2 false recognition, $r(162) = 0.102$, $p = 0.195$, and low-BAS L2 word knowledge did not correlate with low-BAS L2 false recognition, $r(162) = 0.058$, $p = 0.464$. Thus, the false recognition differences across language for both low- and high-BAS lists were not due to L2 word knowledge.

Third, although there were not reliable interactions between L2 proficiency and the other independent variables in this experiment, we sought to evaluate whether there were L2 differences in L2 false recognition directly. In addition, we sought to evaluate if L2 knowledge differences explain L2 proficiency group differences in false recognition. We conducted two ANCOVAs: one that evaluated the effect of L2 proficiency on high-BAS list false recognition in L2, and one that evaluated the effect of L2 proficiency on low-BAS list false recognition in L2, and compared those results to ANOVAs evaluating whether L2 proficiency had effects on false memory. For the ANCOVA analyses, high-BAS study item knowledge in L2 was used as a covariate for the high-BAS analyses and low-BAS study item knowledge in L2 was used as a covariate for the low-BAS analyses. These analyses showed that neither of the covariates explained a significant amount of the variance in false memory [largest $F(1, 160) = 1.42$, $p = 0.235$ for high-BAS L2 word knowledge]. In the analyses evaluating the effect of L2 proficiency on false memory in L2, there were no reliable differences in false recognition across L2 proficiency groups for high-BAS lists in either the ANOVA or ANCOVA, $F(2, 161) = 0.451$, MSE = 0.031, $p = 0.638$, η^2_p = 0.006. Thus, although there was a tendency in the means for the high L2 proficiency group to produce greater false memory ($M = 0.21$) than the other two groups ($M = 0.17$ and $M = 0.19$ for the intermediate and low L2 proficiency groups, respectively), that mean difference was not statistically reliable. However, there was an L2 proficiency group difference in false recognition for low-BAS lists in both the ANOVA and ANCOVA, $F(2, 161) = 3.82$, MSE = 0.023, $p = 0.024$, η^2_p = 0.045. Post-hoc Bonferroni comparisons indicated that the high L2 proficiency group ($M = 0.19$) falsely recognized more words than the intermediate L2 proficiency group ($M = 0.12$, $p = 0.049$, 95% CI [0.001, 0.14]) and the low L2 proficiency group ($M = 0.12$, $p = 0.046$, 95% CI [0.001, 0.15]).

The final way we evaluated the viability of word knowledge to explain variability in false recognition was by correlating each measure of word knowledge with false recogniton in L2. In these analyses, the only covariate that was correlated with false recognition in L2 was high BAS study item knowledge, which was weakly correlated, $r(162) = 0.182$, $p = 0.02$, with L2 low BAS lure false recognition. This correlation is somewhat surprising given that the measures that showed a correlation between L2 knowledege and L2 false recognition were at different BAS levels and thus should not be directly related to a greater degree than word knowledge and false recognition at the same level of BAS. As with Experiment 1, the general lack of correlations between L2 false recognition and L2 word knowledge were not because either variable had properties (e.g., restricted range) that undermined finding reliable correlations. High- and low-BAS false recognition in L2 was positively correlated, $r(162) = 0.291$, $p < 0.001$, and all of the measures of L2 word knowledge were positively correlated with each other [smallest $r(162) = 0.419$, $p < 0.001$ for the correlation between critical lure word knowledge and low-BAS study item knowledge]. Given that the covariates did not explain significant variation in false recognition based upon the ANCOVA results and the correlations between the covariates and false recognition in L2 were largely non-significant, it seems inappropriate to interpret the change in the interaction between language and L2 proficiency in the ANCOVAs relative to the ANOVA as evidence that L2 word knowledge explains the false recognition effects we observed in this experiment.

True recognition

Mean proportions of true recognition are presented in Table 6. We analyzed true recognition (hit rates) using a 2 (language) × 2 (BAS) × 3 (L2 proficiency) ANOVA, with language and BAS as within-subjects factors and L2 proficiency as a between-subjects factor. This analysis revealed reliable main effects of language, $F(1, 161) = 25.80$, MSE = 0.023, $p < 0.001$, $\eta^2_p = 0.138$ and BAS, $F(1, 161) = 35.03$, MSE = 0.016,

$p < 0.001$, $\eta^2_p = 0.179$. The main effect of language showed that hit rates were greater when items were studied in L2 ($M = 0.77$) than in L1 ($M = 0.71$). The main effect of BAS showed that hit rates were greater for high-BAS study items ($M = 0.77$) than low-BAS study items ($M = 0.71$). Both of these main effects replicated true recognition findings in Experiment 1. None of the two- or three-way interactions were statistically reliable.

Baseline false alarm rates

Finally, we analyzed the baseline false alarm rates for unrelated distractors and for unrelated critical distractors using 2 (language) × 2 (BAS) × 3 (L2 proficiency) mixed-model ANOVAs. The analysis of unrelated distractor false alarm rates produced a main effect of language, $F(1, 161) = 25.39$, $MSE = 0.025$, $p < 0.001$, $\eta^2_p = 0.136$, documenting that false alarm rates were higher in L2 than in L1. No other main effects or interactions were significant. The analysis of unrelated critical distractors produced main effects of language, $F(1, 161) = 4.18$, $MSE = 0.038$, $p = 0.043$, $\eta^2_p = 0.025$, and BAS, $F(1, 161) = 25.82$, $MSE = 0.036$, $p < 0.001$, $\eta^2_p = 0.138$. The main effect of language occurred because unrelated critical distractor false alarms were greater in L2 ($M = 0.10$) than in L1 ($M = 0.07$), and the main effect of BAS occurred because false alarms were higher for high-BAS unrelated critical distractors ($M = 0.12$) than low-BAS unrelated critical distractors ($M = 0.05$). No other main effects or interactions were reliable.

The results of this experiment replicated those from Experiment 1 and further support associative activation theories' explanation of false recognition, both in general and in terms of their explanation of experience-based changes in false recognition. This support stems from the findings that (1) BAS impacted false recognition in both L1 and L2, (2) L2 proficiency increased false recognition following the study of low-BAS lists, and (3) knowledge of words in L2 did not explain the effects of L1 vs. L2 or L2 proficiency on false recognition. Thus, these results are most readily explained by the claim, derived from associative activation theories of false memory, that experience impacts the extent to which semantic memory, and thus critical lure representations, are activated by studying words. In particular, greater activation should occur when people have greater experience with a language, such as is the case for studying words in L1 or when they have greater proficiency in L2.

General discussion

The results of these experiments replicated findings from previous research showing that false memory was greater in L1 compared to L2 (Anastasi et al, 2005; Arndt & Beato, 2017; Howe et al, 2008; Sahlin et al, 2005), and that false memory increased with L2 proficiency (Arndt & Beato, 2017). Furthermore, the results of these experiments extended these findings in four important ways. First, greater false recognition in L1 than L2 occurred for both high-BAS lists, which are similar to those that have been used in prior research (Anastasi et al., 2005; Arndt & Beato, 2017; Howe, Wimmer, Gagnon, et al., 2009; Sahlin et al., 2005), and low-BAS lists. Second, BAS increased false recognition in both L1, as has been found in much prior work (Arndt, 2012a, 2015; Arndt & Gould, 2006; Gallo & Roediger, 2003; Howe, Wimmer, Gagnon, et al., 2009; McEvoy et al., 1999; Roediger et al., 2001), and in L2. Third, the increase in false recognition as L2 proficiency increased occurred for lists composed of low-BAS associates, which extends prior findings showing increased false recognition as L2 proficiency increased when lists composed of high-BAS associates were studied (Arndt & Beato, 2017). Fourth, we used L2 to L1 translation data to evaluate the extent to which participants possessed knowledge of stimulus words in L2 (English). As reviewed above, an implicit assumption in most prior work examining language proficiency effects on false memory is that participants have full knowledge of the words they are studying in L2. This assumption was not supported by the present data, where even participants that were highly proficient in L2 did not know all of the words they studied in L2. Despite participants having less than perfect knowledge of stimuli in L2, lower word knowledge in L2 was unable to explain the effects of L1 vs. L2 or L2 proficiency on false recognition.

Taken together, these results are consistent with the view that increases in false recognition with linguistic experience are driven by changes in the degree to which lexical representations activate conceptual representations in semantic memory, which in turn activate representations of related, but unstudied, critical lures (Howe, Wimmer, Gagnon, et al., 2009). Four facets of the present results favor this view. First, studying high-BAS lists produced greater false recognition than studying low-BAS lists, and this effect occurred in both L1 and L2. Second, L1 false recognition was greater than L2 false recognition for both study lists composed of high-BAS items and study lists composed of low-BAS items. Third, differences in false recognition as L2 proficiency increased occurred for study lists composed of low-BAS associates, and showed a mean tendency to be greater for highly proficient L2 participants when study lists were composed of high-BAS associates. Fourth, analyses that accounted for participants' L2 word knowledge did not impact the effects of L1 vs. L2, BAS, or L2 proficiency on false recognition. Next, we highlight how these outcomes, individually and collectively, support the view that false recognition results from how extensively study items activate semantic memory representations, which is the core proposal of associative activation theories' explanation of why false memory increases with linguistic experience.

A core assumption of associative activation theories of false memory is that the more an unstudied concept (e.g., a critical lure) is activated during encoding, the more likely it is to be falsely remembered later (Howe, Wimmer, Gagnon, et al., 2009; Roediger et al., 2001). A commonly employed test of this claim has been to manipulate the extent to which study items generate critical lures in free association (i.e., BAS). In the view of associative activation theories, high-BAS associates activate critical lures' representations to a greater extent than low-BAS associates, and thus produce greater levels of false memory. This effect has been found consistently in the literature (e.g., Arndt & Gould, 2006; Gallo & Roediger, 2003; Howe, Wimmer, Gagnon, et al., 2009; McEvoy et al., 1999; Roediger et al., 2001) and is a key finding that supports associative activation theories' view that false memory stems from how extensively study items activate critical lures' representations in semantic memory. Both studies we conducted further tested this explanation by examining if BAS impacts false recognition when lists are studied in L2. The significance of examining the effects of BAS in L2 can be found in theories of bilingual representation, which generally suggest that lexical terms in L1 and L2 access common semantic memory representations (Dijkstra & Van Heuven, 2002; Francis, 1999; Green, 1998; Kroll & Stewart, 1994; Li et al., 2004). Thus, because associative activation theories of false memory claim that the effects of BAS are based in how semantic memory is structured, they should be observed when study lists are presented in both L1 and L2, precisely the outcome observed in these experiments.

Associative activation theories explain why false memory is greater in L1 than in L2 by arguing that semantic memory is more readily and automatically accessible when information is studied in L1. This claim stems from the view that lexical terms and their conceptual representations in semantic memory are more strongly associated in L1 compared to L2, which is a core feature of bilingual language representation models (Dijkstra & Van Heuven, 2002; Green, 1998; Kroll & Stewart, 1994; Li et al., 2004). A necessary consequence of this explanation is that any set of study items, as long as they activate critical lures' representations in semantic memory, should produce greater false recognition in L1 compared to L2. This is exactly the result found in both of these studies—L1 false memory was greater than L2 false memory, both for study lists composed of high-BAS associates and for those composed of low-BAS associates. Importantly, as in our previous work (Arndt & Beato, 2017), we took extensive steps to ensure that lists in L1 (Spanish) had the same levels of BAS as lists in L2 (English), using norms of free association (see Table 1). Thus, the most likely reason that false recognition differed between L1 and L2 is that study lists presented in L1 activated semantic memory, and thus critical lures' representations, to a greater degree than study lists presented in L2.

This same basic theoretical formulation can be used to explain the effects of L2 proficiency on false recognition. In particular, the more proficient a person is in their L2, the more completely and automatically lexical terms in that language (i.e., studies items) will activate their conceptual representations in semantic memory. In turn, the more fully and automatically study items' representation are activated, the more activation will spread to critical lures' representations, thereby increasing false memory. This explanation is supported by the finding that false recognition in L2 increased with language proficiency for low-BAS study lists and showed a qualitative, but non-significant, tendency to increase with L2 proficiency for high-BAS study lists.

Finally, and critical to supporting associative activation theories' view of how L1 vs. L2 and L2 proficiency impact false memory, participants' knowledge of words in L2 failed to explain the effects of L1 vs. L2, L2 proficiency, and BAS effects on false memory in these studies. The inability of word knowledge to explain the false recognition effects observed in these studies occurred despite the fact that L2 word knowledge offers a cogent and plausible interpretation of these effects. Specifically, knowledge of the stimuli in participants' L2 was far from perfect and increased with L2 proficiency, making L2 word knowledge a potential explanation of L1 vs. L2 and L2 proficiency effects on false memory. Despite this, ANCOVA analyses and correlational analyses failed to support a role for word knowledge contributing to the false memory effects observed in these studies. This outcome further supports the view that false memory is based upon the characteristics of representations in semantic memory, as is claimed by associative activation theory (Howe, Wimmer, Gagnon, et al., 2009; Roediger et al., 2001). Further, this outcome contradicts the view that changes in false memory with experience can be understood based upon participants' linguistic knowledge (Carneiro & Fernandez, 2010).

Beyond the theoretical implications of these studies, the present results underscore the importance of evaluating assumptions underlying empirical tests of theory, particularly in situations where participant-based variables are being examined. In the case of the prior studies that investigated of how language proficiency impacted false memory, all implicitly assumed that stimuli were equally understood in participants' L1 and L2 (Anastasi et al., 2005; Arndt & Beato, 2017; Howe et al., 2008; Sahlin et al., 2005). In particular, no prior studies evaluated the extent to which participants knew the meaning of stimuli in their L2, despite that knowledge providing a plausible explanation L1 vs. L2 differences as well as L2 proficiency differences. As documented in the present studies, participants' knoweldge of stimuli in L2 was far from perfect, in addition to varying with L2 proficiency. Thus, future studies that examine such effects with the goal of understanding the effects of L1 vs. L2 and L2 proficiency on memory should include a measure of L2 word knowledge to evaluate the validity of alternative explanations for memory effects, such as the hypothesis that L2 word knowledge differences across groups or across languages can explain some or all language proficiency effects on memory. Collecting such data will either validate the assumption that stimulus knowledge does not differ across groups or participants' dominant/non-dominant languages or will provide key data that can then be used to evaluate the extent to which word knowledge differences can explain some or all of the memory effects observed in those studies.

While the focus of these studies was on false recognition, it is notable that the true recognition data in both experiments showed higher hit rates in L2 than in L1, replicating prior work with unrelated words as stimuli (Francis & Gutiérrez, 2012; Francis & Strobach, 2013). However, it is important to note that the baseline false alarm rates in our studies did not replicate prior work, where false alarms were lower in L2 than L1, producing a "mirror effect" (Glanzer, Hilford, & Kim, 2004) of L2 hits and false alarms compared to L1 (Francis & Gutiérrez, 2012; Francis & Strobach, 2013). It is possible that the nature of the stimuli used in the present studies, with inter-related study items and unstudied words, as well as related critical lures, produced differences in the pattern of false alarms found between L1 and L2 in prior work. Future studies should investigate the extent to which these different patterns occur because of stimulus material differences (unrelated study and test lists vs. related study and test lists) or other procedural differences between the present studies and previous work.

Potential limitations and alternate explanations

Although we designed these studies to evaluate one potential alternate explanation for the effects of language dominance and L2 proficiency on false memory (L2 word knowledge), it is important to recognize that there may be other alternate explanations of our results that were not directly evaluated. One potential concern is that the stimuli used in L1 and L2 were not the same words, which could complicate interpreting these results as a function of language dominance (L1 vs. L2). While we are unable to rule out this factor as a potential limitation of the present studies, our prior work (Arndt & Beato, 2017) has documented that lists that were constructed in a similar fashion to those used in these studies, where BAS and FAS were rigorously controlled across L1 and L2, showed the same effects when participants were native English speakers (L1) learning Spanish (L2) and when participants were native Spanish speakers (L1) learning English (L2). Thus, prior work that has evaulated this possibility has shown that the same L1 vs. L2 effects occur for the same stimuli regardless of participants' L1 and L2. This outcome suggests that it is language dominance per se, and not a unique characteristic of the stimuli or native Spanish speakers, that is driving the L1 vs. L2 effects documented in the present studies.

A second potential concern is that we used relatively short lists of four associates for each critical lure, which tends to produce a less robust false memory illusion compared to DRM lists composed of larger numbers of associates (Arndt, 2010; Robinson & Roediger, 1997). Despite the tendency for shorter lists to produce less robust DRM false memory, it is important to note that with lists of four or fewer associates, several basic characteristics of critical lure false memories occur. For example, participants show greater false memory following the study of high-BAS compared to low-BAS associates (Arndt, 2012a, 2015; Arndt & Gould, 2006), a finding that was replicated in both of the studies reported here. Similarly, people show a bias to judge critical lures as coming from the source used to present its associates, even when as few as two associates were studied (Arndt, 2010). Finally, increasing retrieval time, which should increase the use of recollective information (e.g., about items' encoding source or format), increases false alarms to critical lures when four associates were studied (Arndt, 2012a). This latter finding has been interpreted as supporting the view that at least some of the false memory observed in the DRM paradigm is underlain by false recollection, and not simply enhanced familiarity, of criticial lures (Arndt, 2012a, b). Thus, when research has examined parallels between DRM lists using a large number of associates and lists with smaller numbers of associates, there are a number of key parallels in the empirical effects, which suggest the core psychological phenomenon being studied arises from the same mechanisms.

Conclusion

In summary, these studies replicated prior results showing that false memory was greater in L1 than L2 (Anastasi et al., 2005; Arndt & Beato, 2017; Howe et al., 2008; Sahlin et al., 2005), and that false memory in L2 increased with L2 proficiency (Arndt & Beato, 2017). In addition, these studies added to these prior findings in two key ways. First, the effects of L1 vs. L2 and L2 proficiency on false memory generalized to the study of low-BAS lists while all prior studies have examined lists composed of high-BAS associates of critical lures. Second, we measured participants' knowledge of terms in their L2 to evaluate how well this knoweldge could explain some or all of the effects of L1 vs. L2, L2 proficiency, and BAS on false memory in L2. As reviewed above, knowledge of stimuli in L2 did not explain the impact of these variables on false memory. Taken together, these outcomes favor the explanation offered by associative activation theories of false memory, particularly their application to explaining the effects of linguistic experience on false memory, over views that highlight the role that differences in knowledge of stimuli explains the effects of linguistic experience on false memory.

Author contributions All authors contributed equally to this work.

Funding This research was supported by grant SA052G18, awarded to Maria Soledad Beato by Junta de Castilla y León.

Compliance with ethical standards

Conflict of interest The authors declare they have no competing interests in conducting this work.

Ethical approval The research protocol was approved by the Bioethics Committee at the University of Salamanca and has been performed in accordance with the ethical standards laid down in the 1964 Declaration of Helsinki and its later amendments. Informed consent was obtained from all individual participants included in the study. The authors declare that they have no conflict of interest.

Appendix

The DRM lists constructed as stimuli for these studies. BAS list strengths are reported for each list.

Spanish lists

Critical lure	Associated words	BAS	BAS condition
Corazón (heart)	Latido, miocardio, bombear, órgano (beat, myocardium, pump, organ)	0.475	High
Corazón (heart)	Arteria, amor, sentimiento, roto (artery, love, feeling, broken)	0.072	Low
Árbol (tree)	Rama, tronco, abeto, raíz (branch, log, fir, root)	0.483	High
Árbol (tree)	Leña, naturaleza, monte, pájaro (firewood, nature, hill, bird)	0.062	Low
Coche (car)	Volante, maletero, rueda, conducir (steering-wheel, trunk, wheel, drive)	0.533	High
Coche (car)	Moto, rápido, nuevo, pinchazo (motorcycle, fast, new, puncture)	0.067	Low
Basura (trash)	Desperdicios, vertedero, papelera, porquería (scraps, garbage-dump, basket, garbage)	0.420	High
Basura (trash)	Contaminar, mierda, suciedad, desagradable (contaminate, shit, dirt, unpleasant)	0.045	Low

Critical lure	Associated words	BAS	BAS condition
Cárcel (jail)	Rejas, prisionero, prisión, celda (bars, prisoner, prison, cell)	0.452	High
Cárcel (jail)	Culpable, ladrón, corrupto, libertad (guilty, thief, corrupt, freedom)	0.057	Low
Cura (pastor)	Sotana, reverendo, sacerdote, párroco (soutane, reverend, priest, parish-priest)	0.430	High
Cura (pastor)	Diócesis, papa, religión, capilla (diocese, pope, religion, chapel)	0.067	Low
Pelo (hair)	Peine, coleta, melena, patillas (comb, ponytail, mane, sideburns)	0.457	High
Pelo (hair)	Bigote, depilación, largo, corto (mustache, depilation, long, short)	0.070	Low
Luz (light)	Bombilla, lámpara, claridad, vela (bulb, lamp, clarity, candle)	0.492	High
Luz (light)	Día, sol, penumbra, cortina (day, sun, semidarkness, curtain)	0.072	Low
Tonto (silly)	Bobo, imbécil, estúpido, idiota (fool, imbecile, stupid, idiot)	0.456	High
Tonto (silly)	Inocente, insulto, antipático, grosero (gullible, insult, unfriendly, gross)	0.053	Low
Perro (dog)	Ladrar, ladrido, gato, hueso (bark, barking, cat, bone)	0.497	High
Perro (dog)	Caza, fiel, collar, colmillo (hunting, faithful, collar, fang)	0.048	Low
Libro (book)	Capítulo, lectura, prólogo, manual (chapter, reading, prologue, manual)	0.439	High
Libro (book)	Interesante, hojas, cultura, letras (interesting, sheets, culture, letters)	0.060	Low
Muerte (death)	Tanatorio, pésame, sepultura, ataúd (morgue, condolence, burial, coffin)	0.468	High
Muerte (death)	Ceniza, pena, esqueleto, desgracia (ash, pity, skeleton, misfortune)	0.079	Low
Puerta (door)	Pomo, llave, cerrar, blindada (knob, key, close, armored)	0.465	High
Puerta (door)	Entrada, ventana, salida, salir (entrance, window, exit, leave)	0.057	Low
Rey (king)	Monarca, majestad, reina, príncipe (monarch, majesty, queen, prince)	0.474	High
Rey (king)	Infante, poder, reino, castillo (infante, power, kingdom, castle)	0.068	Low
Dinero (money)	Cheque, hucha, ahorro, ingreso (check, piggybank, savings, deposit)	0.482	High

Critical lure	Associated words	BAS	BAS condition
Dinero (money)	Cuenta, recibo, perder, comprar (bank-account, receipt, lose, buy)	0.063	Low
Teatro (theater)	Escenario, interpretación, representación, escena (stage, interpretation, performance, scene)	0.431	High
Teatro (theater)	Aplauso, público, actriz, butaca (applause, public, actress, armchair)	0.052	Low

The approximate English translation for Spanish critical lures and Spanish associated (studied) words are reported in parenthesis

English lists

Critical lure	Associated words	BAS	BAS condition
Board	Chalk, bulletin, dart, skate	0.483	High
Board	Stiff, surf, committee, game	0.055	Low
Bug	Beetle, insect, termite, pest	0.442	High
Bug	Irritate, ant, tick, annoy	0.077	Low
Color	Hue, crayon, pigment, maroon	0.472	High
Color	Favorite, bright, pink, yellow	0.063	Low
Ear	Lobe, q-tips, earring, hearing	0.470	High
Ear	Listen, eye, noise, wax	0.060	Low
Hot	Spicy, humid, chili, fever	0.435	High
Hot	Iron, sexy, flame, burn	0.078	Low
Hungry	Starving, famished, starve, thirsty	0.492	High
Hungry	Eating, stomach, sandwich, food	0.050	Low
Lie	Fib, perjury, untruthful, deception	0.481	High
Lie	Betray, honest, myth, false	0.079	Low
Number	Digit, serial, count, seven	0.475	High
Number	Name, hundred, two, one	0.074	Low
Party	Celebration, celebrate, festival, slumber	0.454	High
Party	Fun, banquet, dance, beer	0.073	Low
Pig	Hog, pork, sow, ham	0.456	High
Pig	Squeal, cop, policeman, mud	0.065	Low
Ring	Diamond, doorbell, bell, engagement	0.466	High
Ring	Phone, ruby, finger, marriage	0.067	Low
Same	Differ, identical, alike, similar	0.436	High
Same	Usual, twin, constant, exact	0.064	Low
Shoe	Sneaker, sock, shoelace, heel	0.483	High
Shoe	Galoshes, tie, salesman, fit	0.059	Low

Critical lure	Associated words	BAS	BAS condition
Story	Tale, fairytale, fable, plot	0.466	High
Story	Fiction, tell, paragraph, novel	0.065	Low
Street	Avenue, boulevard, main, sidewalk	0.467	High
Street	Parkway, cross, highway, block	0.056	Low
Wood	Lumber, splinter, timber, carpenter	0.461	High
Wood	Cabinet, ax, chop, paper	0.067	Low

References

Anastasi, J. S., & Rhodes, M. G. (2008). Examining differences in the levels of false memories in children and adults using child-normed lists. *Developmental Psychology, 44*(3), 889–894. https://doi.org/10.1037/0012-1649.44.3.889.

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. https://doi.org/10.1080/09658210444000421.

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. https://doi.org/10.1037/a0017394.

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

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

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

Arndt, J., & Beato, M. S. (2017). The role of language proficiency in producing false memories. *Journal of Memory and Language, 95,* 146–158. https://doi.org/10.1016/j.jml.2017.03.004.

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

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. https://doi.org/10.7334/psicothema2014.79.

Beato, M. S., & Cadavid, S. (2016). Normative study of theme identifiability: Instructions with and without explanation of the false memory effect. *Behavior Research Methods, 48*(4), 1252–1265. https://doi.org/10.3758/s13428-015-0652-6.

Brainerd, C. J., Forrest, T. J., Karibian, D., & Reyna, V. F. (2006). Development of the false-memory illusion. *Developmental Psychology, 42*(5), 962–979. https://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. https://doi.org/10.1111/1467-8624.00477.

Cabeza, R., & Lennartson, E. R. (2005). False memory across languages: Implicit associative response vs fuzzy trace views. *Memory, 13*(1), 1–5. https://doi.org/10.1080/09658210344000161.
Calado, B., Otgaar, H., & Muris, P. (2018). Are children better witnesses than adolescents? Developmental trends in different false memory paradigms. *Journal of Child Custody: Research, Issues, and Practices, 15*(4), 330–348. https://doi.org/10.1080/15379418.2019.1568948.
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. https://doi.org/10.1016/j.jecp.2009.09.004.
Cadavid, S., & Beato, M. S. (2016). Memory distortion and its avoidance: An event-related potentials study on false recognition and correct rejection. *PLoS ONE*, *11*(10), e0164024. https://doi.org/10.1371/journal.pone.0164024.
Deese, J. (1959). On the prediction of occurrence of particular verbal intrusions in immediate recall. *Journal of Experimental Psychology, 58*(1), 17–22. https://doi.org/10.1037/h0046671.
Delaney, H., & Maxwell, S. (1981). On using analysis of covariance in repeated measures designs. *Multivariate Behavioral Research, 16,* 105–128.
Dewhurst, S. A., & Robinson, C. A. (2004). False memories in children: Evidence for a shift from phonological to semantic associations. *Psychological Science, 15*(11), 782–786. https://doi.org/10.1111/j.0956-7976.2004.00756.x.
Dijkstra, T., & Van Heuven, W. J. B. (2002). The architecture of the bilingual word recognition system: From identification to decision. *Bilingualism: Language and Cognition, 5*(3), 175–197. https://doi.org/10.1017/S1366728902003012.
Fernández, A., Díez, E., Alonso, M. A. (2009). Normas de Asociación libre en castellano, online database [Free Association norms in Spanish]. http://campus.usal.es/gimc/nalc.
Francis, W. S. (1999). Cognitive integration of language and memory in bilinguals: Semantic representation. *Psychological Bulletin, 125*(2), 193–222. https://doi.org/10.1037/0033-2909.125.2.193.
Francis, W. S., & Gutiérrez, M. (2012). Bilingual recognition memory: Stronger performance but weaker levels-of-processing effects in the less fluent language. *Memory and Cognition, 40*(3), 496–503. https://doi.org/10.3758/s13421-011-0163-3.
Francis, W. S., & Strobach, E. N. (2013). The bilingual L2 advantage in recognition memory. *Psychonomic Bulletin and Review, 20*(6), 1296–1303. https://doi.org/10.3758/s13423-013-0427-y.
Gallo, D. A., & Roediger, H. L. (2003). The effects of associations and aging on illusory recollection. *Memory and Cognition, 31*(7), 1036–1044. https://doi.org/10.3758/BF03196124.
Glanzer, M., Hilford, A., & Kim, K. (2004). Six regularities of source recognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 30*(6), 1176–1195. https://doi.org/10.1037/0278-7393.30.6.1176.
Green, D. W. (1998). Mental control of the bilingual lexico-semantic system. *Bilingualism: Language and Cognition, 1*(2), 67–81. https://doi.org/10.1017/S1366728998000133.
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. https://doi.org/10.1111/j.1467-8624.2006.00922.x.
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. https://doi.org/10.1016/j.jml.2007.09.001.
Howe, M. L., Wimmer, M. C., & Blease, K. (2009). The role of associative strength in children's false memory illusions. *Memory, 17*(1), 8–16. https://doi.org/10.1080/09658210802438474.
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. https://doi.org/10.1016/j.jml.2008.10.002.
Huff, M. J., & Bodner, G. E. (2013). When does memory monitoring succeed versus fail? Comparing item-specific and relational encoding in the DRM paradigm. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 39*(4), 1246–1256. https://doi.org/10.1037/a0031338.
Knott, L. M., Howe, M. L., Wimmer, M. C., & Dewhurst, S. A. (2011). The development of automatic and controlled inhibitory retrieval processes in true and false recall. *Journal of Experimental Child Psychology, 109*(1), 91–108. https://doi.org/10.1016/j.jecp.2011.01.001.
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. https://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. https://doi.org/10.1016/j.neunet.2004.07.004.
Lim, L. C. L., & Goh, W. D. (2019). False recognition modality effects in short-term memory: Reversing the auditory advantage. *Cognition, 193,* 104008. https://doi.org/10.1016/j.cognition.2019.104008.
McEvoy, C. L., Nelson, D. L., & Komatsu, T. (1999). What is the connection between true and false memories? The differential roles of interitem associations in recall and recognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 25*(5), 1177. https://doi.org/10.1037/0278-7393.25.5.1177.
Metzger, R. L., Warren, A. R., Shelton, J. T., Price, J., Reed, A. W., & Williams, D. (2008). Do children "DRM" like adults? False memory production in children. *Developmental Psychology, 44*(1), 169–181. https://doi.org/10.1037/0012-1649.44.1.169.
Nelson, D. L., McEvoy, C. L., Schreiber, T. A. (1998). The University of South Florida word association, rhyme, and word fragment norms. http://w3.usf.edu/FreeAssociation/.
Otgaar, H., de Ruiter, C., Howe, M. L., Hoetmer, L., & van Reekum, P. (2017). A case concerning children's false memories of abuse: Recommendations regarding expert witness work. *Psychiatry, Psychology and Law, 24*(3), 365–378. https://doi.org/10.1080/13218719.2016.1230924.
Otgaar, H., Howe, M. L., Muris, P., & Merckelbach, H. (2019). Associative activation as a mechanism underlying false memory formation. *Clinical Psychological Science, 7*(2), 191–195. https://doi.org/10.1177/2167702618807189.
Robinson, K. J., & Roediger, H. L. (1997). Associative processes in false recall and false recognition. *Psychological Science, 8*(3), 231–237. https://doi.org/10.1111/j.1467-9280.1997.tb00417.x.
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. https://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 and Review, 8*(3), 385–407. https://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 and Cognition, 33*(8), 1414–1421. https://doi.org/10.3758/BF03193374.

Sugrue, K., & Hayne, H. (2006). False memories produced by children and adults in the DRM paradigm. *Applied Cognitive Psychology, 20*(5), 625–631. https://doi.org/10.1002/acp.1214.

Wang, J., Otgaar, H., Howe, M. L., Lippe, F., & Smeets, T. (2018). The nature and consequences of false memories for visual stimuli. *Journal of Memory and Language, 101,* 124–135. https://doi.org/10.1016/j.jml.2018.04.007.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.