

# New in, old out: Does learning a new language make you forget previously learned foreign languages?



Anne Mickan<sup>1,2,3</sup>, Ekaterina Slesareva<sup>3</sup>, James M McQueen<sup>1,3</sup>  
and Kristin Lemhöfer<sup>3</sup> 

Quarterly Journal of Experimental Psychology  
1–21

© Experimental Psychology Society 2023



Article reuse guidelines:  
sagepub.com/journals-permissions  
DOI: 10.1177/17470218231181380  
qjep.sagepub.com



## Abstract

Anecdotal evidence suggests that learning a new foreign language (FL) makes you forget previously learned FLs. To seek empirical evidence for this claim, we tested whether learning words in a previously unknown L3 hampers subsequent retrieval of their L2 translation equivalents. In two experiments, Dutch native speakers with knowledge of English (L2), but not Spanish (L3), first completed an English vocabulary test, based on which 46 participant-specific, known English words were chosen. Half of those were then learned in Spanish. Finally, participants' memory for all 46 English words was probed again in a picture naming task. In Experiment 1, all tests took place within one session. In Experiment 2, we separated the English pre-test from Spanish learning by a day and manipulated the timing of the English post-test (immediately after learning vs. 1 day later). By separating the post-test from Spanish learning, we asked whether consolidation of the new Spanish words would increase their interference strength. We found significant main effects of interference in naming latencies and accuracy: Participants speeded up less and were less accurate to recall words in English for which they had learned Spanish translations, compared with words for which they had not. Consolidation time did not significantly affect these interference effects. Thus, learning a new language indeed comes at the cost of subsequent retrieval ability in other FLs. Such interference effects set in immediately after learning and do not need time to emerge, even when the other FL has been known for a long time.

## Keywords

Foreign language attrition; retroactive interference; consolidation

Received: 5 August 2022; revised: 21 April 2023; accepted: 21 April 2023

Most multilinguals share the intuition that learning words in a new foreign language (L3) comes at the cost of retrievability of words in other, previously learned foreign languages (L2). Studies on third language acquisition have paid surprisingly little attention to these frustrating side effects of learning a new language. In fact, to date, there is little experimental evidence documenting them (though see Gollan & Silverberg, 2001), and there have been few attempts to provide an explanation for why learning a new language might negatively affect previously learned ones. The present study seeks to fill this research gap and asks whether learning a new foreign language indeed hampers accessibility to other known foreign languages in the very early stages of vocabulary learning. It also aims at providing insights into under which circumstances, and thus why, such effects emerge.

A possible explanation for why a new foreign language might interfere with older ones is language competition. The languages of a multilingual are thought to interact and compete with one another (Kroll et al., 2008): when a Spanish-English bilingual wants to refer to a *table* in

<sup>1</sup>Max Planck Institute for Psycholinguistics, Nijmegen, The Netherlands

<sup>2</sup>International Max Planck Research School for Language Sciences, Max Planck Institute for Psycholinguistics, Nijmegen, The Netherlands

<sup>3</sup>Donders Institute for Brain, Cognition and Behaviour, Radboud University, Nijmegen, The Netherlands

## Corresponding author:

Kristin Lemhöfer, Donders Institute for Brain, Cognition and Behaviour, Radboud University, Thomas van Aquinostraat 4, 6525 GD Nijmegen, The Netherlands.

Email: kristin.lemhofer@donders.ru.nl

English, the Spanish word *mesa* will supposedly be activated along with the English target word. The resulting between-language lexical competition can delay selection and hence production of the target language word, and can in extreme cases even lead to complete retrieval failure (Costa & Santesteban, 2004; Linck et al., 2012). Based on these documented online language competition effects, it has recently been proposed that between-language competition can, in the long run, lead to language forgetting (i.e., attrition). In a study by Mickan et al. (2020), for example, participants first learned a set of new L3 Spanish words. A day later, participants repeatedly retrieved half of these words in either L1 Dutch or L2 English. In a subsequent Spanish recall test on all originally learned words, recall proved less accurate and slower for Spanish words that had been interfered with (i.e., retrieved in L1/L2) compared with Spanish words that had not been interfered with. These interference effects persisted for an entire week, at least in reaction times, thus linking language competition to long-lasting changes in retrieval ease. These results were replicated in a three-day ERP study with Italian as to-be-learned language and L2 English as the only interference language, providing additional evidence from the N2 component that inhibition is indeed an important mechanism involved in these effects (Mickan et al., 2021). The behavioural findings are supported by a study from Bailey and Newman (2018), who showed that newly learned L2 Welsh words also take longer to be retrieved after retrieval practice of their translation equivalents in L1 English. Together, these studies as well as others (e.g., Isurin & McDonald, 2001) clearly point towards a role for between-language competition in language forgetting.

In the memory literature, effects of new learning on previously learned material are commonly known as “retroactive interference” (RI) effects (McGeoch & McDonald, 1931; Müller & Pilzecker, 1900; Osgood, 1948). In a typical RI study, participants in the experimental group learn a list of paired associations A-B (between syllables, nouns, or other materials) and then another list where the first members of each pair are associated with new stimuli (A-C). Participants in the control group do not learn such a second list and are simply asked to rest. When memory for list 1 is tested again, retrieval accuracy is typically reduced in the experimental compared with the control group, which is—despite a debate on the exact mechanism—generally interpreted as interference from the second to the first list during retrieval (hence “retroactive interference”). Thus, learning new content that shares cues with previous memories impairs memory for these older memories.

Critically, though, classical RI studies test for competition effects on only recently acquired material (list 1 that has been learned just before list 2). However, in the case of foreign languages, the “old” language that is possibly “overwritten” by a newly learnt one has mostly been learned years ago. It remains unclear whether such “old” memories are also affected by such competition dynamics.

A central issue when it comes to the fate of a memory trace over time is *consolidation*. Already Müller and Pilzecker (1900) recognised that memories may need to be consolidated during an “idle mind state” following learning; the detrimental effect of a second list on the recall of a first may simply be due to the lack of opportunity for such a consolidation (see also Dewar et al., 2007, for a summary of Müller & Pilzecker’s insights). Indeed, memory consolidation that occurs over time with a central role of sleep and rest (Dewar et al., 2012; Inostroza & Born, 2013) is now thought to be central for the transition of detailed, but short-lived episodic memories into more stable and durable, mostly semantic memories (Nadel et al., 2012; Squire et al., 2015). Classical RI experiments are concerned with memories that are probably short-lived and purely episodic in nature. The question thus is whether memories that have been transferred to the more permanent semantic memory system, such as the knowledge of foreign language words, are also susceptible to RI-like phenomena.

Similar to traditional RI studies, the above studies on language attrition allowed for very little, if any, consolidation of the new words (Bailey & Newman, 2018; Isurin & McDonald, 2001; Mickan et al., 2020, 2021). Thus, these studies again only demonstrate the impact of language competition for relatively fresh, and most likely still episodic memories, rather than established semantic knowledge, such as the words of an already known L2. It is conceivable that well-established old knowledge is less vulnerable to interference than the fresh L3 knowledge tested in the studies above. For this reason, testing for interference on long ago consolidated words, as will be done in the present study, is a much stricter test of the role of between-language competition during language attrition.

The notion that memory consolidation may play a decisive role in forgetting is supported by some RI studies that have allowed for (at least partial) consolidation of items in the first list before the learning of the second list. The obtained findings suggest that RI has little impact on new knowledge once it is consolidated. For instance, Ellenbogen et al. (2006) showed that an interfering list had no detrimental effects on retention when it was administered after a night of sleep, in contrast to after 12 hr of wakefulness. Landauer (1974) showed that the more time for consolidation there is between a critical item and interfering information, the better its retention. On the grounds of these insights, one would thus expect no or only limited interfering effects of learning a new language on an already known—and typically consolidated—foreign language. This however speaks against the anecdotal reports that learning a new foreign language can make it harder to speak a previously acquired foreign language.

In this article, we report on two experiments that investigate if, and under which circumstances, new language learning hampers access to previously learned and well-consolidated foreign language words. In the first

experiment, we tested a group of Dutch native speakers with good command of L2 English. Participants first completed a picture-based English vocabulary test, on the basis of which a set of participant-specific, *known* English words was chosen. For half of those words, their L3 Spanish translations were subsequently learned, and hence the English words were supposedly interfered with, while this was not the case for the other half. Finally, participants' picture naming accuracy and speed for all L2 English words was measured. If the learning of a new language comes at the cost of remembering already known languages, we should see longer naming latencies and—if this cost is as severe as has been shown for recently acquired L2 knowledge (Mickan et al., 2020, 2021)—lower accuracy in English word productions for which Spanish translations were learned, compared with words for which no translation equivalents have been learned.

In the second experiment, we followed up on the results of the first one under the viewpoint of consolidation, varying the time intervals between the English testing, the Spanish learning, and the final English testing. We will motivate those variations after we have discussed whether any interference effect from Spanish learning can be observed at all, as evident from Experiment 1.

## Experiment 1

### Method

**Participants.** Thirty-one Dutch native speakers with normal or corrected-to-normal vision and without a history of neurological or language-related impairments were recruited from the Radboud University participant pool. Four of

them had to be excluded because they did not know enough English words in the pre-test to construct a matched item list (see Item Selection). One additional person had to be excluded due to a technical failure. The remaining 26 participants (16 female) were aged between 18 and 27 ( $M=21.77$ ) and had Dutch as their only mother tongue.

As determined via an online language background questionnaire completed before participants came to the laboratory, none of the participants, with the exception of one, had any knowledge of Spanish prior to the experiment. The one participant who did report having learned Spanish had just started doing so using a language learning app (Duolingo) 3 weeks prior to participating in the experiment, and judged their Spanish as very poor (one out of seven in all domains, that is, reading, writing, listening, and speaking).<sup>1</sup> All participants reported that English was their first and most frequently spoken foreign language. We asked for proficiency self-ratings on a Likert-type scale from 1 to 7 and participants' frequency of use of English per day in minutes, separately for the four language domains (speaking, listening, reading and writing). After the main experiment, we also measured participants' English vocabulary size using LexTALE (Lemhöfer & Broersma, 2012). The results of these measures are summarised, for this experiment and Experiment 2, in Table 1. Other foreign languages participants had learned included French, German and Latin. For the purpose of this article, we refer to Spanish as an L3, regardless of how many other foreign languages a participant had learned before the study. Participants gave informed consent and received either course credit or vouchers for participation (10€/hr). The study was approved by the Ethics Committee of the Faculty of Social Sciences, Radboud University.

**Table 1.** Participant characteristics as collected in the language background questionnaire for both experiments.

	Experiment 1			Experiment 2					
	M	SD	range	Consolidation group (N=41)			No-consolidation group (N=41)		
M				SD	range	M	SD	range	
English AoA	10.85	1.52	8–14	10.56	1.61	6–14	10.90	1.56	7–14
English LoE (years)	10.69	3.39	6–18	9.54	4.24	2–22	10.20	2.84	4–15
English frequency of Use (min/day)									
Speaking	15.92	27.33	0–120	16.81	23.82	0–127	21.49	52.84	0–240
Listening	153.12	119.75	0–480	145.12	73.24	10–300	148.78	112.79	0–600
Reading	48.81	38.23	0–120	84.76	65.36	0–240	83.17	79.43	0–300
Writing	16.69	22.37	0–60	29.00	43.46	0–240	28.02	56.60	0–300
English proficiency <sup>a</sup>									
Speaking	5.04	0.87	3–7	5.27	1.00	3–7	4.98	1.19	1–7
Listening	6.04	0.72	5–7	6.10	0.77	4–7	5.68	1.08	2–7
Reading	5.73	0.96	4–7	5.90	0.89	3–7	5.83	0.86	3–7
Writing	5.12	1.07	2–7	5.15	1.04	2–7	4.71	1.25	1–7
English LexTALE	74.89	12.04	51–95	74.3	12.03	51–95	72.9	11.00	48–92

M: mean; SD: standard deviation; AoA: Age of acquisition; LoE: length of exposure.

<sup>a</sup>Proficiency self-ratings were given on a scale from 1 (*very bad*) to 7 (*like a native speaker*).

**Materials.** The complete item set consisted of 103 nouns referring to concrete objects and animals (see the online Supplementary Material A). They were non-cognates between Dutch, English and Spanish and were one or two syllables long in English ( $M=1.33$ ,  $SD=0.47$ ) and between two and four syllables long in Spanish ( $M=2.59$ ,  $SD=0.66$ ). Lemma frequencies of the corresponding Dutch words ranged from 0 to 200 per million ( $M=35.90$ ,  $SD=49.48$ ) according to CELEX (Baayen et al., 1995). To match items in terms of frequency between the interference and no-interference sets for each participant, we used the corresponding log frequencies, which ranged from 0 to 2.32 ( $M=1.08$ ,  $SD=0.63$ ). Pictures representing the nouns were photographs taken from Google images or the BOSS database (Brodeur et al., 2010) and were presented on a white background (occupying a maximum of 400 px in either width or length). Each Spanish noun was recorded spoken by a female Spanish native speaker from Andalucía (Spain).

**Item selection.** Per participant, 46 nouns were selected on the basis of the participant's pre-test results, which would be divided into two subsets: 23 words that would be learned in Spanish (interference set) and 23 words that would not (no-interference set). All nouns had to be known in English, that is correctly named in the pre-test at first attempt. The first 46 words from the pre-test were considered the "ideal" item set. If a participant did not know one or more words from this set, these were subsequently replaced with known words from the remaining pre-test items. A MATLAB script (v.8.6, R2015b, The Math Works, Inc.) took care of the replacement and made sure that replacements were as similar to the original item from the base set as possible in terms of the item matching

criteria (mean words replaced = 13.03; range = 1–30), such that the two subsets remained matched. These criteria were Spanish and English word length (measured in syllables), Dutch lemma log frequency according to CELEX (Baayen et al., 1995), mean phonological similarity with all other items in a participant's entire set, both in English and Spanish (Levenshtein distances based on phonemes; Levenshtein, 1966), and average within- and across-set semantic similarity. These semantic similarity values based on cosine similarity between semantic vectors were obtained using the open source online tool *snaut* (<http://meshugga.ugent.be/snaut/>; Mandera et al., 2017). Table 2 indicates the averages of these measures. Assignment of each subset to an interference condition was counterbalanced across participants.

**Procedure.** Participants were tested individually in a quiet room adjacent to the experimenter's room. The door between the rooms was kept open at all times for communication and response coding. The experiment consisted of three parts: an English pre-test to determine an item set for the remainder of the experiment, a Spanish learning phase, and finally a surprise English post-test to assess the effect of the Spanish learning phase on the accessibility of the corresponding English words. Participants were led to believe that the study was about learning Spanish. There was no mention of English in the study description other than the fact that participants would need to take an English vocabulary test at the beginning of the study. The post-test thus came as a complete surprise to all participants (as also confirmed in post-experiment interviews).

**English pre-test.** To select a matched participant-specific item set, participants saw 103 pictures of everyday objects,

**Table 2.** Item characteristics in Experiment 1.

	Interference set (=learned in Spanish)			No-interference set (=not learned in Spanish)		
	<i>M</i>	<i>SD</i>	range	<i>M</i>	<i>SD</i>	range
Spanish word length (in syllables)	2.66	0.69	2–4	2.72	0.70	2–4
English word length (in syllables)	1.39	0.49	1–2	1.40	0.50	1–2
Dutch Celex log frequency	0.94	0.56	0–2.32	0.93	0.55	0–2.32
Dutch Celex lemma frequency per million	22.09	35.68	0–208	20.48	34.37	0–208
Within-set semantic distance <sup>a</sup>	0.81	0.09	0.31–1.10	0.81	0.09	0.39–1.10
	<i>M</i>		<i>SD</i>			range
Spanish Levenshtein distance <sup>b</sup>	6.10		1.49			2–10
English Levenshtein distance <sup>b</sup>	5.18		1.30			2–9
Across-set semantic distance	0.81		0.10			0.31–1.08

Item sets differed across participants, as described in the Item selection section. Mean values (*M*) and standard deviations (*SD*) were first calculated per subject and set, and subsequently averaged over participants. Ranges show the absolute minimal and maximal values per group and set.

<sup>a</sup>Small values for semantic distance mean high similarity.

<sup>b</sup>Small Levenshtein distances mean high similarity. Levenshtein distances were calculated across both word sets.

one at a time, and their task was to name them in English to the best of their knowledge. We encouraged participants to answer as quickly as they could, but there was no time limit. Each trial started with a 500-ms blank screen, followed by the presentation of the picture in the centre of the screen. The experimenter coded participants' answers for accuracy via a button press, which immediately initiated the next trial. Participants did not receive feedback on their answers. An item was considered known, and thus suitable for the experiment, only if the participant was able to name the picture correctly on their first try. Synonyms and partial answers were not considered correct. This strictness was applied to exclude hard-to-retrieve items, or those with a preferred alternative name (synonym), and be able to replace them with better items. Next to accuracy, naming speed was measured and later used as a baseline for the English post-test.

*Spanish learning tasks.* To teach the participants the 23 individually selected Spanish words from the "interference set," a series of learning tasks, moving from rather easy (recognition-based) to more difficult (retrieval/production-based), was administered (see Kang et al., 2013, for a demonstration of the benefits of retrieval-based training): (1) a familiarisation phase; (2) a two-alternative forced-choice task between two words for a given picture; (3) a word completion task; (4) a written naming task; and (5) an adaptive oral picture naming task (continuing until all items were produced correctly). These learning tasks were followed by a final test (picture naming). None of the data of this learning phase were later analysed, except for the final picture naming test that assessed the success of the learning phase.

The first task was a *familiarisation task*. Each trial started with a 500-ms blank screen, followed by the presentation of a picture slightly above the centre of the screen. After 300 ms, the corresponding Spanish word appeared below the picture. After another 300 ms, the audio recording was presented via headphones. Participants were instructed to repeat the words out loud, and were furthermore asked to inform the experimenter if they knew any of the Spanish words already. Spanish words that participants knew already were later excluded from analysis (see Response coding and exclusion criteria section below).

Subsequently, participants completed two rounds of a *two-alternative forced choice task*: After a blank screen (500 ms), a picture was presented at the centre of the screen together with two Spanish words from the 23 to-be-learned words printed below it and to the left and right of the centre, each surrounded by a black circle. Participants had to select the label that corresponded to the picture by clicking on one of the circled words with the mouse. The response was followed by a 500-ms blank screen, after which feedback was presented in form of the circle surrounding the clicked word presented in either green (when the answer

was correct) or red colour. This colour feedback stayed on the screen for 500 ms, after which only the picture and the correct word were displayed together in the centre of the screen and the corresponding audio was played. The picture and its label stayed on screen for another 500 ms after the end of the audio recording, after which the next trial started automatically.

In the second round of this task, participants were asked to try to name the picture before seeing the two Spanish labels. This was done to encourage them to start engaging with the words more actively. After the response was given, the experimenter pressed a key, after which the two word options appeared on screen and the trial continued as in the first round.

This two-alternative forced choice task was followed by two rounds of a *word completion task*. After a blank screen of 500 ms, each picture was presented individually as in the tasks above, but now accompanied by the first syllable of that word (or the first grapheme for monosyllabic words) written below it. Participants had to complete the word by saying it out loud. After a key press by the experimenter, a feedback screen appeared, consisting of either a red or a green square around the picture, together with the full Spanish label and the auditory presentation of the spoken word. 100 ms after the end of the audio file, participants could initiate the next trial by clicking a button.

Next, in a *written naming task*, participants saw (after a 500-ms long blank screen) each picture again and were asked to write its Spanish name down on paper. They subsequently pressed a button to see and hear the correct Spanish word on the screen and correct themselves (on paper) if necessary, and initiated the next trial themselves by a button press.

The final learning task was an *adaptive picture naming task*: participants first went through two rounds of simply naming all pictures again. They saw the pictures and named them orally, coded by the experimenter as correct or incorrect, after which they received feedback by way of red and green frames around the picture as well as the correct answer as in the word completion task above. The next trial was initiated by participants pressing a button. If after these two rounds there were still words left that they were unable to name properly, the task would continue and those unknown words would be repeated until the participant had named each of them correctly at least twice in a row. These optional, subsequent naming rounds always consisted of at least 10 pictures; thus, the set of still to-be-learned words was typically complemented by already learned items. The task script, however, took care to repeat each of the already known words equally often. The adaptive task continued until either all items had been learned, or otherwise until 50 min of the learning phase had passed.

The learning phase then ended in one final round of *picture naming without feedback* as a final test of Spanish word learning.

There was no time limit for participants' responses in any of the above tasks. The presentation order of words was semi-random, such that it was different for every learning task and every participant, but the same across (within-participant) repetitions of a task. This pseudo-random order was chosen to avoid order effects, while at the same time keeping the distance between repetitions within a task constant. The learning phase resulted in a minimum of nine exposures per word with feedback before the final test without feedback. The number of additional exposures to each item depended on the length of the adaptive picture naming task. On average, participants required 12.96 exposures per item (mean  $SD=3.18$ , range = 10–32).

In total, the learning session took a maximum of 1 hr. To continue with the experiment, participants had to correctly produce 18 of the 23 words in the final round of picture naming. Note that even though all words had been learned in the previous adaptive picture naming task, it occasionally happened that words were forgotten again before this final task. All participants satisfied this criterion: on average, participants produced 22 out of the 23 new Spanish words correctly in the final naming task (range = 19–23; see Results section for more details).

**English post-test.** Finally, participants were tested again in English on all 46 words chosen in the pre-test. The procedure was identical to the pre-test.

**LexTALE.** After the word naming tasks, participants completed the English version of the LexTALE, a vocabulary test based on lexical decision (Lemhöfer & Broersma, 2012).<sup>2</sup>

#### Response coding and exclusion criteria

**Accuracy.** Participants' English productions in the post-test were coded as either correct or incorrect/unknown. When participants corrected themselves, or otherwise needed multiple attempts to name a picture, only the last utterance was scored. Synonyms of intended answers were counted as correct (0.26% of correct responses were synonyms: 0.4% in the interference condition and 0.2% in the no-interference condition). Words that were not successfully learned in Spanish ( $M=0.77$  out of a total of 23, that is, on average 3%, range = 0%–4%) or that were known in Spanish prior to the experiment ( $M=0.19$  out of a total of 23 words, on average 1%, range = 0%–2%; as determined in the familiarisation of the Spanish learning phase), were excluded from analysis.

**Naming latencies.** Naming latencies in the English pre- and post-tests were measured manually in Praat (Boersma, 2001) based on response recordings, and reflect the time from picture presentation to speech onset. In the analysis, to take initial between-item differences into account, we calculated difference scores using the English pre-test latencies as the baseline for latencies during the post-test. Because post-

test latencies were typically shorter than pre-test latencies, reflecting an effect of repetition, we subtracted post-test latencies from pre-test latencies.<sup>3</sup> The resulting difference scores thus reflect a *speed-up* in naming latencies from pre- to post-test in English; we hypothesise this speed-up to be smaller for words that were learned in Spanish.

Next to the exclusions mentioned above for accuracy, for the naming latency analysis, we additionally excluded trials in which participants were unable to name a picture, named it incorrectly or took multiple attempts at naming. Trials in which participants corrected themselves, coughed or laughed were also excluded. Smacks and hesitations were accepted though; naming latencies for these trials were measured at the onset of the actual word production. These extra exclusions resulted in an average of 3% additional data loss per participant ( $M_{Int}=4\%$ , range = 0%–17%;  $M_{NoInt}=3\%$ , range = 0%–13%). Finally, trials in which participants used an article before the noun in the pre-test, but not the post-test, or vice versa, were also excluded from the latency analysis, because the pre-post naming latency comparison was impossible for these trials. Exclusions due to these article trials resulted in an additional loss of on average 8% of RT data ( $M_{Int}=9\%$ , range = 0%–56%;  $M_{NoInt}=7\%$ , range = 0%–56%). Participants for whom all exclusions taken together resulted in more than 30% data loss in either the interference or the no-interference condition were excluded from the naming latency analysis ( $N=3$ ). The remaining 23 participants had an average of 41 out of 46 trials left for analysis (range 33–46 trials,  $M_{Int}=20.04$ ,  $M_{NoInt}=21.30$ ). Note that this means that the naming latency analysis is based on fewer participants than the accuracy analysis.

**Modelling.** We analysed the data using (generalised) mixed-effects models with the lme4 package (version 1.1-21, Bates et al., 2015) in R (Version 3.5.1, R Core Team, 2019). The accuracy data were analysed using a generalised mixed-effects model of the binomial family, fitted by maximum likelihood estimation, using the logit link function and the optimiser “bobyqa.” The dependent variable consisted of 1's and 0's for correct and incorrect words, respectively. The only fixed effects variable was Interference (two levels: no interference, interference) and it was effects coded (–0.5, 0.5). This means that negative beta coefficients reflect lower accuracy for interfered items than not-interfered items. Random effects were fitted to the maximum structure justified by the experimental design (Barr et al., 2013), which included random intercepts for both Subjects and Items, as well as random slopes by Subject for Interference. Random slopes that correlated highly ( $r > 0.95$ ) with their respective intercepts were removed to avoid over-fitting. All  $p$ -values were calculated by model comparison, omitting one factor at a time, using chi-square tests.

Naming latencies were analysed using a linear mixed-effects model fitted by restricted maximum likelihood

estimation (using Satterthwaite approximation to degrees of freedom). Raw naming latencies were first log-transformed and then difference scores (pre-test RT—post-test RT) were calculated and entered into the model. The fixed and random effects structure was identical to the accuracy model. Interference was again effects coded with the contrasts (−0.5, 0.5), meaning that negative beta coefficients in the RT model reflect a smaller latency speed-up from pre- to post-test in English for interfered as compared with not interfered items.

## Results

The full data of this study are available under <https://doi.org/10.34973/x2dw-jr34>.

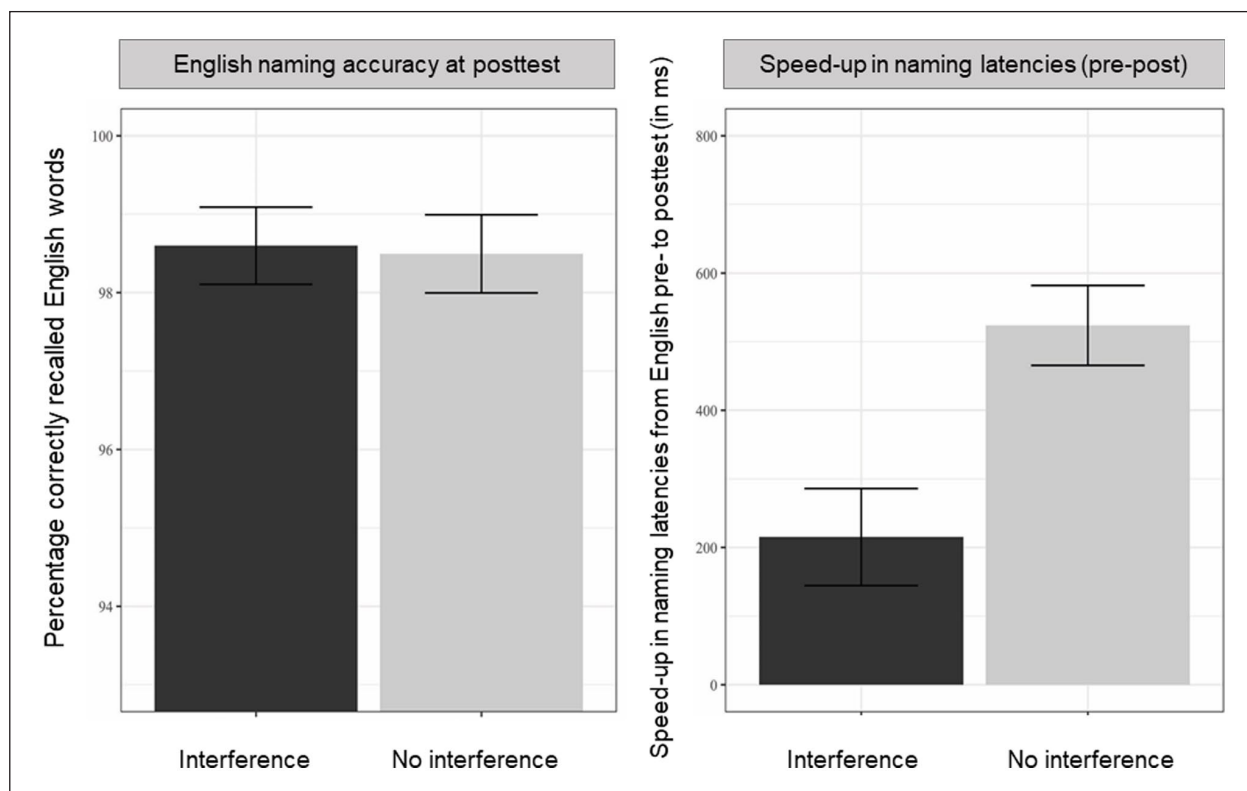
**English pre-test performance.** On average, participants knew 77% ( $SD=14\%$ , range=49%–93%) of all 103 words from the English pre-test. As explained above, only known items were chosen as the 46 experimental items for each individual participant.

**Spanish learning performance.** On average, participants learned 97% ( $SD=5\%$ , range=83%–100%) of the 23 new Spanish words.

## English final test performance

**Naming accuracy.** The mean percentage of correctly recalled English words for the two interference conditions at final test in English is shown in the left panel of Figure 1. The model output is shown in Table 3. The model included random intercepts for subjects and for items, but because of a too high correlation with the intercept, it did not include random slopes for interference over participants (nor item, which would not make sense as the items differed between conditions and often also participants). As can be seen in Table 3, there was no effect of Interference on participants' English post-test production accuracy. Words for which the Spanish translation equivalents had been learned were thus recalled as often as words for which that was not the case.

**Naming latencies.** The mean naming latency speed-up from pre- to post-test in English for the two interference conditions is plotted in the right panel of Figure 1. Model outcomes can be found in Table 3. The model included random intercepts for subjects and items, and random slopes for interference over participants. Here, we did observe the expected main effect of Interference, such that the difference in naming latencies from pre- to post-test was smaller (i.e., less speeding up) in the interference than in the no-interference condition.



**Figure 1.** Accuracy scores and naming latency speed-up (pre-test—post-test; in ms) in English productions at final test in English. “Interference” and “No Interference” means that the Spanish translation equivalent of the word had or had not been learned. Error bars reflect standard errors.

**Table 3.** Mixed-effect model output for naming accuracy and naming latencies in Experiment 1.

Fixed effects	Naming accuracy				Naming latencies			
	Estimate	SE	z	$p(\chi^2)$	Estimate	SE	t	$p(\chi^2)$
Intercept	5.58	0.75	7.43	<.001	0.22	0.03	7.63	<.001
Interference	0.38	1.03	0.37	0.708	-0.24	0.03	-7.28	<.001
Random effects	Groups	Var	SD	Corr	Groups	Var	SD	Corr
Item	Intercept	1.92	1.39		Intercept	0.02	0.15	
Subject	Intercept	0.65	0.80		Intercept	0.01	0.09	
	Interference	4.24	2.06	0.29	Interference	0.00	0.06	0.63

Significant effects are marked in bold.

SE: standard error; SD: standard deviation;  $p(\chi^2)$ : chi-square test statistic; Var: variance; Corr: correlation.

## Discussion

In Experiment 1, we asked whether learning a new foreign language comes at the expense of retrieval of other, previously learned foreign languages. Participants learned new Spanish words, and we assessed whether this new learning affected the accessibility of their (previously known) English translation equivalents. While English words were still recalled correctly at post-test regardless of whether they had been learned in Spanish or not, participants speeded up less from pre- to post-test for words for which they had learned Spanish translations compared with words for which they had not. Although we did not find evidence for interference in recall accuracy, Spanish learning thus hindered the otherwise expected latency speed-up in subsequent English productions. Hence, we can conclude that learning words from a new language does come at the cost of at least retrieval *ease* for words in previously learnt foreign languages.

To our knowledge, we are the first to experimentally demonstrate a detrimental effect of learning a new foreign language on a previously learned FL. While previous research had shown that repeatedly retrieving already known translation equivalents hampers subsequent access to newly learned L3 words (e.g., Bailey & Newman, 2018; Mickan et al., 2020, 2021), these results show that the opposite is also true: new learning negatively affects retrieval ease in a long before acquired foreign language. This study thus differs from earlier studies in two crucial ways: first, instead of the retrieval of known words, it is the *new learning of L3 words* that leads to interference. Furthermore, the English target words were long known to the participants rather than taught to them during the experimental session. Thus, they were no longer episodic memories, but instead “old,” semantic memories, and hence presumably much harder to interfere with than the newly learned items in, for example, Mickan et al. (2020).

Given these differences, it is perhaps not surprising that we did not observe accuracy effects in this experiment. In

fact, some RI studies found that *consolidated* material is less susceptible, if at all, to interference from subsequent learning of new information than fresh, unconsolidated knowledge is (e.g., Ellenbogen et al., 2006; Landauer, 1974). In general, it is assumed that the more time a memory has to consolidate, the less it will suffer from subsequent interfering tasks (see Müller & Pilzecker, 1900, for the original formulation of this argument). Whether this is really the case is still being debated (see “General discussion,” and Wixted, 2004, for a comprehensive discussion of the role of consolidation in RI). Regardless though, the idea that consolidated memories are *resistant* to interference is at odds with the fact that we do find clear evidence for interference of Spanish word learning on English words in reaction times. Reaction times have often been ignored in the RI literature, despite their potential to uncover nuances of retrieval difficulty that go unnoticed with a dichotomous “remembered—not remembered” response coding (see Postman & Kaplan, 1947, for a more elaborate account of this argument). Clearly, new learning can lead to retrieval difficulties of well consolidated material, just possibly not to the extreme extent (i.e., complete retrieval failure) that is usually probed in RI studies.

Outside of the RI literature, a slow-down in retrieval speed has been interpreted as evidence for the early stages of forgetting (see Mickan et al., 2020, for a discussion). Yet, we clearly did not make our participants completely *forget* English words by teaching them their Spanish translation equivalents. Are such extreme interference effects impossible to induce in the lab, or do they simply take longer to set in? Even in real life, it is very possible that the detrimental effects of learning a new language do not surface until a few months into the learning process. The newly learned Spanish words, being fragile and not yet consolidated themselves, might not yet interact with translation equivalents from other languages in the way necessary for maximal interference effects to arise (see the “General discussion” for a more elaborate consideration of RT vs accuracy effects).



Earlier, we explained between-language interference effects in production in terms of lexical competition between translation equivalents (e.g., Mickan et al., 2020). From research on novel word learning, however, it appears that some types of lexical competition do not emerge immediately, but instead require consolidation. Dumay and Gaskell (2007), for example, showed that during word recognition, newly learned words only compete with phonological neighbours after a night of sleep (see also Bakker et al., 2014; Bowers et al., 2005). Similarly, Maciejewski et al. (2020) showed that after learning novel (additional) meanings of L1 words across several days, participants were slower to retrieve the original meaning of these words; however, note that in the study by Fang and Perfetti (2019), this slow-down was only temporary. While these studies investigate competition effects between form-similar or form-overlapping words in L1 in perception rather than competition between translation equivalents in production, it is possible that the same principles hold for between-translation competition, hence, that newly acquired vocabulary also needs to consolidate first before it can compete with translation equivalents. Davis and Gaskell (2009) explain their findings (with form-similar word pairs) by assuming that novel words, just like any new memories, are initially encoded as episodic traces that are heavily dependent on the hippocampus. Through offline consolidation, and particularly through sleep, these episodic traces become gradually integrated into the existing neo-cortical memory network, and only then start to compete with related words, such as phonological neighbours, in the mental lexicon. Research suggests that this consolidation process is aided by the first night of sleep (Dumay & Gaskell, 2007), but also that it is a gradual process that can take up to multiple weeks to complete (Takashima et al., 2006).

If novel words require such a slow integration process, it might seem puzzling that we found any interference effects at all, even in reaction times. There is, however, also evidence for immediate lexical integration of newly learned words (e.g., Borovsky et al., 2012; Coutanche & Thompson-Schill, 2014; Lindsay & Gaskell, 2013). For instance, Elias and Degani (2022) showed that the learning of Arabic interlingual homophones to Hebrew (words with the same pronunciation, but different meaning) by native speakers of Hebrew slowed down subsequent lexical decision RTs to these Hebrew words, compared with control words. Thus, having learnt a new, competing meaning to an already known word form immediately slowed down the processing of that word form that had already been in the mental lexicon for a long time.

Interestingly, outside the lexical domain, integration has been shown to be especially fast when the newly

acquired knowledge fits in with existing knowledge (e.g., Bellana et al., 2021; van Kesteren et al., 2010). Since newly learned translation equivalents share their concepts with existing words, it is very possible that they can benefit from such fast integration. The RT effect, in our eyes, then reflects the immediate beginning of the lexical integration process, allowing the Spanish words to interact with English words and causing a slow-down in subsequent retrieval of the latter. Possibly, however, they were not yet integrated *enough* to entirely block access to their English translation equivalents (as would be apparent in an accuracy effect). Assuming that consolidation is a gradual process that is crucially aided by sleep, the newly learned Spanish words might become much stronger interferers after a longer consolidation time window including a night of sleep. Experiment 2 addresses this possibility by separating the final English test from the Spanish learning by roughly 24 hr.

In addition to manipulating the consolidation time window for the newly learned Spanish words, we also changed the timing of the pre-test. The reason for this change relates to an alternative explanation for the RT effect. Research has shown that despite the initial consolidation process that makes memories less prone to interference, the *retrieval* of a consolidated memory in fact can make it fragile and hence susceptible to interference again (Walker et al., 2003). The destabilised memory then needs to be “re-consolidated,” a process that is faster than the original consolidation, but that can still take up to 6 hr (Stickgold & Walker, 2005). If retrieving a memory destabilises it, we might have artificially increased the chances of interference arising for the English words by having participants retrieve them in the pre-test, immediately before some of them were learned in Spanish. To avoid this potential problem and to assess the robustness of our interference effect, we thus separated the English pre-test from the Spanish learning phase by roughly 24 hr (a time frame that is long enough to allow for complete reconsolidation, should it be necessary; see Stickgold & Walker, 2005).

A final change in Experiment 2 concerned the introduction of an additional Spanish post-test in both groups, directly before the English testing. This was necessary to assess whether consolidation after learning the Spanish words (i.e., a night of sleep) had had any effect on their *retention* (which, in turn, would of course affect their potential to interfere). Note, though, that the mere incidence of testing, according to the “re-consolidation” notion sketched above, could potentially also lead to destabilisation, particular of already consolidated material. If that was the case, the consolidation group’s results might differ from those of the no-consolidation group because of more destabilised Spanish word memory. We will return to this issue in the “General discussion.”

To assess the effects of both the pre- and post-test timing changes, we needed to test two groups of participants in Experiment 2. Both groups differed from the group tested in Experiment 1 in that they completed the English pre-test one day before the Spanish learning phase. The two groups in Experiment 2, however, differed from each other in the timing of the English post-test: the no-consolidation group was tested in English immediately after the Spanish learning phase (i.e., as in Experiment 1), while the consolidation group was tested in English only a day later (see Figure 2). We hypothesised that interference effects would be stronger, as apparent in an effect in accuracy and/or a larger naming latency effect, for the consolidation group compared with the no-consolidation group, because the already consolidated Spanish words in the first group should be stronger interferers for the English words. Furthermore, a comparison of the no-consolidation group with the group tested in Experiment 1 will address whether the RT effects found in Experiment 1 are reliable and still obtained even when an interruption of reconsolidation during pre-test is avoided and thus can no longer be the reason for interference.

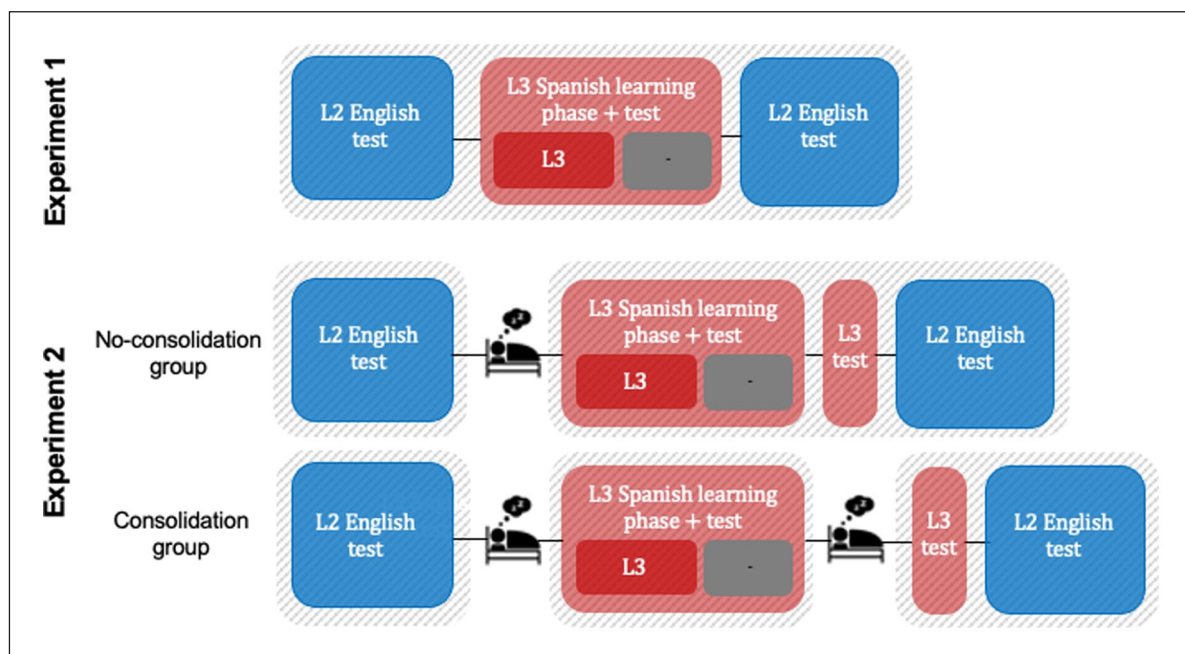
## Experiment 2

The experimental setup of Experiment 2 differed from that of Experiment 1 only in the timing of the English pre-test (1 day before the Spanish learning) and the post-test (consolidation group: 1 day after Spanish learning; no-consolidation group: same day as Spanish learning).

## Method

**Participants.** A total of 86 Dutch native speakers with normal or corrected-to-normal vision and without a history of neurological or language-related impairments participated in Experiment 2. None of them had taken part in Experiment 1. One of them had to be excluded from the analysis because they did not learn enough Spanish words. Two participants were excluded because they failed to show up for the final experimental session. One additional participant had to be excluded due to a technical failure. The remaining 82 participants (57 female) were between 18 and 29 years of age ( $M=21.99$ ) and had Dutch as their only mother tongue. None except four of the participants indicated having any prior knowledge of Spanish. The four participants who did report having learned some Spanish in the past, had done so for very short amounts of time ( $M=4.75$  months, range=1–9), rated their current knowledge of Spanish as very poor ( $M=1.38$ ;  $SD=0.48$ , range=1–2, on a scale from 1 to 7) and stated that they hardly ever spoke Spanish. Their knowledge of Spanish can thus be described as minimal. Furthermore, as in Experiment 1, all participants reported to have English as their first and most frequently spoken foreign language (see Table 1 for reports of frequency of use, proficiency self-ratings and English LexTALE scores). Other foreign languages participants knew again included most prominently French, German, and Latin.

Upon coming to the lab, participants were randomly assigned to either the consolidation group ( $N=41$ ) or the no-consolidation group ( $N=41$ ). As confirmed by



**Figure 2.** Schematic representation of the experimental setup for both Experiment 1 and 2. Striped boxes in the background indicate separate testing days.

**Table 4.** Characteristics of the extended set of items used in Experiment 2.

	Consolidation group						No-consolidation group					
	Interference set			No-interference set			Interference set			No-interference set		
	<i>M</i>	<i>SD</i>	range	<i>M</i>	<i>SD</i>	range	<i>M</i>	<i>SD</i>	range	<i>M</i>	<i>SD</i>	range
Spanish word length (in syllables)	2.60	0.70	2–5	2.62	0.69	2–5	2.60	0.67	2–5	2.60	0.69	2–5
English word length (in syllables)	1.36	0.50	1–4	1.39	0.50	1–3	1.40	0.53	1–4	1.37	0.50	1–3
Dutch Celex log frequency	1.02	0.58	0–2.32	1.02	0.59	0–2.91	1.05	0.58	0–2.91	1.05	0.59	0–2.32
Dutch Celex frequency per million	22.60	31.76	0–208	23.44	35.98	0–818	24.90	35.47	0–818	25.00	34.54	0–208
Within-set semantic distance <sup>a</sup>	0.83	0.09	0.31–1.09	0.83	0.11	0–1.09	0.82	0.09	0.37–1.09	0.81	0.10	0–1.09
	<i>M</i>		<i>SD</i>	range		<i>M</i>		<i>SD</i>	range			
Spanish Levenshtein distance	5.94		1.52	2–10		5.94		1.52	2–10			
English Levenshtein distance	5.05		1.23	2–11		5.08		1.23	2–11			
Across-set semantic distance <sup>a</sup>	0.83		0.08	0.37–1.11		0.83		0.08	0.40–1.11			

Item sets differed across participants, as described in the Item selection section. Mean values (*M*) and standard deviations (*SD*) were first calculated per subject and interference condition, and subsequently averaged over groups. Ranges show the absolute min and max values per group and condition.

<sup>a</sup>Small values for semantic distance mean high similarity.

independent *t*-tests, the two groups did not differ in their average frequency of use or their performance on the English LexTALE and were also comparable on their English proficiency self-ratings (see Table 1).

**Materials.** The stimulus database from Experiment 1 was extended from 103 to 140 concrete and non-cognate words referring to everyday objects or animals (the full list can be inspected in the online Supplementary Material B). We did so to ensure that the item selection script (same as in Experiment 1) would succeed at constructing a participant-specific item list in as many cases as possible, and hence to decrease the drop-out rate at pre-test compared with Experiment 1. These 140 words were between one and five syllables long in Spanish ( $M=2.56$ ,  $SD=0.68$ ), and between one and four syllables long in English ( $M=1.42$ ,  $SD=0.60$ ). The CELEX lemma frequencies of the corresponding Dutch translation equivalents ranged from 0 to 818 occurrences per million ( $M=29.02$ ,  $SD=76.69$ , Baayen et al., 1995). To match items in the interference and no-interference sets, we again used the corresponding log frequencies, which ranged from 0 to 2.91 ( $M=0.98$ ,  $SD=0.65$ ). Pictures were identical to those in Experiment 1, with additional pictures taken from Google images. Audios were recorded by the same female Spanish native speaker from Andalucía (Spain) as in Experiment 1.

**Item selection.** As in Experiment 1, if the pre-defined set of 46 words could not be used due to unknown words, a MATLAB script created a participant-specific item set of 46 known English words based on each participant's pre-test performance. The item selection and replacement procedure were identical to that in Experiment 1 (mean words replaced=17.81; range=3–36, mean of 16.12 in consolidation group, mean of 19.49 in no-consolidation group). As can be seen in Table 4, words in the two inter-

ference subsets were matched on the same variables as in Experiment 1.

**Procedure.** Experiment 2 took place over multiple days (see Figure 2 for a schematic representation of the experimental set up for both experiments). On Day 1, participants came to the lab to take the English pre-test. One day later, they returned for the Spanish learning session (time between the two sessions: consolidation group:  $M=23.46$  hr,  $SD=2.23$ , range=19–29; no-consolidation group:  $M=23.29$  hr,  $SD=2.82$ , range=16–29). On the same day, the no-consolidation group also completed the final English post-test. The consolidation group instead was sent home after the Spanish learning phase and only took the final English post-test on the next day (after  $M=24.12$  hr,  $SD=2.75$ , range=19–30). All tasks were administered exactly as in Experiment 1. The only difference in tasks between the two experiments concerns the final Spanish test: in Experiment 2, participants underwent two final Spanish tests without feedback, as opposed to just one in Experiment 1. The consolidation group did the first of those post-tests on Day 2 immediately after learning (i.e., as in Experiment 1) and the second on Day 3 before the final English post-test. The no-consolidation group, in turn, did both Spanish post-tests in immediate succession after the Spanish learning phase on Day 2. The second Spanish test was added primarily to assess whether the participants in the consolidation group had forgotten any words overnight. Note that next to providing a measure of overnight Spanish retention for the consolidation group, the additional Spanish test on Day 3 also served to match the two groups in terms of recency of exposure to Spanish prior to the final English post-test, such that the only difference between the two groups was whether or not the Spanish words had time to consolidate overnight.

### Coding and exclusion criteria

**Accuracy.** Participants' answers were scored as in Experiment 1. Only trials in which participants were either entirely unable to name the picture in English during the post-test (consolidation group: 83% of all errors, no-consolidation group: 92% of all errors), or named it incorrectly (consolidation group: 17% of all errors, no-consolidation group: 8% of all errors) were counted as errors, synonyms were not (consolidation group: 0.5% of all correct answers, no-consolidation group: 0.8% of all correct answers). Moreover, words which participants already knew in Spanish before starting the experiment ( $M_{\text{ConsolGroup}} = .20$  out of 23, 1%,  $\text{range}_{\text{ConsolGroup}} = 0\%–2\%$ ,  $M_{\text{NoConsolGroup}} = .27$  out of 23, 1%,  $\text{range}_{\text{NoConsolGroup}} = 0\%–2\%$ ), words that were not successfully learned in Spanish, as assessed during the second Spanish post-test ( $M_{\text{ConsolGroup}} = 1.85$  out of 23, 8%,  $\text{range}_{\text{ConsolGroup}} = 0–7$ ,  $M_{\text{NoConsolGroup}} = 1.24$  out of 23, 5%,  $\text{range}_{\text{NoConsolGroup}} = 0–7$ ) as well as words that had accidentally been coded as correct in the pre-test, but that were actually unknown to participants in English ( $M_{\text{ConsolGroup}} = .12$  out of 23, 1%,  $\text{range}_{\text{ConsolGroup}} = 0–1$ ,  $M_{\text{NoConsolGroup}} = .12$  out of 23, 1%,  $\text{range}_{\text{NoConsolGroup}} = 0–3$ ) were again excluded from all subsequent analyses.

**Naming latencies.** As in Experiment 1, trials excluded from the accuracy analysis were also excluded from the RT analysis. On top of that, as in Experiment 1, trials in which participants made errors, took multiple attempts at naming, corrected themselves or coughed or laughed were excluded from RT analysis (6% of trials on average,  $M_{\text{ConsolGroup}} = 5\%$ ,  $M_{\text{ConInt}} = 7\%$ ,  $M_{\text{ConNoInt}} = 4\%$ ;  $M_{\text{NoConsolGroup}} = 6\%$ ,  $M_{\text{NoConInt}} = 8\%$ ,  $M_{\text{NoConNoInt}} = 5\%$ ). In addition, trials in which participants inconsistently used articles at pre- but not post-test in English or vice versa were also excluded from RT analysis, resulting in an additional loss of on average 3% of trials per person ( $M_{\text{ConsolGroup}} = 2\%$ ,  $M_{\text{ConInt}} = 2\%$ ,  $M_{\text{ConNoInt}} = 3\%$ ;  $M_{\text{NoConsolGroup}} = 3\%$ ,  $M_{\text{NoConInt}} = 3\%$ ,  $M_{\text{NoConNoInt}} = 3\%$ ). As in Experiment 1, participants who after all exclusions had less than 70% of trials in either of the two interference conditions left, were excluded from RT analysis ( $N_{\text{ConsolGroup}} = 3$ ,  $N_{\text{NoConsolGroup}} = 1$ ). The remaining 78 participants had an average of 41 of 46 trials left ( $M_{\text{ConsolGroup}} = 41.34$ ,  $\text{range} = 37–46$ ,  $M_{\text{ConInt}} = 19.22$ ,  $M_{\text{ConNoInt}} = 21.46$ ;  $M_{\text{NoConsolGroup}} = 40.53$ ,  $\text{range} = 32–46$ ,  $M_{\text{NoConInt}} = 19.24$ ,  $M_{\text{NoConNoInt}} = 21.07$ ).

**Modelling.** As in Experiment 1, the data were analysed using (generalised) mixed-effect models with lme4 in R. Fixed effects were Interference (two levels: no interference, interference) and Group (two levels: consolidation and no-consolidation) as well as their interaction. Both fixed factors were effects coded (−0.5, 0.5). All other model specifications were as in Experiment 1.

### Results

**English pre-test performance.** The accuracy of the two groups in the English pre-test is presented in Table 5. The two groups did not differ significantly in naming performance,  $t(78.11) = 1.75$ ,  $p = .085$ ;  $d_f$ 's Welch-adjusted for unequal variances.

**Spanish learning performance.** Participants in both groups were successful at learning the 23 Spanish words; their accuracy in both Spanish post-tests is indicated in Table 5. The two groups did not significantly differ from each other in either post-test, nor did the two post-test moments differ for the no-consolidation group (all  $p > .10$ ). However, the consolidation group showed some forgetting from the first to the second post-test, taken on the next day;  $t(40) = 4.49$ ,  $p < .001$ .

### English final test performance

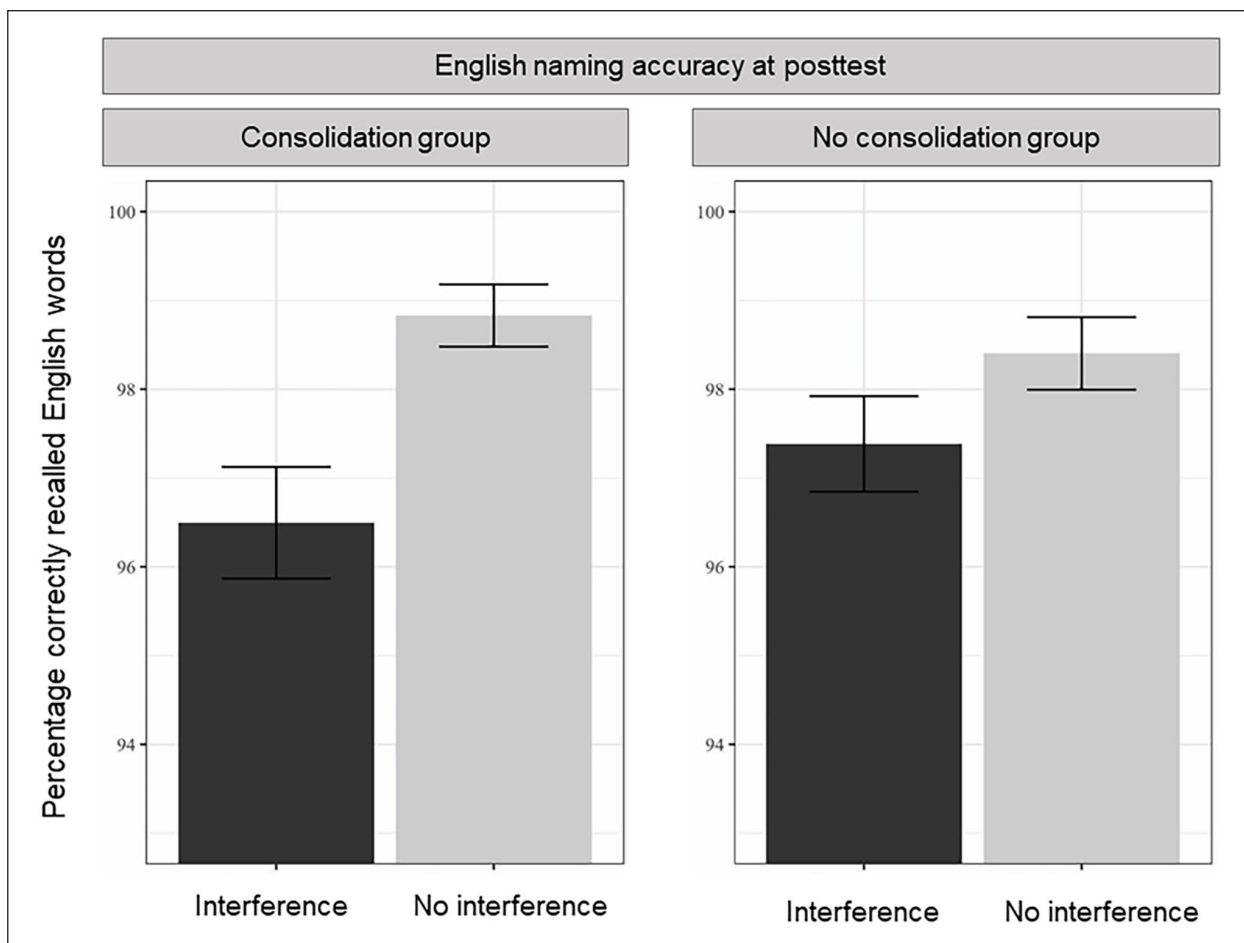
**Naming accuracy.** Average naming accuracy scores per group and interference condition can be found in Figure 3 and model outcomes for accuracy scores can be found in Table 6. As in Experiment 1, the model included random intercepts for subjects and items, but no random slopes for interference over participants due to a too high correlation with the intercept. Unlike in Experiment 1, we observed a main effect of Interference on recall accuracy, such that interfered items were recalled less well than non-interfered items. There was no main effect of Group, nor was the interaction between Interference and Group significant.

**Naming latencies.** Mean latency speed-up from pre- to post-test in English for both groups and interference conditions is shown in Figure 4, and model outcomes for naming latencies are reported in Table 6. As in Experiment 1, the model included random intercepts for subjects and items, and random slopes for interference over participants. Again, we observed a main effect of Interference,

**Table 5.** Accuracy (in %) in English pre-test and the two Spanish post-tests after Spanish learning.

	English pre-test		First Spanish post-test		Second Spanish post-test	
	M (SD)	range	M (SD)	range	M (SD)	range
Consolidation group	67 (12)	41–89	97 (4)	83–100	92 (8)	70–100
No-Consolidation group	62 (14)	34–83	95 (6)	65–100	95 (6)	70–100

SD: standard deviation.



**Figure 3.** Accuracy scores in English productions at final test in English in Experiment 2. Error bars reflect standard errors.

**Table 6.** Mixed-effect model output for naming accuracy and naming latencies in Experiment 2.

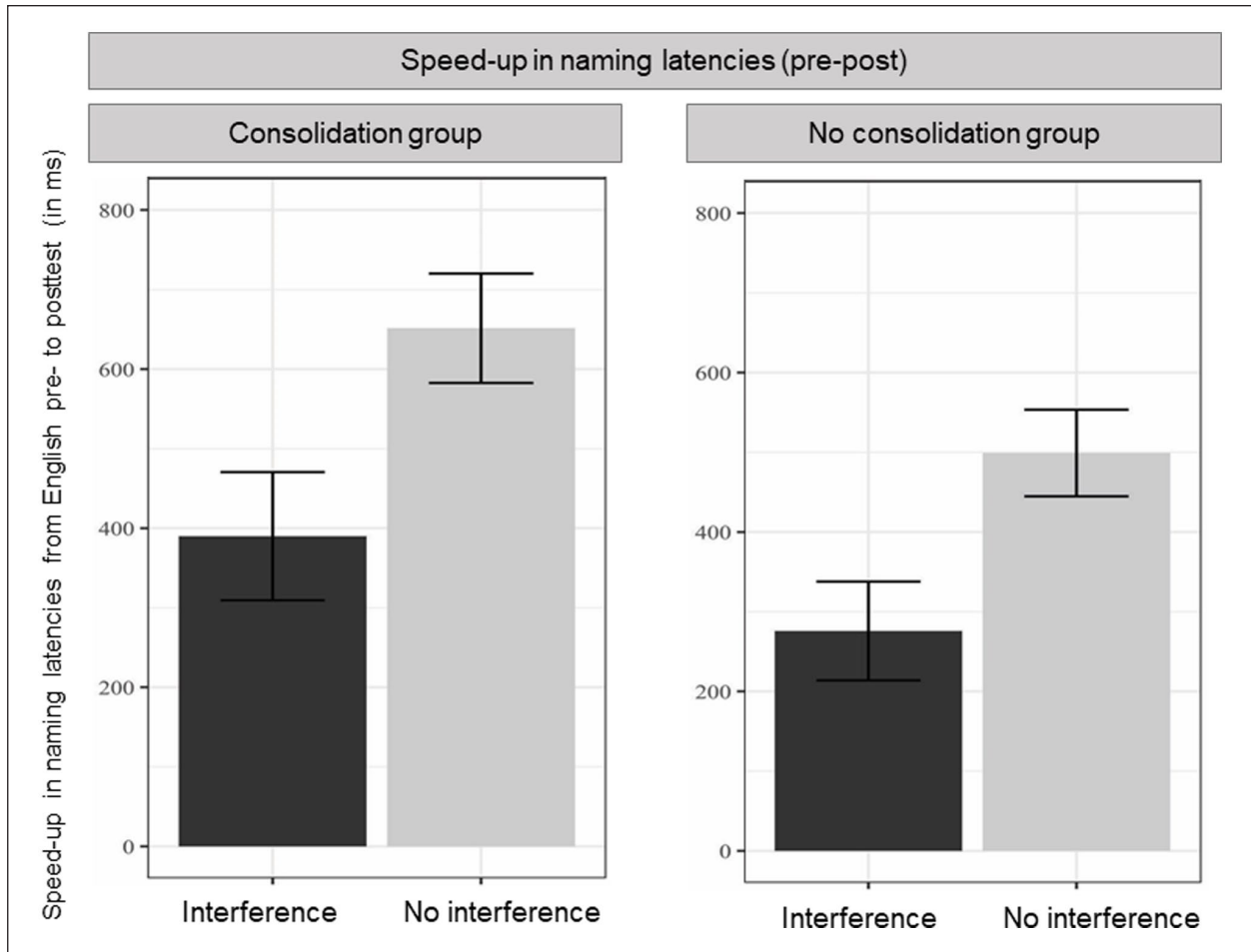
Fixed effects	Naming accuracy				Naming latencies			
	Estimate	SE	z	$p(\chi^2)$	Estimate	SE	t	$p(\chi^2)$
Intercept	4.30	0.24	17.73	<b>&lt;.001</b>	0.23	0.03	8.82	<b>&lt;.001</b>
Interference	-0.80	0.25	-3.23	<b>.001</b>	-0.17	0.02	-7.42	<b>&lt;.001</b>
Group	0.05	0.26	0.18	.856	0.01	0.04	0.22	.821
Interference*Group	-0.58	0.50	-1.17	.241	0.01	0.05	0.09	.926
Random effects	Groups	Var	SD	Corr	Groups	Var	SD	Corr
Item	Intercept	1.20	1.09		Intercept	0.02	0.15	
Subject	Intercept	0.08	0.28		Intercept	0.03	0.17	
					Interference	0.02	0.12	0.54

Significant effects are marked in bold.

SE: standard error; SD: standard deviation;  $p(\chi^2)$ : chi-square test statistic; Var: variance; Corr: correlation.

indicating that participants speeded up more from pre- to post-test in the no-interference condition than the interference condition. There was no main effect of Group and no interaction between the two fixed factors.

*Joint analysis of Experiments 1 and 2.* To test for differences in interference magnitude between the three groups tested in both experiments, we ran mixed-effect models for both accuracy and naming latencies with Group as a 3-level



**Figure 4.** Naming latency speed-up (pre-post; in ms) in English productions at final test in English in Experiment 2. Error bars reflect standard errors.

**Table 7.** Contrast vectors for the comparison of three groups, including the one from Exp. 1.

	No Consolidation vs. Exp. 1	Exp. 1 vs. Consolidation
No Consolidation	-2/3	-1/3
Exp. 1	1/3	-1/3
Consolidation	1/3	2/3

factor (forward difference contrast coded, first contrast compares the no-consolidation group with the group from Experiment 1, and the second contrast compares the consolidation group with the group from Experiment 1; see Table 7), and Interference (no interference, interference) effects coded (-0.5, 0.5) as usual.<sup>4</sup>

Recall that the only difference between the no-consolidation group and the group of Experiment 1 was the time of the English pre-test (on same day vs. one day before Spanish learning). The consolidation group additionally differed from the other two groups in that the English post-test was administered one day after the Spanish

learning, rather than immediately after. The model outcomes can be inspected in Table 8. Both for naming accuracy and naming latencies, we found a significant main effect of Interference, indicating that overall, participants were less accurate and sped up less from English pre- to post-test in naming pictures for which they had learned Spanish translation equivalents. We observed no main effects of Group in either model, nor did any of the groups differ in the magnitude of their interference effect in naming latencies. In naming accuracy, the consolidation group had a numerically larger interference effect than the group in Experiment 1; the interaction, however, did not quite reach statistical significance ( $p = .06$ ). The no-consolidation group did not differ from the group in Experiment 1.

Finally, to check whether the apparent difference in interference effects regarding accuracy between experiments is in fact meaningful, we compared the interference effect from Experiment 1 with the combined interference effect (i.e., the overall main effect in accuracy) obtained in Experiment 2. More specifically, we ran a model with a contrast coding that compares the group from Experiment

**Table 8.** Mixed-effect model output for naming accuracy and naming latencies for both experiments combined.

Fixed effects	Naming accuracy				Naming latencies			
	Estimate	SE	z	$p(\chi^2)$	Estimate	SE	t	$p(\chi^2)$
Intercept	4.41	0.22	20.20	<.001	0.22	0.02	10.21	<.001
Interference	-0.52	0.23	-2.21	.028	-0.19	0.02	-9.47	<.001
GroupC1: NoConsol-vs-Exp1	0.50	0.34	-1.50	.127	-0.01	0.05	-0.18	.856
GroupC2: Consol-vs-Exp1	-0.46	0.34	-1.34	.178	0.02	0.05	0.36	.711
Interference*GroupC1	0.57	0.60	0.94	.345	-0.06	0.05	-1.12	.255
Interference*GroupC2	-1.15	0.61	-1.88	.060	0.07	0.05	1.26	.202
Random effects	Groups	Var	SD		Groups	Var	SD	Corr
Item	Intercept	1.05	1.03		Intercept	0.02	0.14	
Subject	Intercept	0.16	0.40		Intercept	0.02	0.16	
					Interference	0.01	0.11	0.54

Significant effects are marked in bold.

SE: standard error; SD: standard deviation;  $p(\chi^2)$ : chi-square test statistic; Var: variance; Corr: correlation; Consol: Consolidation; Exp1: group from Experiment 1.

1 to the average of both groups from Experiment 2 (Helmert contrast) and found no significant interaction,  $\beta = 1.14$ ,  $z = 1.55$ ,  $p(\chi^2) = .122$ , meaning that the numerical difference between experiments is not substantial and very possibly just due to chance.

## Discussion

Experiment 2 was designed to answer two questions: first and foremost, we wanted to know whether allowing the newly learned Spanish words time to consolidate would make them stronger interferers with English translation equivalents, especially with respect to an effect in naming accuracy, that is, production failures, that had been absent in Experiment 1. Second, we were interested in whether the interference effect in naming latencies observed in Experiment 1 would persist and replicate even if we separated the English pre-test from the Spanish learning phase. This design allowed for full reconsolidation of the English target words unlike in Experiment 1, where the lack thereof may have favoured the emergence of interference effects. To that end, we compared two groups of participants, both of whom took the English pre-test a day before learning Spanish, but for whom the English post-test took place on the same day or one day after they had learned the Spanish words. Results showed an overall interference effect in naming latencies, thus replicating Experiment 1, despite the separation of the English pre-test from the Spanish learning phase. Interestingly, unlike in Experiment 1, we also observed an overall interference effect in naming accuracy: learning Spanish translations thus actually made participants forget some of the corresponding English words. Neither of those effects, however, was substantially modulated by consolidation time for the newly learned Spanish words. Based on the present study, we conclude

that consolidation, at least after one night of sleep, does not seem to make newly learned words stronger interferers.

## General discussion

In this study, we asked whether learning a new language can make you forget previously learned foreign languages, and whether such detrimental effects set in immediately after learning, or only later in the learning process. In Experiment 1, we showed that new word learning comes at the cost of subsequent retrieval ease of words in another foreign language: while participants were still able to recall English words after learning their Spanish translation equivalents, the expected speed-up in naming latencies from pre- to post-test was smaller for those words compared with words for which they had not learned Spanish translations. The goal of Experiment 2 was two-fold: first, we aimed at replicating the result from Experiment 1 while removing the English pre-test from the Spanish learning by a day to eliminate the possibility of artificially facilitating the emergence of interference effects. Second, we asked whether interference effects would become stronger (and hence also be found in accuracy) if we allowed the newly learned words time to consolidate overnight.

We indeed replicated the naming speed effect from Experiment 1, thus reinforcing the finding that learning words from a new foreign language is detrimental for subsequent retrieval ease of words in other foreign languages. Interestingly, unlike in Experiment 1, we also observed a significant interference effect in naming accuracy in Experiment 2, suggesting that learning a new language can under some circumstances actually make you forget corresponding words in another foreign language. These

interference effects, however, were *not* significantly modulated by consolidation, and the difference in accuracy main effects between experiments, in fact, proved insignificant as well. A statistical model with both experiments together showed overall main effects of interference in accuracy and naming latencies in the absence of significant interactions between groups. Hence, we conclude that learning a new language does hamper subsequent retrieval ease in other foreign languages. Moreover, it appears that these effects can be observed immediately after learning, and do not need time (or overnight consolidation) to emerge, although they were descriptively smaller when no time for such consolidation was given (in Experiment 1).

### *Between-language interference as a cause of attrition*

In general, our findings fit in well with previous research on interference-based forgetting: engaging with new learning materials affects the ability to retrieve materials learned earlier. However, the present study extends this finding to foreign language learning and thus contributes to our knowledge of foreign language attrition. First and foremost, we show interference not from the *use* of other foreign languages, but instead from the *new learning* of foreign language material. In doing so, we provide the first empirical evidence for the anecdotal reports mentioned in the Introduction, namely that learning a new language negatively affects the retrieval of previously learned foreign languages.

Second, the present experiments demonstrate new learning-induced retrieval difficulty for *well-consolidated* foreign language vocabulary, that is vocabulary learned long ago. Showing interference effects for old, semantic memory content is arguably more difficult and hence a more convincing proof of the role of between-language competition in language forgetting than comparable effects on just recently acquired vocabulary (e.g., Bailey & Newman, 2018; Mickan et al., 2020, 2021). English is a language that our participants started learning a long time ago and which they have achieved considerable proficiency in. It is thus quite remarkable that a Spanish learning session of just 1 hr on average made some participants entirely unable to retrieve some of the corresponding English translations.

Admittedly, the accuracy effect is numerically very small: though statistically robust, it amounts to a difference between interference conditions of on average less than one word (out of 23). It would be interesting for future research to test whether a longer learning session, spread out over days or weeks, allowing not only for consolidation but also for repeated rehearsal of the new language, would show (numerically) larger interference effects in the lab. Such an experimental setup would also be closer to real-life foreign language learning.

The combination of the above mentioned two design aspects (inducing interference through *new learning* and on L2 vocabulary *learned long ago*) is what makes the present experiments novel and different from previous research on language forgetting. Mickan et al. (2020, 2021) and Bailey and Newman (2018) showed forgetting effects on newly learned foreign language words induced via retrieval practice of translation equivalents. Isurin and McDonald (2001) were also interested in how new learning affects a previously learned foreign language: they had participants first learn a list of English-Hebrew translation pairs, followed by a list of English-Russian word pairs (or vice versa). While they did find evidence for interference of learning the second list on subsequent retrieval success for the first learned list, the two languages in their experiment were learned in immediate succession. Their results thus demonstrate RI from new learning on still episodic memories. Levy et al. (2007), in turn, showed that retrieval of L2 English words impaired subsequent recall of L1 translation equivalents. Words in the participants' mother tongue were—like the English words in our study—known long before the experimental session. However, Levy and colleagues probed their participants' memory of L1 words in a rather indirect way (testing whether these previously named words were produced as response to a cue in a subsequent rhyme generation or a semantic or phonological relatedness generation task) which might have underestimated participants' actual L1 knowledge. Moreover, critically, their study differs from ours in that their interference phase consisted of retrieval of known words rather than new learning.

New word learning has also been shown to make old knowledge less accessible, e.g., in terms of processing times. Elias and Degani (2022) showed that learning the Arabic meanings of Hebrew-Arabic interlingual homophones slowed down the subsequent recognition of the Hebrew meanings of the same words in Hebrew speakers, as measured in a lexical decision task. Maciejewski et al. (2020) and Fang and Perfetti (2019) let participants learn new meanings for existing L1 words across several days, and found that this learning rendered the original meanings temporarily less accessible. Note however that all these studies are concerned with learning a new meaning for an already known form, and not, as in our case, a new form (translation) for a given meaning. Altogether, however, our results add to a slowly emerging body of literature that shows the interference of new experiences even with well-consolidated old knowledge.

### *Retroactive interference and foreign language attrition*

Apart from informing research in the language domain, our study also adds to the memory literature on retroactive interference (RI) effects. As mentioned in the Discussion



of Experiment 1, it is not common for traditional RI studies to report naming latencies. Quite in contrast to those studies, most robust evidence for RI in our experiments comes from naming latency effects. Although these effects might seem more subtle than effects on retrieval ability, we think they provide just as much evidence for interference processes. Especially in a design where participants do not have a response time limit, and maybe especially when looking at old, well consolidated semantic knowledge rather than “fresh” episodic memory traces, retrieval difficulties are likely to surface in naming latencies first. Of course, this does not necessarily mean that every single item with slow latencies will later be forgotten; but, in reverse, every forgotten item will have displayed long naming latencies at an earlier stage. This means that naming latency can be considered a proxy for “forgetting risk.” Without measuring response latencies in Experiment 1, for example, we would have erroneously concluded that learning a new language does not impact previously learned languages. Thus, RTs clearly provide a more nuanced picture of RI effects than accuracy.

Another important implication of the present results concerns the origin of RI effects. Remember that in a classical RI experiment, after learning two lists in immediate succession, retrieval of items on list 1 is impeded at final test, compared with a condition in which no second list is learned. It has commonly been assumed that this happens because list 2 learning interferes with the consolidation of list 1 material, essentially overwriting the latter (Dewar et al., 2007; Müller & Pilzecker, 1900; Wixted, 2004). However, in our case, English words were learned long ago, and so learning Spanish (list 2) should have no effect if RI was merely due to an interruption of the consolidation process of list 1 (i.e., English) items. Thus, the interference effects that we report clearly challenge the consolidation account of RI, and are at odds with studies that have failed to find RI effects on consolidated material (e.g., Ellenbogen et al., 2006; Sheth et al., 2012).

Interestingly though, we are not the only ones to report RI-like effects with consolidated material (e.g., Houston, 1967; McGeoch & Nolen, 1933; see Wixted, 2004, for an overview, and Pöhlchen et al., 2021, and Bailes et al., 2020, for failed replication attempts of Ellenbogen et al., 2006). Such findings are better accounted for by a competing view of RI which has been called “cue-overload” or “transfer theory” (McGeoch & McDonald, 1931; Watkins & Watkins, 1976; Wixted, 2004). According to that account, items from list 1 and 2 that have been associated to the same cues compete for activation upon presentation of these cues during retrieval, with list 2 items having a recency advantage over list 1 items, thereby (partially) blocking the original list 1 associations. As Skaggs (1933) already pointed out, both accounts may actually hold to some extent, with both cue competition and consolidation blocking partially responsible for RI effects, even though

Wixted (2004) attributes by far the most “daily life” relevance to consolidation blocking. However, only the competition account can explain RI effects on consolidated material.

The competition during retrieval account is in fact very similar to how Mickan et al. (2020) as well as other lab-based language attrition studies explain their findings (Bailey & Newman, 2018; Levy et al., 2007), and for which Mickan et al. (2021), for the situation of new L3 word learning impaired by L2 translation activation, have found electrophysiological evidence in the form of N2 effects. Just like items in the two lists in RI studies are connected to one another via a shared cue, so are the English and Spanish words in our study connected to one another via their shared concept. At final test, the Spanish words, having been retrieved more recently for a given concept, have a competition advantage over their English competitors and hence make it harder for participants to recall the English words, compared with English words for which no Spanish translations were learned. In a nutshell, while it is true that consolidated knowledge, such as the English words in our study, is more difficult to interfere with than unconsolidated material, it is not immune to interference.

Memory researchers continue to discuss the factors that determine whether RI effects do or do not occur for consolidated materials, like, for instance, the timing of the final (list 1) test relative to the last (list 2) learning (for an overview see Wixted, 2004). However, the consolidation status of the interfering, rather than the originally learned, material, is rarely systematically varied. Wixted (2004) reports that across studies (e.g., Postman & Alper, 1946; Sisson, 1939), findings seem to point at larger interference effects when the time interval between list 2 learning and list 1 test is short. Different to this observation though, separating Spanish (“list2”) learning and English (“list 1”) post-test by a night of sleep (Experiment 3) made no significant difference for—and if anything, increased rather than diminished—the magnitude of interference effects in our study. It is possible that the stimulus materials used in the present study, that is vocabulary lists that are meaningful also outside of the context of the experiment and that have been learned and consolidated a long time ago, adhere to slightly different principles than the rather artificial materials tested in such memory studies on RI.<sup>5</sup> However, much more research seems necessary to systematically assess the role of consolidation of interfering material on previously learned (general or vocabulary) knowledge.

### *The role of consolidation in lexical competition*

Assuming that the interference effects that we observed are based on competition between translation equivalents in different languages, and knowing that some competition effects (specifically, lexical competition in spoken word recognition) have been shown to depend on the integration

of newly learned words into the existing mental lexicon (Davis & Gaskell, 2009), one might expect interference effects to increase after the newly learned Spanish words have had the chance to consolidate. We addressed this possibility in Experiment 2, yet were unable to confirm this hypothesis. The interference effect in accuracy was numerically larger in the consolidation group and was in fact only statistically reliable in that group (as tested in post hoc models for each group separately; consolidation group:  $\beta = -1.02$ ,  $z = -2.63$ ,  $p(\chi^2) = .008$ ; no-consolidation group:  $\beta = -0.53$ ,  $z = -1.64$ ,  $p(\chi^2) = .101$ ), but the difference in interference magnitude between the two groups was not big enough and hence did not reach statistical significance. Given the current pattern of results, we thus conclude that consolidation of the interfering knowledge does not significantly increase interference strength and that the interference effects that we observed, both in naming latencies and accuracy (in Experiment 2), emerge immediately after or during learning, and do not appear to need time to evolve.

As with any null result, the question remains whether consolidation really does not affect interference strength, or whether—especially in the light of descriptive differences—we simply failed to detect the difference in this study setup. From research on novel word consolidation, it appears that while lexical integration starts immediately during or after learning (e.g., Lindsay & Gaskell, 2013), it can take multiple days or even weeks to complete (Takashima et al., 2006). It is, thus, possible that a consolidation time window of just one night was too short to result in large enough differences in interference magnitude between consolidated and unconsolidated Spanish words. Conversely, it may be the case that the Spanish words were very easy and fast to integrate into the mental lexicon given that they refer to already existing concepts. Maybe vocabulary learning then resembles schema-consistent learning via fast (cortical) mapping (e.g., Coutanche & Thompson-Schill, 2014; van Kesteren et al., 2010). If this is the case, new foreign language vocabulary might not need the slow consolidation process that is otherwise required and thus would not benefit much from overnight consolidation.

Finally, we also have to consider the possibility that the interference effects we observe are not actually caused by lexical competition at final test, but rather via inhibitory dynamics during the learning of the Spanish words. The long-known English words might have been interfering with the acquisition of their Spanish translation equivalents and hence might have been suppressed to allow for efficient learning of the Spanish words (see Anderson, 2003, for a more in-depth discussion of inhibition as a mechanism driving forgetting). Quite possibly then, this lasting inhibition may have hindered recall of these English words, compared with English words which did not need to be suppressed because no Spanish translations were learned. If our effects are indeed caused by inhibition *during* learning, consolidation *after* the learning phase would indeed have

little extra effect and hence explain why we failed to find a difference between the two groups in Experiment 2. Whatever the explanation, we encourage future research into the role of consolidation for both “new” and “old” language words in foreign language attrition.

### Reconsolidation effects

According to previous reports in the memory literature, retrieval of old memories may make these memories more vulnerable to interference from new learning, unless the old memories are reconsolidated (Hupbach et al., 2007; Stickgold & Walker, 2005). Due to this observation, we had hypothesised to observe a smaller interference effect in the no-consolidation group of Experiment 2—who took the English test a day before learning the Spanish words—than in Experiment 1, where both phases had been administered on the same day. However, our data on the comparison of these two groups showed that there were no differences. In naming latencies, the two groups showed comparable interference effects. In naming accuracy, neither of the two groups showed an effect; if anything, the interference effect was numerically larger, rather than smaller, in the no-consolidation group than in the group from Experiment 1.

While the result of stable interference effects, independent of the time interval between English pre-test and Spanish learning, is reassuring, one may wonder why we did not observe even the slightest indication for such destabilisation / reconsolidation effects. In humans, many studies on reconsolidation have focussed on procedural memory (e.g., Hardwicke et al., 2016; Walker et al., 2003) and fear memory (e.g., Kindt & Soeter, 2013; Schiller et al., 2010, see Agren, 2014 for a review). Whether reactivation of a *declarative* memory also makes it labile and in need of subsequent reconsolidation is less clear (compare Forcato et al., 2009; Hupbach et al., 2007; Strange et al., 2010; Wichert et al., 2011). In fact, the study most comparable in setup and materials to ours did not find evidence for such a testing-induced destabilisation either: Potts and Shanks (2012) observed that learning English-Finnish translations a day after learning English-Swahili translation pairs lead to RI (from Finnish to Swahili). However, these effect disappeared when Day 2 (the Finnish learning day) started with another test of the Swahili words. Thus, retrieving the previously learned knowledge just before new learning seemed to protect, rather than destabilise, that old knowledge. The authors attribute their finding to the fact that their “reminder test” required active retrieval of the English-Swahili word pairs, as opposed to just restudying them passively (as is the case in most other studies on reconsolidation), and speculate that under these circumstances, testing is beneficial rather than detrimental (in line with the “testing effect,” see Antony et al., 2017).

Even though none of the groups in our studies differed from one another in statistical terms, the pattern of results is compatible with that reported by Potts and Shanks

(2012). The English pre-test was an active retrieval test, and, numerically speaking, interference effects in accuracy rates were smallest in Experiment 1, where the English pre-test and the Spanish learning took place in immediate succession. Separating the English pre-test from the Spanish learning phase in Experiment 2 numerically increased the interference effect. Destabilisation and the need for reconsolidation thus do not seem to be an issue during L2 word retrieval in situations like ours.

## Conclusion

This study provides the first empirical evidence of the detrimental effects that learning words from a new language can have on remembering words from previously learned foreign languages. Multilinguals are thus not wrong in their perception that adding a language to their repertoire will, at least during the first stages of learning the new language, hamper access to other foreign languages, even when those languages were learned a long time ago and to a high level of proficiency. Our results, furthermore, suggest that these effects emerge immediately, and do not critically depend on consolidation time for the newly learned language words.

## Acknowledgements

We would like to thank Panthea Nemat and Cami Geenen for help with data acquisition and coding for Experiment 1.

## Data accessibility statement



The data and materials from the present experiment are publicly available at the Open Science Framework website: <https://doi.org/10.34973/x2dw-jr34>.

## Declaration of conflicting interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

## Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This work was supported by an International Max-Planck Research School for Language Sciences PhD fellowship awarded to A.M. (grant period: 2016–2020).

## ORCID iD

Kristin Lemhöfer  <https://orcid.org/0000-0003-0430-3968>

## Supplemental material

The supplementary material is available at [qjep.sagepub.com](http://qjep.sagepub.com).

## Notes

1. When this participant was excluded in the analysis, there were no changes at all regarding the significance levels of the effects, with only minor changes in the parameters of the model outcome.
2. There was one word item in the LexTALE test (turtle) that appeared as a reserve item in Exp. 2.
3. We used difference scores for the analysis because we conducted the analysis of naming latencies in parallel to the accuracy analysis. In the latter, the dependent variable (DV) was automatically a “difference score,” considering that accuracy at pre-test was by definition 0% (i.e., accuracy at post-test was used as DV).
4. This coding matrix may seem counterintuitive, but is correct according to the logic of forward contrast coding, see <https://stats.oarc.ucla.edu/r/library/r-library-contrast-coding-systems-for-categorical-variables/#forward> and [https://bookdown.org/pingapang9/linear\\_models\\_bookdown/contrasts.html#connection-between-contrast-and-coding-schemes](https://bookdown.org/pingapang9/linear_models_bookdown/contrasts.html#connection-between-contrast-and-coding-schemes).
5. Even consolidated list 1 material in RI studies always refers to material learned within the context of the experiment, one (or at most a few days) before interference induction, but never, at least to our knowledge, to material learned years before, outside of the experimental context.

## References

- Agren, T. (2014). Human reconsolidation: A reactivation and update. *Brain Research Bulletin*, *105*, 70–82. <https://doi.org/10.1016/j.brainresbull.2013.12.010>
- Anderson, M. C. (2003). Rethinking interference theory: Executive control and the mechanisms of forgetting. *Journal of Memory and Language*, *49*, 415–445. <https://doi.org/10.1016/j.jml.2003.08.006>
- Antony, J. W., Ferreira, C. S., Norman, K. A., & Wimber, M. (2017). Retrieval as a fast route to memory consolidation. *Trends in Cognitive Sciences*, *21*, 573–576. <https://doi.org/10.1016/j.tics.2017.05.001>
- Baayen, R. H., Piepenbrock, R., & Gulikers, L. (1995). *The CELEX lexical database* [Webcelex]. Linguistic Data Consortium, University of Pennsylvania.
- Bailes, C., Caldwell, M., Wamsley, E. J., & Tucker, M. A. (2020). Does sleep protect memories against interference? A failure to replicate. *PLOS ONE*, *15*(2), Article e0220419. <https://doi.org/10.1371/journal.pone.0220419>
- Bailey, L., & Newman, A. J. (2018). Retrieval-induced forgetting and second language vocabulary acquisition: Insights from a Welsh-language training study. In J. M. Fawcett (Ed.), *Control processes in human memory: The role of retrieval suppression and retrieval practice*. Symposium conducted at the joint meeting of the Canadian Society for Brain, Behaviour and Cognitive Science and the Experimental Psychology Society, St. John’s, CA.
- Bakker, I., Takashima, A., van Hell, J. G., Janzen, G., & McQueen, J. M. (2014). Competition from unseen or unheard novel words: Lexical consolidation across modalities. *Journal of Memory and Language*, *73*, 116–130. <https://doi.org/10.1016/j.jml.2014.03.002>
- Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it

- maximal. *Journal of Memory and Language*, *68*, 255–278. <https://doi.org/10.1016/j.jml.2012.11.001>
- Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). *lme4: Linear mixed-effects models using Eigen and S4* (Version 1.1-15) [R package]. <https://cran.r-project.org/web/packages/lme4/index.html>
- Bellana, B., Mansour, R., Ladyka-Wojcik, N., Grady, C. L., & Moscovitch, M. (2021). The influence of prior knowledge on the formation of detailed and durable memories. *Journal of Memory and Language*, *121*, 104264. <https://doi.org/10.1016/j.jml.2021.104264>
- Boersma, P. (2001). Praat, a system for doing phonetics by computer. *Glott International*, *5*(9–10), 341–345.
- Borovsky, A., Elman, J. L., & Kutas, M. (2012). Once is enough: N400 indexes semantic integration of novel word meanings from a single exposure in context. *Language Learning and Development*, *8*, 278–302. <https://doi.org/10.1080/15475441.2011.614893>
- Bowers, J. S., Davis, C. J., & Hanley, D. A. (2005). Interfering neighbours: The impact of novel word learning on the identification of visually similar words. *Cognition*, *97*, B45–B54. <https://doi.org/10.1016/j.cognition.2005.02.002>
- Brodeur, M. B., Dionne-Dostie, E., Montreuil, T., & Lepage, M. (2010). The bank of standardized stimuli (BOSS), a new set of 480 normative photos of objects to be used as visual stimuli in cognitive research. *PLOS ONE*, *5*(5), Article e10773. <https://doi.org/10.1371/journal.pone.0010773>
- Costa, A., & Santesteban, M. (2004). Lexical access in bilingual speech production: Evidence from language switching in highly proficient bilinguals and L2 learners. *Journal of Memory and Language*, *50*, 491–511. <https://doi.org/10.1016/j.jml.2004.02.002>
- Coutanche, M. N., & Thompson-Schill, S. L. (2014). Fast mapping rapidly integrates information into existing memory networks. *Journal of Experimental Psychology: General*, *143*, 2296–2303. <https://doi.org/10.1037/xge0000020>
- Davis, M. H., & Gaskell, M. G. (2009). A complementary systems account of word learning: Neural and behavioural evidence. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, *364*, 3773–3800. <https://doi.org/10.1098/rstb.2009.0111>
- Dewar, M. T., Alber, J., Butler, C., Cowan, N., & Della Sala, S. (2012). Brief wakeful resting boosts new memories over the long term. *Psychological Science*, *23*(9), 955–960. <https://doi.org/10.1177/0956797612441220>
- Dewar, M. T., Cowan, N., & Sala, S. D. (2007). Forgetting due to retroactive interference: A fusion of Müller and Pilzecker's (1900) early insights into everyday forgetting and recent research on anterograde amnesia. *Cortex*, *43*(5), 616–634. [https://doi.org/10.1016/S0010-9452\(08\)70492-1](https://doi.org/10.1016/S0010-9452(08)70492-1)
- Dumay, N., & Gaskell, M. G. (2007). Sleep-associated changes in the mental representation of spoken words. *Psychological Science*, *18*, 35–39. <https://doi.org/10.1111/j.1467-9280.2007.01845.x>
- Elias, M., & Degani, T. (2022). Cross-language interactions during novel word learning: The contribution of form similarity and participant characteristics. *Bilingualism: Language and Cognition*, *25*, 548–565. <https://doi.org/10.1017/S1366728921000857>
- Ellenbogen, J. M., Hulbert, J. C., Stickgold, R., Dinges, D. F., & Thompson-Schill, S. L. (2006). Interfering with theories of sleep and memory: Sleep, declarative memory, and associative interference. *Current Biology*, *16*, 1290–1294. <https://doi.org/10.1016/j.cub.2006.05.024>
- Fang, X., & Perfetti, C. A. (2019). Learning new meanings for known words: Perturbation of original meanings and retention of new meanings. *Memory & Cognition*, *47*, 130–144. <https://doi.org/10.3758/s13421-018-0855-z>
- Forcato, C., Argibay, P. F., Pedreira, M. E., & Maldonado, H. (2009). Human reconsolidation does not always occur when a memory is retrieved: The relevance of the reminder structure. *Neurobiology of Learning and Memory*, *91*, 50–57. <https://doi.org/10.1016/j.nlm.2008.09.011>
- Gollan, T. H., & Silverberg, N. B. (2001). Tip-of-the-tongue states in Hebrew–English bilinguals. *Bilingualism: Language and Cognition*, *4*, 63–83. <https://doi.org/10.1017/S136672890100013X>
- Hardwicke, T. E., Taqi, M., & Shanks, D. R. (2016). Postretrieval new learning does not reliably induce human memory updating via reconsolidation. *Proceedings of the National Academy of Sciences of the United States of America*, *113*, 5206–5211. <https://doi.org/10.1073/pnas.1601440113>
- Houston, J. P. (1967). Retroactive inhibition and point of interpolation. *Journal of Verbal Learning and Verbal Behavior*, *6*, 84–88.
- Hupbach, A., Gomez, R., Hardt, O., & Nadel, L. (2007). Reconsolidation of episodic memories: A subtle reminder triggers integration of new information. *Learning & Memory*, *14*, 47–53. <https://doi.org/10.1101/lm.365707>
- Inostroza, M., & Born, J. (2013). Sleep for preserving and transforming episodic memory. *Annual Review of Neuroscience*, *36*(1), 79–102. <https://doi.org/10.1146/annurev-neuro-062012-170429>
- Isurin, L., & McDonald, J. L. (2001). Retroactive interference from translation equivalents: Implications for first language forgetting. *Memory & Cognition*, *29*, 312–319. <https://doi.org/10.3758/BF03194925>
- Kang, S. H. K., Gollan, T. H., & Pashler, H. (2013). Don't just repeat after me: Retrieval practice is better than imitation for foreign vocabulary learning. *Psychonomic Bulletin & Review*, *20*, 1259–1265. <https://doi.org/10.3758/s13423-013-0450-z>
- Kindt, M., & Soeter, M. (2013). Reconsolidation in a human fear conditioning study: A test of extinction as updating mechanism. *Biological Psychology*, *92*, 43–50. <https://doi.org/10.1016/j.biopsycho.2011.09.016>
- Kroll, J. F., Bobb, S. C., Misra, M., & Guo, T. (2008). Language selection in bilingual speech: Evidence for inhibitory processes. *Acta Psychologica*, *128*, 416–430. <https://doi.org/10.1016/j.actpsy.2008.02.001>
- Landauer, T. K. (1974). Consolidation in human memory: Retrograde amnesic effects of confusable items in paired-associate learning. *Journal of Verbal Learning and Verbal Behavior*, *13*, 45–53. [https://doi.org/10.1016/S0022-5371\(74\)80029-0](https://doi.org/10.1016/S0022-5371(74)80029-0)
- Lemhöfer, K., & Broersma, M. (2012). Introducing LexTALE: A quick and valid lexical test for advanced learners of English. *Behavior Research Methods*, *44*, 325–343. <https://doi.org/10.3758/s13428-011-0146-0>
- Levenshtein, V. I. (1966). Binary codes capable of correcting deletions, insertions, and reversals. *Soviet Physics Doklady*, *10*, 707–710.
- Levy, B. J., McVeigh, N. D., Marful, A., & Anderson, M. C. (2007). Inhibiting your native language: The role of retrieval-induced forgetting during second language

- acquisition. *Psychological Science*, 18, 29–34. <https://doi.org/10.1111/j.1467-9280.2007.01844.x>
- Linck, J. A., Schwieter, J. W., & Sunderman, G. (2012). Inhibitory control predicts language switching performance in trilingual speech production. *Bilingualism: Language and Cognition*, 15, 651–662. <https://doi.org/10.1017/S136672891100054X>
- Lindsay, S., & Gaskell, M. G. (2013). Lexical integration of novel words without sleep. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 39, 608–622. <https://doi.org/10.1037/a0029243>
- Maciejewski, G., Rodd, J. M., Mon-Williams, M., & Klepousniotou, E. (2020). The cost of learning new meanings for familiar words. *Language, Cognition and Neuroscience*, 35, 188–210. <https://doi.org/10.1080/23273798.2019.1642500>
- Mandera, P., Keuleers, E., & Brysbaert, M. (2017). Explaining human performance in psycholinguistic tasks with models of semantic similarity based on prediction and counting: A review and empirical validation. *Journal of Memory and Language*, 92, 57–78. <https://doi.org/10.1016/j.jml.2016.04.001>
- McGeoch, J. A., & McDonald, W. T. (1931). Meaningful relation and retroactive inhibition. *American Journal of Psychology*, 43, 579–588. <https://doi.org/10.2307/1415159>
- McGeoch, J. A., & Nolen, M. E. (1933). Studies in retroactive inhibition. IV. Temporal point of interpolation and degree of retroactive inhibition. *Journal of Comparative Psychology*, 15, 407–417. <https://doi.org/10.1037/h0072751>
- Mickan, A., McQueen, J. M., & Lemhöfer, K. (2020). Between-language competition as a driving force in foreign language attrition. *Cognition*, 198, 104218. <https://doi.org/10.1016/j.cognition.2020.104218>
- Mickan, A., McQueen, J. M., Valentini, B., Piai, V., & Lemhöfer, K. (2021). Electrophysiological evidence for cross-language interference in foreign-language attrition. *Neuropsychologia*, 155, 107795. <https://doi.org/10.1016/j.neuropsychologia.2021.107795>
- Müller, G. E., & Pilzecker, A. (1900). Experimentelle Beiträge zur Lehre vom Gedächtnis [Experimental Contributions to the Theory of Memory]. *Zeitschrift für Psychologie: Ergänzungsband*, 1, 1–300.
- Nadel, L., Hupbach, A., Gomez, R., & Newman-Smith, K. (2012). Memory formation, consolidation and transformation. *Neuroscience & Biobehavioral Reviews*, 36(7), 1640–1645. <https://doi.org/10.1016/j.neubiorev.2012.03.001>
- Osgood, C. E. (1948). An investigation into the causes of retroactive interference. *Journal of Experimental Psychology*, 38, 132–154. <https://doi.org/10.1037/h0055753>
- Pöhlchen, D., Pawlizki, A., Gais, S., & Schönauer, M. (2021). Evidence against a large effect of sleep in protecting verbal memories from interference. *Journal of Sleep Research*, 30, e13042. <https://doi.org/10.1111/jsr.13042>
- Postman, L., & Alper, T. G. (1946). Retroactive inhibition as a function of the time of interpolation of the inhibitor between learning and recall. *American Journal of Psychology*, 59, 439–449. <https://doi.org/10.2307/1417613>
- Postman, L., & Kaplan, H. L. (1947). Reaction time as a measure of retroactive inhibition. *Journal of Experimental Psychology*, 37, 136–145. <https://doi.org/10.1037/h0055703>
- Potts, R., & Shanks, D. R. (2012). Can testing immunize memories against interference? *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 38, 1780–1785. <https://doi.org/10.1037/a0028218>
- R Core Team (2019). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. <https://www.R-project.org/>.
- Schiller, D., Monfils, M.-H., Raio, C. M., Johnson, D. C., LeDoux, J. E., & Phelps, E. A. (2010). Preventing the return of fear in humans using reconsolidation update mechanisms. *Nature*, 463, 49–53. <https://doi.org/10.1038/nature08637>
- Sheth, B. R., Varghese, R., & Thuy Truong, B. S. (2012). Sleep shelters verbal memory from different kinds of interference. *Sleep*, 35, 985–996. <https://doi.org/10.5665/sleep.1966>
- Sisson, E. D. (1939). Retroactive inhibition: The temporal position of interpolated activity. *Journal of Experimental Psychology*, 25, 228–233.
- Skaggs, E. B. (1933). A discussion on the temporal point of interpolation and degree of retroactive inhibition. *Journal of Comparative Psychology*, 16, 411–414.
- Squire, L. R., Genzel, L., Wixted, J. T., & Morris, R. G. (2015). Memory consolidation. *Cold Spring Harbor Perspectives in Biology*, 7(8), a021766. <https://doi.org/10.1101/cshperspect.a021766>
- Stickgold, R., & Walker, M. P. (2005). Memory consolidation and reconsolidation: What is the role of sleep? *Trends in Neurosciences*, 28, 408–415. <https://doi.org/10.1016/j.tins.2005.06.004>
- Strange, B. A., Kroes, M. C., Fan, J., & Dolan, R. J. (2010). Emotion causes targeted forgetting of established memories. *Frontiers in Behavioral Neuroscience*, 4, Article 175. <https://doi.org/10.3389/fnbeh.2010.00175>
- Takashima, A., Petersson, K. M., Rutters, F., Tendolkar, I., Jensen, O., Zwartz, M. J., & McNaughton, B. L. (2006). Declarative memory consolidation in humans: A prospective functional magnetic resonance imaging study. *Proceedings of the National Academy of Sciences of the United States of America*, 103, 756–761. <https://doi.org/10.1073/pnas.0507774103>
- van Kesteren, M. T. R., Rijpkema, M., Ruiter, D. J., & Fernandez, G. (2010). Retrieval of associative information congruent with prior knowledge is related to increased medial prefrontal activity and connectivity. *The Journal of Neuroscience*, 30, 15888–15894. <https://doi.org/10.1523/JNEUROSCI.2674-10.2010>
- Walker, M. P., Brakefield, T., Allan Hobson, J., & Stickgold, R. (2003). Dissociable stages of human memory consolidation and reconsolidation. *Nature*, 425, 616–620. <https://doi.org/10.1038/nature01930>
- Watkins, M. J., & Watkins, O. C. (1976). Cue-overload theory and the method of interpolated attributes. *Bulletin of the Psychonomic Society*, 7, 289–291.
- Wichert, S., Wolf, O. T., & Schwabe, L. (2011). Reactivation, interference, and reconsolidation: Are recent and remote memories likewise susceptible? *Behavioral Neuroscience*, 125, 699–704. <https://doi.org/10.1037/a0025235>
- Wixted, J. T. (2004). The psychology and neuroscience of forgetting. *Annual Review of Psychology*, 55, 235–269. <https://doi.org/10.1146/annurev.psych.55.090902.141555>
- Zhang, Q., Popov, V., Koch, G. E., Calloway, R. C., & Coutanche, M. N. (2018). Fast memory integration facilitated by schema consistency. *bioRxiv*. Advance online publication. <https://doi.org/10.1101/253393>