Psychological Test and Assessment Modeling, Volume 59, 2017 (4), 405-441

Deterioration and recovery in verbal recall: Repetition helps against pro-active interference

Chris Lange-Kuettner[1], Monika Markowska[2] & Ridhi Kochhar[2]

Abstract

The current study tests whether memory deterioration due to pro-active interference (PI) in verbal recall could be halted via block repetition potentially leading to an increased memory consolidation. We also tested whether bilinguals would be better shielded against memory deterioration than monolinguals because they constantly need to enrich their vocabulary to compensate for their smaller lexica in either language. We tested monolinguals and balanced bilinguals with an N-Back and a free verbal recall task. Repetition showed a significant main effect with a large effect size. In Study 1 (N=45), monolingual men showed less improvement in the repetition blocks, while bilingual men showed a significant doubling of their word recall on each repetition. In Study 2 (N=78), monolingual women were less likely to use the repetition opportunity to improve the word score. Thus, in both studies, a significant monolingual disadvantage showed. When the two data sets were merged (N=123), statistical effects showed that the single word list repetition had successfully and significantly increased resistance to PI, but all individual differences due to bilingualism and sex had disappeared. This supported a previous meta-analysis showing that a monolingual disadvantage does not hold in large samples with $N > 100$ (Paap effect).

Key words: memory deterioration, bilingualism, free verbal recall, proactive inhibition, rehearsal, memory consolidation

[1] *Correspondence concerning this article should be addressed to:* Chris Lange-Kuettner, PhD, London Metropolitan University, Tower Building T6-20, 166–220 Holloway Road, London N7 8DB, United Kingdom; email: c.langekuettner@londonmet.ac.uk

[2] London Metropolitan University, United Kingdom

Introduction

Proactive-interference (PI) is a longstanding research topic investigated in cognitive psychology. It consists of a memory decline during a memory experiment because items from previous lists intrude into the memorization of a current list (Underwood, 1945). PI also occurs in animals such as rats (Cohen & Armstrong, 1996; Dunnett & Martel, 1990; Dunnett, Martel, & Iversen, 1990) and pigeons (Thomas, Burr, & Vogt, 1982; Wilkie, 1986). As such, PI gives important clues to the nature of memory deterioration. The degree of pro-active interference depends on the amount of items that are recalled and the length of the delay before retrieval (Keppel & Underwood, 1962). PI can be prevented or reduced by changing the categorical membership of the memory items (Wickens, 1970) and hence their long-term memory representations are involved (Oberauer, Awh, & Sutterer, 2016). Pro-active interference (as well as retro-active interference) was found to be reduced after sleep which plays a role in memory consolidation (Abel & Bäuml, 2014). However, it is also possible to show resistance against PI by blocking out intrusive thoughts and having a good reading span (Friedman & Miyake, 2004). By allowing more time for encoding a word, PI is also reduced (Loess & Waugh, 1967). When more than two minutes were allowed for encoding, PI was negligible. Accordingly, because young children are slow information processors (Lange-Küttner, 2012), the extent of PI decreases as their cognition becomes more efficient with age (Kail, 2002), and increases again with aging (Carretti, Mammarella, & Borella, 2012).

The current study tests the rehearsal hypothesis that repetition of the words counter-acts pro-active interference in verbal recall (Schendel, 1976). Resistance against PI builds up especially with repeated testing as this prevents intrusions (Szpunar, McDermott, & Roediger, 2008, 2009; Wahlheim, 2015). Each memory list was once repeatedly assessed in order to foster memory consolidation. We predicted that while word memory would deteriorate, verbal recall would recover on each repetition, and could eventually halt memory deterioration during the experiment. We also expected that bilinguals would tend to be better in resisting PI (Bialystok & Feng, 2009; Dillon, McCormack, Petrusic, Cook, & Lafleur, 1973). Bilinguals' switching between languages trains their executive function which in turn is a protective factor against aging and even dementia (Bialystok, Craik, & Freedman, 2007; Woumans et al., 2015). However, we tested young adults where no aging effect could be expected. Instead, we hypothesized that bilingual young adults may be better in resisting PI because of their constant practice of upgrading their vocabulary in their languages.

Proactive interference (PI), rehearsal and repetition

Previous research found a release from PI via a categorical change of the stimulus items (Wickens, Born, & Allen, 1963) which indicates that people build up a pool of items across experimental blocks and subsequently confuse the items in this pool during retrieval. Hence, changing to a new pool (semantic item category) solves the problem as the previous items are neatly gathered elsewhere. In fact, this spatial metaphor seems to be more than just a metaphorical image that we use to illustrate the mental classification process because even changing the size of the display area produced release from PI (Turvey & Egan, 1969) and could improve memory performance (Lange-Küttner, 2013).

Already in very early studies, it was suspected that rehearsal of memory items may be more important for release from PI – and thus prevent memory deterioration – than the novelty of a different class of memory items (Reutener, 1972; Schendel, 1976). The early rehearsal hypothesis is in accordance with the working memory model which predicts that memory deterioration in verbal recall is prevented by rehearsal processes in the phonological loop (Baddeley, Gathercole, & Papagno, 1998). There can be covert and overt rehearsal (Cowan & Vergauwe, 2015). However, rehearsal can also be forced by asking participants to repeat experiments blocks (Lange-Küttner & Sykorova, 2015)

While the inconsistent results of the early studies were most probably due to the self-report methodology of rehearsal, more clear results were obtained from studies with enforced rehearsal by subjecting participants to tests which yielded more release from PI than just studying words (Szpunar et al., 2008, 2009; Wahlheim, 2015). These authors suggest that memory testing of word lists would promote consolidation and enhance long-term retention because well-learned material facilitates list integration (Bäuml & Kliegl, 2013; Wahlheim, 2015) as well as list discrimination (Szpunar et al., 2008) and thus counteracts PI. Bäuml and Kliegl (2013) suggest that repetition allows the narrowing down of attention towards the target and the elimination of non-relevant distracter information. Repetition in word learning is particularly beneficial both for children (Horst, 2013; Horst, Parsons, & Bryan, 2011) and in adults especially for less familiar words (Francis & Sáenz, 2007; Hernandez & Reyes, 2002; Lange-Küttner & Sykorova, 2015). Learning words via repetition is also called the Hebb repetition effect (Cumming, Page, & Norris, 2003; McKelvie, 1987; Mosse & Jarrold, 2008; Page & Norris, 2009).

That repetition appears to be an excellent way to prevent the build-up of PI was also shown in another recent study (Rahimi-Golkhandan, Maruff, Darby, & Wilson, 2012). In this study, an International Shopping List Task (ISLT) was used that provides multiple lists. The second test was the Rey Auditory Verbal Learning Test (RAVLT). Each list was repeated three times for both tests with the result that no significant PI was found for the ISLT, although the same sample did show PI on the RAVLT.

Hence, in the current study, we varied the usual testing of release from PI via a change of semantic category of the memory items to a new design which involved one repetition of each memory list. We expected that the repetition (rehearsal) would be helpful in overcoming PI because the repetition would lead to a stronger memory consolidation.

The current study

The current study adopts a quasi-experimental design insofar as the second hypothesis states that bilingual participants would show stronger resistance against PI. The bilingual advantage resp. a monolingual disadvantage is not tested with the experiment itself, but instead is a between-subjects factor (Cook & Campbell, 1979).

Bilingualism is a research area that encompasses the entire life-span, from infancy into old age (Costa & Gallés, 2014). Learning a second language increases gray matter in the brain area responsible for verbal fluency, and this effect was quantifiable in terms of years of learning as well as level of second language proficiency (Mechelli et al., 2004). Bilinguals may be used to more effortful and less automatic processing because they need to exert more executive control in order to suppress one language when they address a socially relevant monolingual language speaker (Costa & Gallés, 2014; Vaid & Genesee, 1980). A recent large meta-analysis (Adesope, Lavin, Thompson, & Ungerleider, 2010) of 63 studies with 6,022 participants showed that bilingualism was reliably associated with several positive cognitive outcomes, particularly increased attentional control and metalinguistic awareness.

Practice in multiple languages may also contribute to faster articulation in bilinguals as they map multi-lingual phonology in verbal memory (Lange-Küttner, Puiu, Nylund, Cardona, & Garnes, 2013). Faster articulation should make it possible to rehearse comparably more words in the same time (Hulme, Thomson, Muir, & Lawrence, 1984; Roodenrys, Hulme, & Brown, 1993) which was demonstrated in adults especially for shorter word lists of three to four items (Cowan & Saults, 2013).

Both verbal fluency and vocabulary size were found to be important factors (Paap et al., 2017). Verbal rehearsal may also be better developed as bilinguals are striving to improve their smaller vocabulary (Ben Zeev, 1977; Craik & Bialystok, 2010; Oller & Jarmulowicz, 2007). They constantly need to upgrade their smaller bilingual lexica which together may be as large as that of a monolingual person, but are smaller in size per language (Ben Zeev, 1977; Craik & Bialystok, 2010). Thus, there is a constant need to upgrade if bilinguals wanted to match the size of their L1 lexicon to that of a fluent monolingual. This may result in more constant improvements during word list repetitions. Thus, the current study tests whether memory deterioration due to pro-active interference (PI) in verbal recall could be halted because of increased memory consolidation via memory list repetition (rehearsal), and whether bilinguals would use the rehearsal opportunity to better effect against memory deterioration during the experiment than monolingual participants.

We also predicted sex differences in verbal recall. Firstly, men of different languages can communicate well with one another in non-verbal activities (Güvendir, 2013), for instance in games such as football. Men take more time than women before they articulate a word (Lange-Küttner et al., 2013). Thus, they may not feel the need to upgrade their foreign language vocabulary as much. Secondly, women thrive in language tests such as story recall and word fluency despite differences in genetic pool and language culture experience, while men fare better on non-verbal visual imagery tests such as mental rotation (Mann, Sasanuma, Sakuma, & Masaki, 1990). Hence we expected that a bilingualism advantage would be more likely to show in males because of their greater reliance on non-verbal strategies and communication, while females are more likely to use a verbal strategy even if no language stimuli are involved (Merrill, Yang, Roskos, & Steele, 2016). Accordingly, females' language function is less lateralized than in males (Chen et al., 2007; Friederici et al., 2008; Frith & Vargha-Khadem, 2001; Schaadt, Hesse, & Friederici, 2015), but sex differences are often not controlled (Hull & Vaid, 2007). Hence, in the current study we controlled word recall performance with respect to sex differences.

We added conditions of two non-informational eye gaze cues in the second and third block in order to maintain participants' attention and to prevent ceiling effects in the repetition blocks. Driver et al. (1999) found that even uninformative eye gaze cues elicit orienting towards one side. In a static condition, a central eye gaze cue appeared shortly before the test word (pre-cue) and kept looking to the same side when the target and distracter pair appeared, while in a dynamic condition, the eye gaze cue looked to the other side (see also Pfeiffer, Vogeley, & Schilbach, 2013). We used these two eye gaze conditions because we reckoned that the dynamic display would impose an even stronger cognitive load (Lavie, 2005; Lavie, Hirst, de Fockert, & Viding, 2004).

We also controlled for individual differences in selective attention and working memory between the two language groups with an N-back task. This task involves keeping track of the mere appearance of a target (0-back) with selective attention, and monitoring whether it had appeared 1-, 2- or 3-back steps before which additionally involves working memory (Kane, Conway, Miura, & Colflesh, 2007).

Study 1

Method Study 1

Participants

We tested university students from a large inner-city university in London. Forty-five students from 19 to 44 years of age (monolingual *M* = 27 years, range 20-44 years; bilingual *M* = 23 years, range 19-40 years) participated, 21 monolingual speakers of English (10 males, 11 females) and 24 speakers of English and at least one or more languages (11 males, 13 females).

We tested bi- and multilingual participants with a multi-cultural background of 22 languages. The main criterion for inclusion in the bilingual group was whether participants regularly used the second language in addition to their main English language (balanced bilinguals). Languages in addition to English were Arabic, Bengali, Burmese, Cantonese, Chinese, Danish, French, German, Greek, Gujarati, Hindi, Italian, Kachin, Lithuanian, Norwegian, Spanish, Swedish, Polish, Portuguese, Punjabi, Urdu, Ygarlan and Yoruba. None of the monolingual English speakers reported actively speaking another language. Participants rated their fluency and usage of the English and the non-English language on a 3-point scale (daily, sometimes, rarely). They did not have to give a value for a possible third language. Bilinguals had for fluency an average value of *M* = 2.87 for English and *M* = 2.79 for the other language which showed that they were balanced bilinguals. For usage, bilinguals had an average value for English of *M* = 3.0, just like monolinguals, and *M* = 2.62 for the other language which is a value between daily and sometimes that shows that the second language was spoken somewhat less often than English.

The design of the study as well as consent and debrief forms were vetted and approved by the departmental Ethics Committee according to the Ethics Guidelines of the British Psychological Society. All subjects gave written informed consent in accordance with the Declaration of Helsinki.

Apparatus and procedure

The N-Back task (Schleepen & Jonkman, 2010) was programmed using the software Experimental Run Time System (ERTS) (Beringer, 1994). Stimuli were individually randomized sequences of 15 consonants (B, C, F, G, H, K, L, M, P, R, S, T, W, X, Z), presented in Times Font size 14 in white on a black background in the center of the computer monitor. The N-Back task consisted of four difficulty levels, 0-, 1-, 2-, and 3-back, see Figure 1, presented in this order; each level was tested with 60 presentation trials. This entire N-Back task was repeated once.

Stimulus duration was 500 ms with 1500 ms inter-stimulus interval (ISI). In the 0-back condition a target was defined as a letter "X". Correct responses were not dependent on the case of the letter. The 0-back level required sustained attention but no working memory processing. In the 1-, 2-, and 3-back levels, a target event was defined as a letter that was identical to the letter that appeared 1 (e.g., S-S), 2 (e.g., R-G-r), or 3 (e.g., T-s-P-T) trials back in the sequence, see Figure 1. Each level had a target frequency of 33%. The responses to target and non-target letters involved pressing a right CTRL key on the computer keyboard marked as "YES", and a left CTRL key marked as "NO", respectively. Task performance (reaction time and accuracy) were recorded by the experimental software.

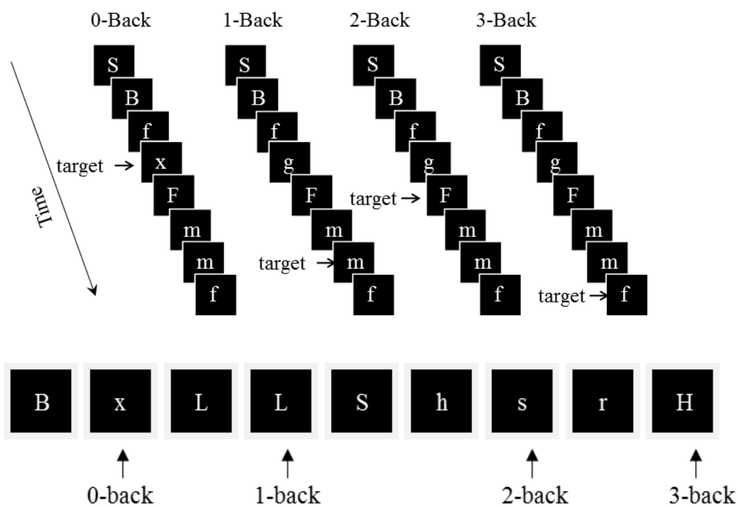

Figure 1:
N-back experiment. In the N-back task, participants need to identify whether the current letter that they are seeing (only one letter is presented on the screen at a time) has occurred 1, 2 or 3 letters before. In the string of letters above, the L occurred one step before, the s occurred as a large S two steps before, and the H occurred as a small s three steps before. In the 0-back condition, only the x has to be discriminated from all other letters

Word Recall Task. The recall task was also programmed with ERTS (Beringer, 1994). There were three memory lists, each involving 8 pairs of words (target word and distracter) which were presented in a randomized order. We used a rhymed non-word distracter because this distracter could cause a phonetic effect (Olson, Davidson, Kliegl, & Davies, 1984) and at the same time preserved much of the letter structure of the target word (Lange-Küttner, 2005; Lange-Küttner & Krappmann, 2011). In each word list, four of the eight target words were on the left-hand side, and the other four target words on the right-hand side, in a randomized sequence. In the repetition, target and distracter had swapped places on the screen and thus were counterbalanced for left-right position for all stimulus pairs.

There were three memory lists and each one was once repeated resulting in six verbal recall memory blocks. The repeated block always followed immediately after the completed recall of the first presentation of the word list (immediate repetition). There were 24 target words (3 lists x 8 target words in a randomized order) to remember, see Table 1. Because each list was once repeated, there were 48 presentation trials in total.

Each presented word pair consisted of a target word, a noun of five or six letters with a frequency of less than 30 per million (Leech, Rayson, & Wilson, 2001), and a distracter. The non-word distracters were created by exchanging the onset of each syllable with a random letter. Words were set in capital letters, Times Font NRC7bit size 12, in white on a black background.

We cued participants with eyes at a central fixation point as a non-informative visual cue, looking at either a target word or at a distracter, see Figure 2.

The first block did not include the eye gaze cues. In the second and third block, the eye cue appeared on the computer screen in the inter-stimulus interval (ISI) for 675ms with a delay (blank screen) of 150ms (stimulus onset asynchrony; SOA) as a pre-cue, followed by a second onset along with the stimulus word pair for a duration of 1000 ms.

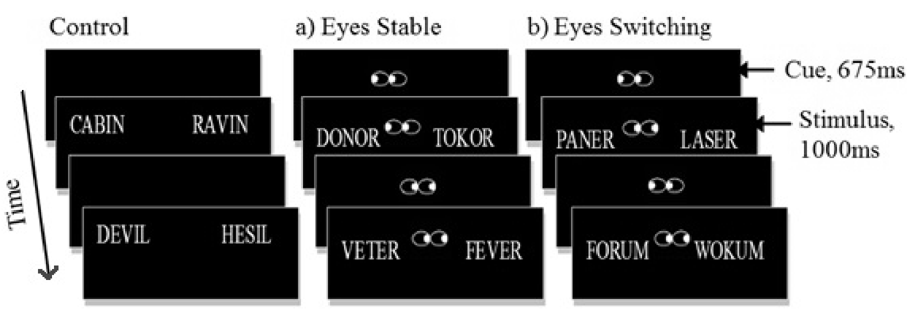

Figure 2:
Verbal Recall experiment. The three conditions of the verbal recall experiment are illustrated. The first word memory list was presented as target and distracter words only. Visual gaze cues were added to the second and third memory list to maintain interest and prevent ceiling effects during the repetition. The two types of eye gaze (switch vs. no switch between pre-cue and presentation cue) was a between-subjects factor

Table 1:
Word Lists with Pairs of Target Words and Non-Word Distracters, Study 1 and 2

Target Word	Non-Word Distracter	Target Word	Non-Word Distracter
List 1A		*List 1B*	
CABIN	RAVIN	RAVIN	CABIN
SEBON	LEMON	LEMON	SEBON
FORUM	WOKUM	WOKUM	FORUM
WAZKET	PACKET	PACKET	WAZKET
TALENT	RAVENT	RAVENT	TALENT
WACAL	CANAL	CANAL	WACAL
ENZYME	ONKYME	ONKYME	ENZYME
HESIL	DEVIL	DEVIL	HESIL
List 2A		*List 2B*	
DONOR	TOKOR	TOKOR	DONOR
SAHER	MAKER	MAKER	SAHER
ELBOW	ILWOW	ILWOW	ELBOW
VELER	FEVER	FEVER	VELER
ANRANT	INFANT	INFANT	ANRANT
RESORT	ZECORT	ZECORT	RESORT
PANMER	BANKER	BANKER	PANMER
MOTIVE	KODIVE	KODIVE	MOTIVE
List 3A		*List 3B*	
BONUS	GOLUS	GOLUS	BONUS
TAUGE	SAUCE	SAUCE	TAUGE
MANOR	GAFOR	GAFOR	MANOR
BAHAGE	GARAGE	GARAGE	BAHAGE
WISDOM	DISGOM	DISGOM	WISDOM
DAFRIX	MATRIX	MATRIX	DAFRIX
LASER	PANER	PANER	LASER
KELEL	REBEL	REBEL	KELEL

The eye gaze direction could either change in the second eye cue presentation phase, or it was stable. Participants were randomly allocated to one of two between-subjects conditions of the recall task: 1) ‘eyes stable’ condition in which the eyes maintained direction (n=23, 11 monolinguals, 12 bilinguals) and 2) ‘eyes switching’ condition in which the eyes changed looking direction between pre-cue and stimulus presentation participants (n=22, 10 monolinguals, 12 bilinguals). The eye gaze towards the target versus the distracter word was counterbalanced in both cue type conditions. Participants were writing down the words in any order (free recall) after each list.

The recall task instructions informed participants that they would see a series of words paired with non-words, one pair at a time, and that their task was to remember only the words for a later recall test. They were requested to fixate at the center of the screen each time the picture of eyes appeared. Participants were informed that although the eyes would sometimes be looking to the left and sometimes towards the right word, this information was uncertain because direction of their gaze would not indicate where the target words would appear.

Participants were tested individually in a laboratory in the basement of the Psychology department on the City Campus. The lab was empty except for a table with a personal computer. After the informed consent form was obtained from participants, they performed the N-Back task first and then the recall task.

Two measures were calculated for each N-Back condition: mean percentage of correctly identified targets (target %) and mean percentage of correctly identified non-targets (non-target %), for both reaction times and accuracy. All N-Back levels were once repeated by the participants, and were averaged before analysis. For the verbal recall task, the percentage of correctly recalled words was calculated per word list and double-checked. Only correctly spelled words were counted, hence there were no disagreements.

Results Study 1

We first report the results of the N-Back task, and then the results for the word recall task. If there was a significant interaction effect with sex, this was followed up with a split sample analysis, that is, the same model was run again separately for men and women. If the Mauchly's test of Sphericity was significant, degrees of freedom were adjusted according to Huynh-Feldt. Because the statistical effects are listed in the Tables, they are not stated again in the text.

The N-Back Task Study 1

Study 1. N-back reaction times. A 4 (N-Back Levels) by 2 (Target/Non-Target) by 2 (Language Group) by 2 (Sex) MANOVA with repeated measures on the first three factors was run, see Table 2. There were no significant effects of the language groups, $p_s > .216$. All between-subject effects were non-significant, $p_s > .118$ showing no differences in selective attention.

A main effect of levels showed that reaction times increased during the task, the more difficult the level (0-back M = 486 ms, 1-back M = 575 ms, 2-back M = 653 ms, 3-back M = 645 ms). Targets triggered faster response times (M = 577 ms) than non-targets (M = 603 ms). Moreover, these two factors interacted; pairwise post-hoc t-tests (two-tailed) showed that target response times were faster than for non-targets at all levels,

Table 2:
Study 1 Statistical MANOVA Effects for the N-Back Task (*N*=45)

Within-subject Effects						
	Reaction Times			Accuracy		
Statistical Effect	*F*	*p*	η2	*F*	*p*	η2
Levels	**40.59**	**.000**	**.497**	**149.164**	**.000**	**.784**
Levels*Language	.347	.792	.008	.462	.709	.011
Levels*Sex	2.611	.054	.060	.448	.719	.011
Levels*Language*Sex	.182	.908	.004	.175	.913	.004
Targets	**14.70**	**.000**	**.264**	**88.713**	**.000**	**.684**
Targets*Language	.486	.490	.012	.030	.864	.001
Targets*Sex	2.423	.127	.056	**4.074**	**.050**	**.090**
Targets*Language*Sex	.006	.937	.000	1.723	.197	.040
Levels*Targets	**9.682**	**.000**	**.191**	**56.000**	**.000**	**.577**
Levels*Targets*Language	1.506	.216	.035	.308	.820	.007
Levels*Targets*Sex	1.296	.279	.031	**3.084**	**.030**	**.007**
Levels*Targets*Language*Sex	.464	.708	.011	.491	.689	.070
Between-subject Effects						
Language	.122	.729	.003	.041	.840	.001
Sex	2.547	.118	.058	.314	.578	.008
Sex*Language	.080	.778	.002	.005	.944	.000

Note. Significant effects are set in bold

t_s (44) > -4.56, p_s < .006, except for the n-back3 task where reaction times were similar for target (*M* = 651 ms) and non-target letters (*M* = 644 ms).

Study 1. N-back accuracy. The same model for N-back accuracy showed again no language group performance differences, p_s > .197. The expected different levels of difficulty of the N-back task also showed in accuracy. A significant effect for the N-back levels showed lower accuracy the more distance there was between the repeated targets (0-back *M* = 92.7%, 1-back *M* = 89.3%, 2-back *M* = 76.6%, 3-back *M* = 66.9%). Targets (*M* = 74.3%) were more difficult to accurately respond to than non-targets (*M* = 88.4%). Moreover, there was a two-way interaction between these two factors; pairwise post-hoc t-tests (two-tailed) showed that target accuracy was less correct than for non-targets at all levels, t_s(44) > -2.84, p_s < .007. This difference was most increased at n-back3 level, *t*(44) =-11.17, *p* < .001, with 51.2% correct for targets and 82.7% for non-targets. This effect varied for men and women. Accuracy in identifying a repeated target letter decreased more in men than women, but this was not the case for non-targets, see Figure 3.

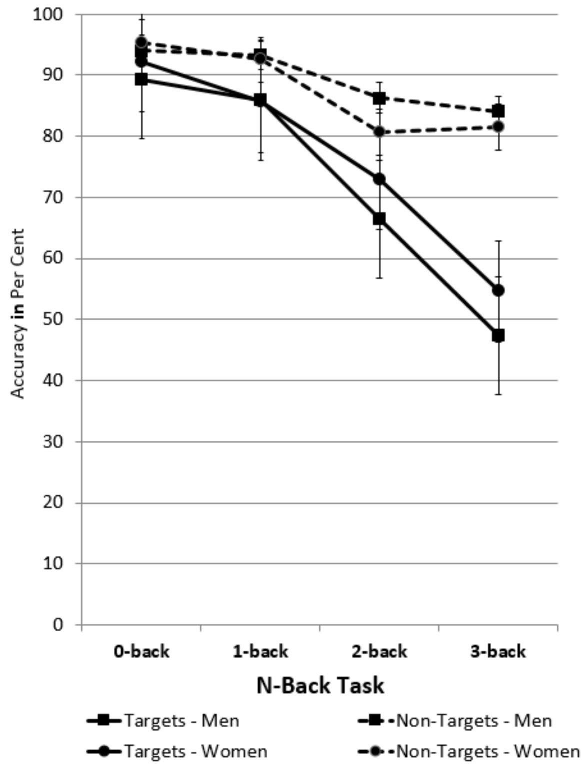

Figure 3:
Study 1. N-Back Task Performance. In the course of the experiment, the gap between detecting targets vs. non-targets widened more in men (square marker) than in women (round marker). Error bars represent the standard error (SE)

The Word Recall Task Study 1

Study 1. Word recall. A 3 (List 1, List 2, List 3) × 2 (Repetition) × 2 (Stable/Switch Eye Gaze Cue) x 2 (Language Group) by 2 (Sex) MANOVA with repeated measures for the first two factors, and eye gaze condition, language and sex as between-subject factors was run, see Table 3 (significant effects are set in bold).

A significant effect of word lists showed the expected deterioration of word memory. The memory score deteriorated from 66.0% to 60.4% to 54.9%. However, a highly significant effect of repetition with a large effect size of $\eta 2 = .86$ showed that participants remembered more in the repeated lists (71.1%) than in the initial presentations (49.7%). Hence, we found the predicted deterioration and recovery.

Table 3:
Study 1 Statistical MANOVA Effects for the Word Recall Task (*N*=45)

Statistical Effect	df	*F*	*p*	η2
		Within-subject Effects		
Lists	**2**	**12.920**	**.000**	**.259**
Lists*Language	2	1.994	.143	.051
Lists*Sex	2	.316	.730	.008
Lists*Gaze	2	.340	.713	.009
Lists*Language*Sex	2	1.033	.361	.027
Lists*Language*Gaze	2	.037	.964	.001
Lists*Sex*Gaze	2	.668	.516	.018
Lists*Language*Sex*Gaze	2	.111	.895	.003
Repetition	**1**	**221.21**	**.000**	**.857**
Repetition*Language	**1**	**15.802**	**.000**	**.299**
Repetition*Sex	1	.945	.337	.025
Repetition*Gaze	1	1.233	.274	.032
Repetition*Language*Sex	1	1.458	.235	.038
Repetition*Language*Gaze	1	.147	.703	.004
Repetition*Sex*Gaze	1	.000	.997	.000
Repetition*Language*Sex*Gaze	**1**	**4.736**	**.036**	**.113**
Lists*Repetition	2	.589	.558	.016
Lists*Repetition*Language	2	.725	.488	.019
Lists*Repetition*Sex	2	1.130	.328	.030
Lists*Repetition*Language*Sex	**2**	**3.439**	**.037**	**.085**
Lists*Repetition*Language*Gaze	2	2.059	.135	.053
Lists*Repetition*Sex*Gaze	2	.912	.406	.024
Lists*Repetition*Language*Sex*Gaze	2	2.404	.097	.061
		Between-subject Effects		
Language	1	1.437	.238	.037
Sex	1	.174	.679	.005
Gaze	1	.133	.717	.004
Language*Sex	1	.125	.726	.003
Language*Gaze	1	1.415	.242	.037
Sex*Gaze	1	.975	.330	.026
Language*Sex*Gaze	1	.725	.400	.019

We also found the predicted interaction effects of the repetition improvements with individual differences. The repetition effect significantly interacted two-way with the two language groups, but also four-way with language, sex and gaze condition, and in a second four-way interaction with the deterioration of the memory score (lists). These four-way interactions were followed up with a split-sample analysis. The results are listed in Table 4.

Significant memory deterioration occurred for both men and women, $p_s < .005$. Also the recovery of the memory score during the repetition was significant for the both men and women sample, $p_s < .001$.

However, in the male sample only, several effects of language group were significant. While women tended to benefit from the repetition independently of whether they spoke one or more languages (monolingual women, first list recall M = 49.9%, second list recall M = 68.7%; bilingual women, first list recall M = 49.6%, second list recall M = 76.2%), in men this was different. Monolingual men benefited significantly less from the

Table 4:
Study 1 Statistical MANOVA Effects for the Word Recall Task (Split Sample by Sex)

Within-subject Effects						
	Male (*n*=21)			Female (*n*=24)		
Statistical Effect	*F*	*p*	η2	*F*	*p*	η2
Lists	**7.426**	**.002**	**.304**	**6.065**	**.005**	**.233**
Lists*Language	2.990	.064	.150	.198	.821	.010
Lists*Gaze	1.026	.369	.057	.038	.962	.002
Lists*Language*Gaze	.034	.967	.002	.112	.894	.006
Repetition	**103.893**	**.000**	**.859**	**121.083**	**.000**	**.858**
Repetition*Language	**14.441**	**.001**	**.459**	3.694	.069	.156
Repetition*Gaze	.667	.425	.038	.591	.451	.029
Repetition*Language*Gaze	1.727	.206	.092	3.161	.091	.136
Lists*Repetition	.830	.445	.047	.922	.406	.044
Lists*Repetition*Language	**4.026**	**.027**	**.191**	.607	.550	.029
Lists*Repetition*Gaze	1.170	.322	.064	.649	.528	.031
Lists*Repetition*Language*Gaze	**3.466**	**.043**	**.169**	1.335	.275	.063
Between-subject Effects						
Language	.393	.539	.023	1.143	.298	.054
Gaze	1.007	.330	.056	.184	.673	.009
Language*Gaze	.063	.805	.004	1.978	.175	.090

Note. Significant effects are set in bold repetition (first list recall M = 51.1%, second list recall M = 62.6%) than bilingual men (first list recall M = 48.3%, second list recall M = 76.7%). Thus, on average, monolingual men benefited the least from the repetitions of the word list.

Moreover, in men only, the memory deterioration was significant in interaction with repetition, language and the eye cue condition. Post-hoc tests (two-tailed) showed that monolingual men overcame the steep decline of their word memory only in the last word list if the gaze cues remained stable (Figure 4A). In contrast, all bilingual men significantly improved during all the repetitions independently of the gaze cue (Figure 4B).

Discussion Study 1

We investigated whether a hypothesized bilingual advantage in PI would become statistically significant if we allowed participants to repeat memorizing a word list in order to consolidate their word memory before presenting them with a new word list. There were some sex differences in the N-back task, but no differences between language groups which was what we wanted to ascertain.

We could show that in our verbal recall task, participants' memory indeed decreased, but it recovered during the repetition when target and distracter had changed places. However, this was significant only in the men sample. The monolingual men started with good verbal recall of more than 60% accuracy in the first verbal recall block, but showed a pronounced deterioration thereafter to around 40% accuracy. They showed no, or only a very weak and non-significant improvement during the repetition blocks. In contrast, while bilingual men's verbal recall in the first presentations of each memory list was modestly just under 50%, they could maintain this level during the experiment and their verbal recall score did not deteriorate. Moreover, in each of the repetition blocks, they significantly upped their performance to over 70% accurate recall on each occasion.

It is very likely that they built verbal rehearsal groups (Lehmann, 2015), with one group in the first list and adding a second one of the unresolved items in the repetition. Their results speak to such a strategy as the first and second list presentation was always remembered to about the same degree. This mechanical creation of rehearsal sets would facilitate the generation of self-contained informational units (Morra & Epidendio, 2015), for instance, a word list with six words could be conveniently grouped into three pairs or two triplets. Likewise, in the current study, eight words per memory list could be grouped into two sets of quadruples. A possible strategy of the bilingual men could have been to learn the first quadruples in the initial recall, and top up the list with the second set of quadruples in the repetition. In contrast, a less efficient strategy would be to recall the words as individual items rather than as sets which would result in more individual units to remember (Watkins & Watkins, 1975).

Thus, these results show that the mono- resp. bilingual effect on PI was associated with weak rehearsal in the repetition. Moreover, the monolingual male sample was also more prone to become distracted by the eye cues. Thus, we found in this first study a significant monolingual disadvantage in verbal rehearsal and recall. Because this was a rela-

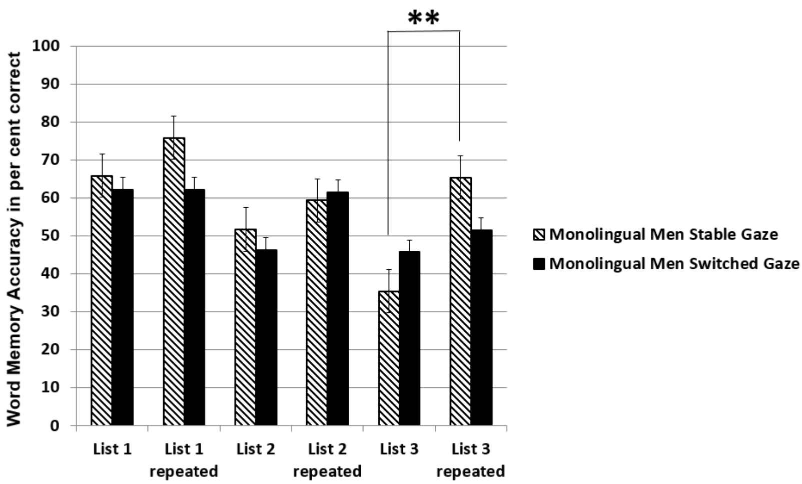

Figure 4A Monolingual Men

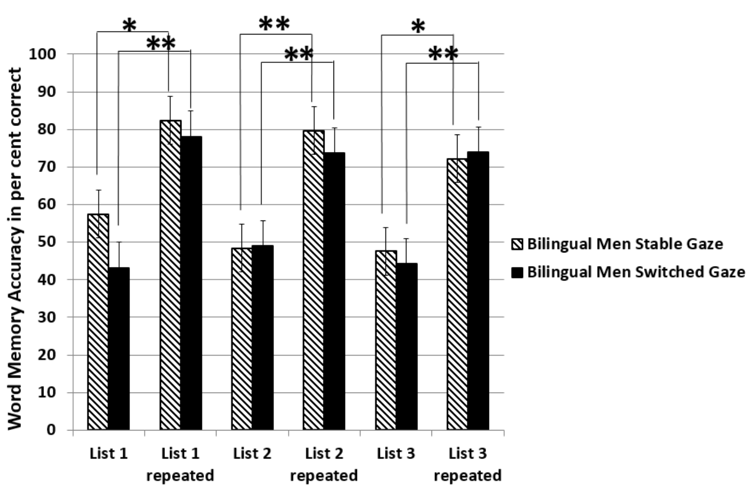

Figure 4B Bilingual Men

Figure 4:

Study 1. No Gaze Cues in List 1, Stable or Switched Gaze Direction between Pre-Cue and Stimulus Cue in Lists 2 and 3. Monolingual men significantly benefited from word list repetition only in the last word list if the gaze cues remained stable so that they overcame the steep decline of their word memory (Figure 4A). All bilingual men improved during the repetition independently of the gaze (Figure 4B). Error bars represent the standard error (SE). * = $p < .05$, ** = $p < .01$

tively small sample, we ran a replication study with a larger sample, with the same verbal recall task, but a shortened N-back task. We report this study separately because we aimed to replicate the study with a different equipment for testing, on a different campus in a different laboratory.

Study 2

Participants

Eighty-five university students on another campus of the same university participated. Five data sets were excluded due to machine recording failure. Two data sets were excluded because the second language was only spoken rarely. In the remaining sample of $N = 78$, there were 39 participants in the stable eye gaze condition (16 monolinguals and 23 bilinguals) and 39 in the switched eye gaze condition (18 monolinguals and 21 bilinguals).

There were 34 monolingual speakers of English (18 males, 16 females) and 47 speakers of English and at least one or more languages (24 males, 20 females). Participants were from 19 to 44 years of age (monolinguals $M = 24$ years, range 19-42 years; bilinguals $M = 24$ years, range 18-44 years).

This sample of bi- and multilinguals had a multi-cultural background of 30 languages. Languages in addition to English were Armenian, Arabic, Bengali, Catalan, Estonian, Farsi, French, German, Greek, Hebrew, Hindi, Hungarian, Italian, Kurdish, Lingala, Malayalee, Maltese, Nepalese, Persian, Polish, Portuguese, Punjabi, Romanian, Russian, Spanish, Tamil, Turkish, Ugandan, Urdu, Yoruba.

None of the monolingual English speakers reported actively speaking another language. For fluency, bilinguals had an average value of $M = 2.84$ for English and $M = 2.79$ for the other language which showed that they were balanced bilinguals. For usage, bilinguals had an average value for English of $M = 3.0$, just like monolinguals, and $M = 2.59$ for the other language which is a value between daily and sometimes that shows that the second language was spoken somewhat less often than English.

Apparatus and procedure

The same *N-back task* (Schleepen & Jonkman, 2010) as in Study 1 was used, but we reduced the test session length by half just testing 1- and 3-back levels once. We also used the same *Word Recall Task.* The same Consent and Debrief forms were used as in Study 1.

The experiment was run on a Windows XP Professional laptop computer (Fujitsu Siemens) and stimuli were presented on a 19 inch (diagonal) Samsung Syncmaster LCD monitor (TFT active matrix), model 940B. The N-back task was presented on the same computer as in Study 1. Participants were tested individually in a laboratory in the Science Centre on the North Campus of the university.

Results Study 2

The N-Back Task Study 2

Study 2. N-back reaction times. A 4 (N-Back Levels) by 2 (Target/Non-Target) by 2 (Language Group) by 2 (Sex) MANOVA with repeated measures on the first three factors was run, see Table 5. None of the between-subject effects were significant, $p_s > .599$.

A main effect of levels showed that reaction times were faster for the 1-back than for the 3back task (1-back M = 626 ms, 3-back M = 672 ms). However, a two-way interaction showed that this was different for targets and non-targets; pairwise post-hoc t-tests (two-tailed) showed that target response times were faster than responses for non-targets at the 1-back level, $t(77) = -3.47$, $p = .001$, (1-back M = 605 ms, 3-back M = 683 ms), but not at the 3-back level, $t(77) = 1.56$, $p = .122$, (1-back M = 651 ms, 3-back M = 661 ms). This is the same kind of result as in Study 1.

Table 5:
Study 2 Statistical MANOVA Effects for the N-Back Task (*N*=78)

Within-subject Effects						
	Reaction Times			Accuracy		
Statistical Effect	*F*	*p*	η2	*F*	*p*	η2
Levels	**10.687**	**.002**	**.126**	**174.679**	**.000**	**.702**
Levels*Language	.876	.352	.004	4.792	.032	.061
Levels*Sex	.002	.963	.000	.040	.842	.001
Levels*Language*Sex	.908	.344	.012	3.617	.061	.047
Targets	1.045	.310	.014	**24.761**	**.000**	**.251**
Targets*Language	.304	.583	.004	.042	.838	.001
Targets*Sex	.004	.950	.000	.256	.615	.003
Targets*Language*Sex	1.578	.213	.021	**4.272**	**.042**	**.042**
Levels*Targets	**17.572**	**.000**	**.192**	**72.164**	**.000**	**.494**
Levels*Targets*Language	.782	.380	.010	1.660	.202	.022
Levels*Targets*Sex	.590	.445	.008	3.163	.079	.041
Levels*Targets*Language*Sex	.971	.328	.013	.123	.726	.002
Between-subject Effects						
Language	.147	.702	.002	.017	.896	.000
Sex	.081	.776	.001	.933	.337	.012
Sex*Language	.278	.599	.004	1.051	.309	.014

Note. Significant effects are set in bold

Study 2. N-back accuracy. The same model for N-back accuracy showed no significant between-subjects group performance differences, $p_s > .337$. The expected different levels of difficulty of the N-back task also showed in accuracy. Like in Study 1, a highly significant effect for the N-back levels showed lower accuracy if there was more distance between the repeated targets (1-back M = 80.2%, 3-back M = 57.4%).

Targets (M = 62.6%) were more difficult to respond to than non-targets (M = 74.6%). There was a two-way interaction between these two factors; pairwise post-hoc t-tests (two-tailed) showed that target (M = 79.9 %) and non-target (M = 80.6 %) accuracy was very similar at the 1-back level, $t(77) = -.394$, $p = .695$. In contrast, this difference was significant at n-back3 level, $t(77) = -7.14$, $p < .001$, with 46.4% correct for targets and 69.1% for non-targets. This was also the same kind of result as in Study 1.

In addition, there was a significant three-way interaction effect of targets, sex, and language group, see Figure 5. Pairwise t-tests (two-tailed) showed that there was no significant difference between targets and non-target accuracy in monolingual females, $t(15) = -.997$ $p = .334$, and bilingual males, $t(23) = -1.863$, $p = .075$. However, the difference was significant in monolingual males, $t(17) = -5.089$, $p < .001$, as well as in bilingual females, $t(19) = -3.135$, $p = .005$, as both these groups performed lower for targets than for non-targets.

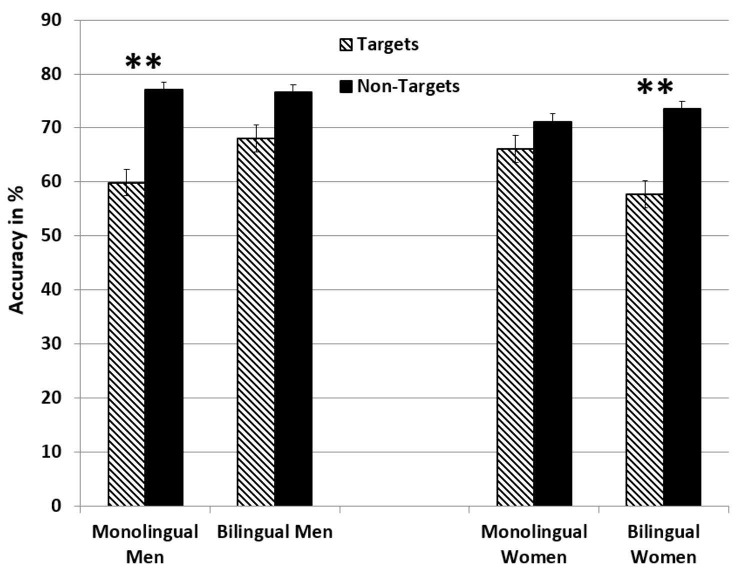

Figure 5:
Study 2. N-Back Task. Monolingual men and bilingual women showed a significantly lower level of accuracy for targets than non-targets. Error bars represent the standard error (SE). Pairwise t-tests (two-tailed) * = $p < .05$, ** = $p < .01$

The Word Recall Task Study 2

Study 2. Word recall. A 3 (List 1, List 2, List 3) × 2 (Repetition) × 2 (Stable/Switch Eye Gaze Cue) x 2 (Language Group) by 2 (Sex) MANOVA with repeated measures for the first two factors was run. Eye gaze cue condition, language group and sex were between-subjects factors. We found no between-subjects differences, $p_s > .292$, see Table 6 for an overview of the statistical effects.

A significant effect of word lists showed the expected significant deterioration of word memory. The memory score deteriorated from 72.3% to 63.2% to 62.0%. This varied in men and women, as women's recall (73.3% to 64.6% to 59.6%) deteriorated more than in men (71.3% to 61.8% to 64.4%). However, memory recovered during the repetition like in Study 1: A highly significant effect of repetition with a large effect size of $\eta^2 = .70$ showed that participants remembered more in the repeated lists (73.5%) than in the initial presentations (58.0%). Again, we obtained interactions with individual differences. We obtained a five-way significant interaction that involved all factors, lists, repetition, language, sex and gaze condition. This was followed up with a split-sample analysis by sex, see Table 7.

Memory deterioration effects occurred in both men and women, $p_s < .001$. Also recovery of the memory score during the repetition occurred in both men and women, $p_s < .001$. In the male sample, the repetition effect interacted with the gaze condition. Men in the stable eye gaze condition showed a stronger repetition increase from a lower baseline (Original M = 56.1%; Repetition 76.5%) than men in the condition when the eye gaze cue changed viewing direction from pre-cue to word stimulus pair (Original M = 58.5%; Repetition 72.2%).

Only in the women, a significant five-way interaction effect of lists, repetition, language and gaze occurred. Post-hoc tests (two-tailed) per word list are illustrated in Figure 6. Overall, bilingual women more often showed an improvement in the repetition of a word list, but the difference between the two female language groups was not as strong as between the men in Study 1. Post-hoc tests (two-tailed) showed that some monolingual women significantly improved their memory score during the repetition of the first word list without eye cues, $t(6) = -3.87$, $p = .008$, but improved their memory score in the Lists 2 and 3 with switched eye cues, List 2: $t(8) = -2.80$, $p = .023$; List 3: $t(8) = -3.16$, $p = .013$, see Figure 6A. Bilingual women in the stable eye gaze condition constantly increased their memory score, List 1, $t(9) = -1.96$, $p = .081$; List 2: $t(9) = -4.64$, $p = .001$; List 3: $t(9) = -9.30$, $p < .001$, see Figure 6B, and so did the bilingual women in the switched condition except in the last block, List 1, $t(9) = -2.75$, $p = .022$; List 2: $t(9) = -2.75$, $p = .022$; List 3: $t(9) = -.921$, $p < .381$.

Table 6:
Study 2 Statistical MANOVA Effects for the Word Recall Task (*N*=78)

Statistical Effect	df	*F*	*p*	η2
Within-subject Effects				
Lists	**2**	**22.318**	**.000**	**.242**
Lists*Language	2	.059	.943	.001
Lists*Sex	**2**	**3.087**	**.049**	**.042**
Lists*Gaze	2	.119	.888	.002
Lists*Language*Sex	2	.383	.682	.005
Lists*Language*Gaze	2	.921	.401	.013
Lists*Sex*Gaze	2	.303	.739	.004
Lists*Language*Sex*Gaze	2	.527	.592	.007
Repetition	**1**	**160.524**	**.000**	**.696**
Repetition*Language	1	1.317	.255	.018
Repetition*Sex	1	2.031	.159	.028
Repetition*Gaze	**1**	**5.027**	**.028**	**.067**
Repetition*Language*Sex	1	1.471	.229	.021
Repetition*Language*Gaze	1	.083	.775	.001
Repetition*Sex*Gaze	1	.298	.587	.004
Repetition*Language*Sex*Gaze	1	.891	.348	.013
Lists*Repetition	2	1.814	.167	.025
Lists*Repetition*Language	2	1.217	.299	.017
Lists*Repetition*Sex	2	.316	.730	.004
Lists*Repetition*Gaze	2	1.279	.282	.018
Lists*Repetition*Language*Sex	2	.481	.619	.007
Lists*Repetition*Language*Gaze	2	2.982	.054	.041
Lists*Repetition*Sex*Gaze	2	.115	.892	.002
Lists*Repetition*Language*Sex*Gaze	**2**	**4.731**	**.010**	**.063**
Between-subject Effects				
Language	1	.787	.378	.011
Sex	1	.000	.994	.000
Gaze	1	.021	.885	.000
Language*Sex	1	.037	.847	.001
Language*Gaze	1	1.125	.292	.016
Sex*Gaze	1	.165	.686	.002
Language*Sex*Gaze	1	.640	.426	.009

Note. Significant effects are set in bold.

Table 7:
Study 2 Statistical MANOVA Effects for the Word Recall Task (Split Sample by Sex)

Within-subject Effects						
	Male (*n*=42)			Female (*n*=36)		
Statistical Effect	*F*	*p*	η2	*F*	*p*	η2
Lists	**8.566**	**.000**	**.184**	**17.284**	**.000**	**.351**
Lists*Language	.371	.691	.010	.074	.929	.002
Lists*Gaze	.391	.678	.010	.034	.967	.001
Lists*Language*Gaze	.364	.696	.009	1.112	.335	.034
Repetition	**140.894**	**.000**	**.788**	**45.854**	**.000**	**.589**
Repetition*Language	.003	.957	.000	2.021	.165	.059
Repetition*Gaze	**5.512**	**.024**	**.127**	1.043	.315	.032
Repetition*Language*Gaze	.306	.584	.008	.550	.464	.017
Lists*Repetition	2.100	.130	.052	.267	.767	.008
Lists*Repetition*Language	.413	.663	.011	1.160	.320	.035
Lists*Repetition*Gaze	.603	.550	.016	.754	.475	.023
Lists*Repetition*Language*Gaze	.254	.776	.007	**6.491**	**.003**	**.169**
Between-subject Effects						
Language	.243	.632	.006	.628	.434	.019
Gaze	.033	.857	.001	.164	.688	.005
Language*Gaze	.033	.857	.000	1.863	.182	.055

Note. Significant effects are set in bold

Discussion Study 2

We could replicate the statistical effects of Study 1 in Study 2 with a sample that was about the double size compared to Study 1. The N-back task showed again no main effect between the language groups. In the verbal recall task, deterioration of the word memory score occurred and recovery effects happened during the repetition. Like in Study 1, the interaction between the deterioration and recovery during repetition showed only along with the control of individual differences. One could say that while in Study 1, we found a pronounced disadvantage only in the monolingual men sample, in Study 2 we found some disadvantage only in the monolingual women sample. In short, these differences in verbal recall were not of the deterministic sort where strong conclusions or predictions with regards to mono- or bilingualism in general could be made.

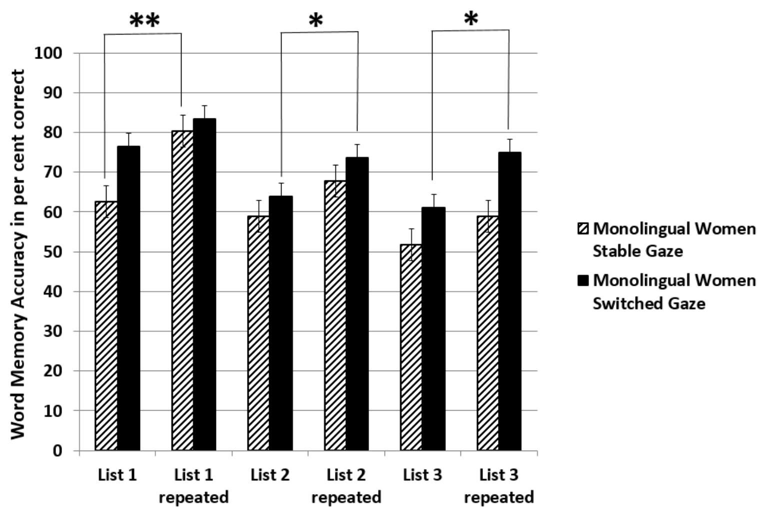

Figure 6A Monolingual Women

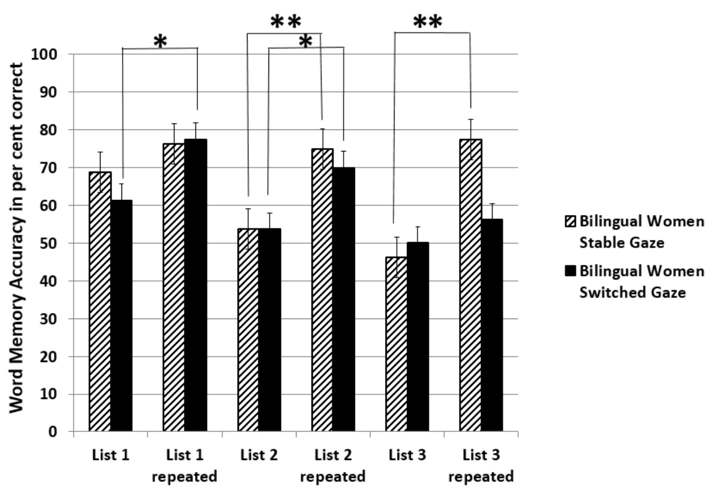

Figure 6B Bilingual Women

Figure 6:

Study 2. No Gaze Cues in List 1, Stable or Switched Gaze Direction between Pre-Cue and Stimulus Cue in Lists 2 and 3. Monolingual women's memory score significantly decreased because of insufficient recovery during repetition (Figure 6A). The bilingual women in the stable gaze condition could maintain their memory score during the experiment in the repeated lists (Figure 6B). Error bars represent the standard error (SE).

* = $p < .05$, ** = $p < .01$

Study 3

Hence, we may need to agree with the argument that executive function can be improved for a variety of reasons other than bilingualism (Bialystok & DePape, 2009; Gathercole, 2015; Paap, Johnson, & Sawi, 2015; Valian, 2015). A recent meta-analysis of 35 studies of inhibitory control and non-verbal interference monitoring (Paap, 2015; Paap & Sawi, 2014) showed that individual differences between mono- and bilinguals could only be found in smaller sample sizes. The advantage of bilinguals appears to be so marginal that it disappears once samples are large enough (Paap & Sawi, 2014). One could argue that Paap and Sawi could find only four studies with more than 100 participants for their meta-analysis. However, in the more numerous studies with less than 70 participants a clear trend was visible that the larger the sample, the less likely it was that a bilingual advantage would occur.

For this reason, we pooled the samples of Studies 1 and 2 into one. We extracted the data for the 1-back and 3-back level of the four-level N-Back task in Study 1 to match the data of Study 2.

Results Study 3

The N-Back Task Study 3

Study 3. N-back reaction times. A 4 (N-Back Levels) by 2 (Target/Non-Target) by 2 (Language Group) by 2 (Sex) MANOVA with repeated measures on the first two factors was run, see Table 8. None of the between-subject effects were significant, $p_s > .240$.

A main effect of levels showed that reaction times were faster for the 1-back than for the 3-back task (1-back M = 618 ms, 3-back M = 668 ms). However, this was dependent on whether a target was remembered or not; pairwise post-hoc t-tests (two-tailed) showed that target response times were faster at the 1-back level, $t(122) = -4.83$, $p < .001$ (1-back M = 597 ms, 3-back M = 642 ms), but at the 3-back level, reaction times were not faster for targets $t(122) = 1.83$, $p = .070$, (1-back M = 677 ms, 3-back M = 658 ms).

Study 3. N-back accuracy. The same model for N-back accuracy showed again no significant between-subjects group performance differences, $p_s > .472$. The expected different levels of difficulty of the N-back task also showed in accuracy. A highly significant effect for the N-back levels with a large effect size of .71 showed lower accuracy if there was more distance between the repeated targets (1-back M = 83.1%, 3-back M = 61.0%).

Targets (M = 64.8%) were more difficult to respond to than non-targets (M = 79.3%). There was a two-way interaction between these two factors; pairwise post-hoc t-tests (two-tailed) showed that target (M = 81.5 %) and non-target (M = 84.6 %) accuracy was less different at the 1-back level, $t(122) = -2.095$, $p = .039$, than at the n-back3 level, $t(122) = -11.472$, $p < .001$, with 48.3% correct targets and 74.1% correct non-targets.

Table 8:
Study 3 MANOVA Results for the N-Back Task (*N*=123)

Within-subject Effects						
	Reaction Times			Accuracy		
Statistical Effect	*F*	*p*	η2	*F*	*p*	η2
Levels	**18.114**	**.000**	**.132**	**295.66**	**.000**	**.713**
Levels*Language	1.009	.317	.008	3.255	.074	.027
Levels*Sex	.166	.685	.001	.004	.948	.000
Levels*Language*Sex	.168	.683	.001	2.050	.155	.017
Targets	2.595	.110	.021	**78.424**	**.000**	**.397**
Targets*Language	1.113	.294	.009	.030	.864	.000
Targets*Sex	.012	.913	.000	.526	.470	.004
Targets*Language*Sex	2.415	.123	.020	**4.371**	**.039**	**.035**
Levels*Targets	**29.169**	**.000**	**.197**	**134.520**	**.000**	**.531**
Levels*Targets*Language	1.647	.202	.014	1.150	.286	.010
Levels*Targets*Sex	.243	.623	.002	.360	.550	.003
Levels*Targets*Language*Sex	1.047	.308	.009	.044	.834	.000
Between-subject Effects						
Language	.188	.666	.002	.018	.895	.000
Sex	1.393	.240	.012	.521	.472	.004
Sex*Language	.056	.814	.009	.578	.449	.005

Note. Significant effects are set in bold

Also in this large sample of $N > 100$, individual differences did not completely disappear, see Table 8. There was a significant interaction effect for target by sex by language. The split-sample analysis by sex showed that the experimental effects were the same in both men and women, but in the women sample, there was a significant interaction between the difficulty of the two N-back levels and language groups, $F(1, 60) = 4.93$, $p = .030$). Pairwise tests (two-tailed) showed that monolingual women were significantly better when letters were close (1-back $M = 85.1\%$) rather than further apart (3-back $M = 58.8\%$), $t(26) = 9.936$, $p < .001$, while in bilingual women the gap was less pronounced. (1-back $M = 79.4\%$, 3-back $M = 61.4\%$), $t(32) = 6.922$, $p < .001$.

The Word Recall Task Study 3

Study 3. Word recall. A 3 (List 1, List 2, List 3) × 2 (Repetition) × 2 (Stable/Switch Eye Gaze Cue) x 2 (Language Group) by 2 (Sex) MANOVA with repeated measures for the first three factors and language and sex as between-subjects factors was run. We found no between-subjects group performance differences, $p_s > .126$, see Table 9.

Table 9:
Study 3 Statistical MANOVA Effects for the Word Recall Task (*N*=123)

Statistical Effect	df	*F*	*p*	η2
Within-subject Effects				
Lists	**2**	**28.921**	**.000**	**.201**
Lists*Language	2	.390	.677	.003
Lists*Sex	2	1.703	.184	.015
Lists*Gaze	2	.063	.939	.001
Lists*Language*Sex	2	.932	.395	.008
Lists*Language*Gaze	2	.437	.647	.004
Lists*Sex*Gaze	2	.536	.586	.005
Lists*Language*Sex*Gaze	2	1.144	.320	.010
Repetition	**1**	**15.548**	**.000**	**.119**
Repetition*Language	1	.009	.923	.000
Repetition*Sex	1	1.737	.190	.015
Repetition*Gaze	1	.766	.383	.007
Repetition*Language*Sex	1	.480	.490	.004
Repetition*Language*Gaze	1	.000	.995	.000
Repetition*Sex*Gaze	1	.151	.698	.001
Repetition*Language*Sex*Gaze	1	.012	.914	.000
Lists*Repetition	**2**	**7.751**	**.001**	**.063**
Lists*Repetition*Language	2	.706	.495	.006
Lists*Repetition*Sex	2	.701	.497	.006
Lists*Repetition*Language*Sex	2	1.269	.283	.011
Lists*Repetition*Language*Gaze	2	.785	.458	.007
Lists*Repetition*Sex*Gaze	2	.145	.865	.001
Lists*Repetition*Language*Sex*Gaze	2	1.503	.225	.013
Between-subject Effects				
Language	1	.229	.663	.002
Sex	1	.000	.983	.000
Gaze	1	.004	.947	.000
Language*Sex	1	.019	.891	.000
Language*Gaze	1	2.378	.126	.020
Sex*Gaze	1	.894	.346	.008
Language*Sex*Gaze	1	1.378	.243	.012

Note. Significant effects are set in bold.

For word recall, the large sample had indeed the effect that all individual differences were erased, and only the experimental effects were significant. A significant effect of word lists showed the expected significant deterioration of word memory, from 68.7% to 60.2% to 62.1%. The repetition effect was comparably diminished to an effect size of .12; participants remembered only somewhat more in the repeated lists (66.7%) than in the initial presentations (60.7%).

However, the memory deterioration that occurred in the three lists during the experiment significantly interacted with the recovery that occurred during the repetition without any involvement of individual differences due to multilingualism or sex. Post-hoc tests (pairwise) showed that there was a significant improvement in the repetition of List 1 (List 1A M = 65.5%, List 1B M = 71.8%, $t(32)$ = .3.99, p <.001), and in the repetition of List 2 (List 2A M = 54.8%, List 2B M = 65.3%), $t(122)$ = -6.53, p <.001), with no further deterioration in List 3 which was equally well remembered on the first and second

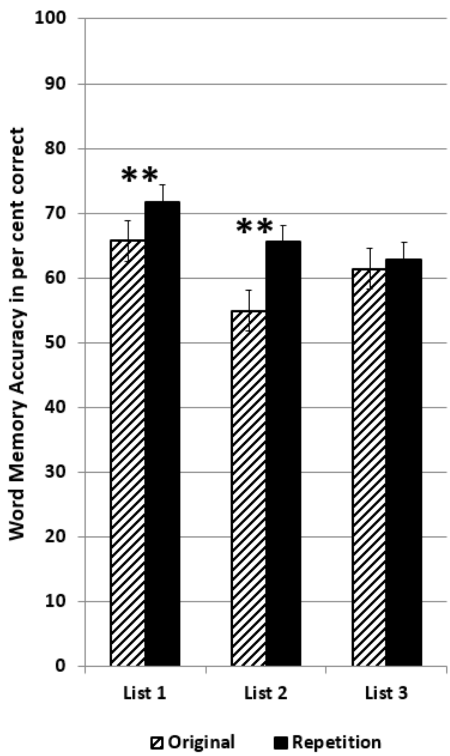

Figure 7:
Study 3. Deterioration and Recovery of Verbal Recall. In the large sample of N = 123 participants, the effect of repetition and recovery from PI is significant for the first and second list. For list 3, recall on first presentation has recovered. Error bars represent the standard error (SE). * = p < .05, ** = p < .01

presentation (List 3A M = 61.4%, List 3B M = 63.0%), $t(122) = -.619$, $p = .537$). A planned comparison showed that the verbal recall of the very first list (List 1A) and the very last list (List 3B) in the experiment was not significantly different, $t(122) = 1.395$, $p = .166$. Thus, while participants did not improve during repetition in the last memory block, the deterioration due to PI was halted, see Figure 7.

Discussion and Conclusion

Pro-active interference usually occurs when memory items from previous lists interfere with the recall of the items from a new list, leading to a pronounced deterioration of memory during a verbal recall experiment. This memory deterioration had also started to begin in the current experiment, but was halted. The important new result from the current studies was that this time, release from PI was not achieved by giving participants a new class of memory items in the last block, but by an immediate repetition of each word list that lead to enhanced memory consolidation of word lists.

The role of statistical power in bilingual research

The analyses of the two experimental studies showed a predicted small but significant monolingual disadvantage in using the rehearsal opportunity during the experiment. Monolinguals were also more likely to be distracted by the uninformative gaze cues which confirmed the relevance of intrusive stimuli for lower resistance to PI (Friedman & Miyake, 2004). In contrast, the better repeated performance of bilinguals during the experiment significantly prevented memory deterioration for the men-only sample in Study 1, and for the women-only sample in Study 2. Performance in the N-back tasks also showed sex differences, accordingly. Hence, the hypothesis that immediate repetition of word lists benefits especially bilinguals and would be useful in overcoming PI could be confirmed.

However, when these samples of the two studies were pooled into a large sample with more than one hundred individuals, the repetition effect mediated the memory deterioration without any individual differences. In short, a monolingual disadvantage in using the rehearsal opportunity during the memory list repetition was replicable – albeit in interaction with the sex of the participants – in the current study, but the disappearance of this effect in the large merged sample poses a research methods question about the status of replicability vs. testing large samples.

The absence of individual differences in the large sample confirms the Paap effect which predicts that only smaller samples are more likely to show significant differences due to multilingualism (Paap, 2015). The Paap effect could be explained with the argument that also monolinguals' executive functions may be improved by active practicing leading to better executive skills, e.g. learning to play musical instruments (Bialystok & DePape, 2009; Cox et al., 2016; Paap, Johnson, & Sawi, 2014; Valian, 2015). However, in the current study the large sample consisted of the very same individuals as in the two smaller studies. Only statistical power was increased in the merged sample.

In experimental psychology, it is common practice to test small samples with within-subject design involving many trials with the reckoning that the extensive testing of a small sample will amount to a comparable reliability as the processing of short testing of a large sample. Replications without (direct replication) or with small variations in the experimental stimuli (conceptual replications) are carried out in order to ascertain the reliability of the results (Pashler & Harris, 2012). Another main aim is to prove that while results may be replicable in one lab, the results must also be replicable by different labs (Pashler & Wagenmakers, 2012). For others, the theoretical background is more important than the mere replication of a phenomenon (Stroebe & Strack, 2014). The current study shows that while results may be replicable and confirm a monolingual disadvantage, this result was dependent on statistical power (LeBel, Campbell, & Loving, 2017). Thus, from the perspective of bilingualism research, this result appears to tilt the methodological balance scales not towards replicability, but towards larger sample sizes. This is in accordance with Paap and his group (Paap, Johnson, & Sawi, 2016) who suggested that the statistical power of large samples is of the essence in bilingualism research because confounding variables cannot completely be accounted for (Gathercole, 2015).

However, another important take-away message from the current set of studies is that small samples can show results like under a microscope that cannot be demonstrated in the large sample because certain groups may show specific characteristics. Even a single case may give information about the possibilities of the human mind (Normand, 2016), e.g. a student who was not distractible at all in an experiment had a pianist mother who would practice all day at their home. Moreover, there is the Jackknife method which states that one could draw small samples from a large sample by randomly eliminating data sets and the results should stay the same (Meyer, Ingersoll, McDonald, & Boyce, 1986; Miller, 1974). It is assumed that testing a larger sample makes it more homogeneous so that ‘distortions’ in smaller samples are averaged out. This was also the case in the current study, and it is up to us to judge whether the results of smaller samples may give us information that would disappear in a larger population like a drop in the ocean. The question then is whether we are interested in particular mechanisms of the mind of smaller groups such as diverse ethnic minorities who together may nevertheless be in the majority in local urban environments.

Resistance to proactive interference via repetition

While the language effects due to individual differences in bilingualism disappeared in the large sample, the experimental effect of a build-up and release of PI was robust. The results support the conclusion that repetition is a forced and tested rehearsal that can lead to self-contained memory representations which counteract memory deterioration in verbal recall (Bäuml & Kliegl, 2013; Szpunar et al., 2008; Wahlheim, 2015). The mere list repetition could function like a Hebb effect that supports learning and consolidation in long-term memory (Mosse & Jarrold, 2008; Page & Norris, 2009; Szmalec, Duyck, Vandierendonck, Mata, & Page, 2009).

The role of rehearsal in pro-active interference has been investigated from early on. Schendel (1976) showed that during build-up and release of PI in an experimental condition using a general-word to number-word list transition, self-reported rehearsal was continuously reported via button presses. In contrast, in the number-word to general-word list transition condition, PI was still released, but verbal rehearsal had declined during build-up and only recovered at the release from PI. Hence, rehearsal was tied to the more varied word lists with less connected words, probably in an effort to create chunks. Grouping was also shown to be relevant for diverse memory items in the visual domain (Lange-Küttner & Küttner, 2015).

Release from PI does not only occur in short-term memory (Turvey, Cremins, & Lombardo, 1969; Turvey, Fertig, & Kravetz, 1969), it also occurs in an action context (Nilsson & Bäckman, 1991). For instance, a decline could be seen in children's performance IQ as measured by repeated assessments of the Draw-A-Person (DAP) test which counts the details that the child cares to remember to draw (Lange-Küttner, Küttner, & Chromekova, 2014). When the more general task to draw a person was given first, until age 11, the DAP IQ score deteriorated in the first three repetitions as they would draw nearly the identical figures, but the score recovered once a more specialized instruction to draw a police man (to the boys) or woman (to the girls) was given. However, when the more specific police task was given first and the person task second, the DAP IQ did not recover but continued to deteriorate.

With respect to words, familiarity and novelty in terms of word frequency have no effect on PI (Underwood, Broder, & Zimmerman, 1973; Underwood & Ekstrand, 1967). Instead, similar words in the lists are more conducive to the build-up of PI (Underwood, 1983) as lists can be more easily interwoven with each other (Underwood, 1982), and the more lists, the stronger the PI build up (Underwood & Ekstrand, 1967).

Pro-active interference appears to be an immediate effect in an experiment which does not occur when learning each list is separated by 24 hours (Underwood et al., 1973). This indicates that the immediate activation of the material plays a role in the build-up and release from PI. In the current study, the word material was rather general, and the eye gaze cues and the swapping of the places with the distracter on the screen during the repetition added some challenge. There is, however, a crucial difference between the release from PI due to a change in the taxonomy of items versus repetition and consolidation. While the release from PI due to category change is immediate and rapid, the release from PI due to the forced repetition of the material was shown to be of a more gradual nature. The elevated memory performance resulting from the list repetitions seemed to transfer into a more efficient memory strategy for participants with the result that the repetition gain became smaller for the last word list as the memory score itself had recovered. A practice effect on PI was also demonstrated in other recent research (Persson & Reuter-Lorenz, 2008; Wahlheim & Jacoby, 2011). Hence, if it is possible that humans could use a single repetition to optimize their memory strategies in such a way that build-up of PI can be overcome without taxonomically new material, this may be of much importance for future memory training studies.

References

Abel, M., & Bäuml, K.-H. T. (2014). Sleep can reduce proactive interference. *Memory, 22*(4), 332-339. doi:10.1080/09658211.2013.785570

Adesope, O. O., Lavin, T., Thompson, T., & Ungerleider, C. (2010). A systematic review and meta-analysis of the cognitive correlates of bilingualism. *Review of Educational Research, 80*(2), 207-245. doi:10.3102/0034654310368803

Baddeley, A., Gathercole, S., & Papagno, C. (1998). The phonological loop as a language learning device. *Psychological Review, 105*(1), 158-173. doi:10.1037/0033-295X.105.1.158

Bäuml, K.-H. T., & Kliegl, O. (2013). The critical role of retrieval processes in release from proactive interference. *Journal of Memory and Language, 68*(1), 39-53. doi:10.1016/j.jml.2012.07.006

Ben Zeev, S. (1977). The influence of bilingualism on cognitive strategy and cognitive development. *Child Development, 48*, 1009-1018. doi:10.2307/1128353

Beringer, J. (1994). Experimental Run Time System, version 3.0 and Experimental Run Time Coding System, version 2.20. Frankfurt, www.erts.de.

Bialystok, E., Craik, F. I. M., & Freedman, M. (2007). Bilingualism as a protection against the onset of symptoms of dementia. *Neuropsychologia, 45*(2), 459-464. doi:10.1016/j.neuropsychologia.2006.10.009

Bialystok, E., & DePape, A.-M. (2009). Musical expertise, bilingualism, and executive functioning. *Journal of Experimental Psychology: Human Perception and Performance, 35*(2), 565-574. doi:10.1037/a0012735

Bialystok, E., & Feng, X. (2009). Language proficiency and executive control in proactive interference: Evidence from monolingual and bilingual children and adults. *Brain and Language, 109*(2-3), 93-100. doi:10.1016/j.bandl.2008.09.001

Carretti, B., Mammarella, I., & Borella, E. (2012). Age differences in proactive interference in verbal and visuospatial working memory. *Journal of Cognitive Psychology, 24*(3), 243-255. doi:10.1080/20445911.2011.603695

Chen, C., Xue, G., Dong, Q., Jin, Z., Li, T., Xue, F., . . . Guo, Y. (2007). Sex determines the neurofunctional predictors of visual word learning. *Neuropsychologia, 45*(4), 741-747. doi:10.1016/j.neuropsychologia.2006.08.018

Cohen, J. S., & Armstrong, D. L. (1996). Sources of intertrial proactive interference in rats' short-term memory in a delayed successive matching-to-sample modality discrimination. *Learning and Motivation, 27*(4), 485-499. doi:10.1006/lmot.1996.0029

Cook, T. D., & Campbell, D. T. (1979). *Quasi-experimentation. Design & analysis issues for field settings*. Boston: Houghton Mifflin.

Costa, A., & Gallés, N. S. (2014). How does the bilingual experience sculpt the brain? *Nature Reviews Neuroscience, 15*(5), 336-345. doi:10.1038/nrn3709

Cowan, N., & Saults, J. S. (2013). When does a good working memory counteract proactive interference? Surprising evidence from a probe recognition task. *Journal of Experimental Psychology: General, 142*(1), 12-17. doi:10.1037/a0027804

Cowan, N., & Vergauwe, E. (2015). Applying how adults rehearse to understand how rehearsal may develop. *Frontiers in Psychology, 5*. doi:10.3389/fpsyg.2014.01538

Cox, S. R., Bak, T. H., Allerhand, M., Redmond, P., Starr, J. M., Deary, I. J., & MacPherson, S. E. (2016). Bilingualism, social cognition and executive functions: A tale of chickens and eggs. *Neuropsychologia, 91*(Supplement C), 299-306. doi:10.1016/j.neuropsychologia.2016.08.029

Craik, F. I. M., & Bialystok, E. (2010). Bilingualism and aging: Costs and benefits. In L. Bäckman & L. Nyberg (Eds.), *Memory, aging and the brain: A festschrift in honour of Lars-Göran Nilsson.* (pp. 115-131). New York, NY US: Psychology Press.

Cumming, N., Page, M., & Norris, D. (2003). Testing a positional model of the Hebb effect. *Memory, 11*(1), 43. doi:10.1080/741938175

Dillon, R., McCormack, P., Petrusic, W., Cook, G. M., & Lafleur, L. (1973). Release from proactive interference in compound and coordinate bilinguals. *Bulletin of the Psychonomic Society, 2*(5), 293-294. doi:10.3758/BF03329277

Driver, J., Davis, G., Ricciardelli, P., Kidd, P., Maxwell, E., & Baron-Cohen, S. (1999). Gaze perception triggers reflexive visuospatial orienting. *Visual Cognition, 6*(5), 509-540. doi:10.1080/135062899394920

Dunnett, S. B., & Martel, F. L. (1990). Proactive interference effects on short-term memory in rats: I. Basic parameters and drug effects. *Behavioral Neuroscience, 104*(5), 655-665. doi:10.1037/0735-7044.104.5.655

Dunnett, S. B., Martel, F. L., & Iversen, S. D. (1990). Proactive interference effects on short-term memory in rats: II. Effects in young and aged rats. *Behavioral Neuroscience, 104*(5), 666-670. doi:10.1037/0735-7044.104.5.666

Francis, W. S., & Sáenz, S. P. (2007). Repetition priming endurance in picture naming and translation: Contributions of component processes. *Memory & Cognition, 35*(3), 481-493. doi:10.3758/bf03193288

Friederici, A. D., Pannekamp, A., Partsch, C.-J., Ulmen, U., Oehler, K., Schmutzler, R., & Hesse, V. (2008). Sex hormone testosterone affects language organization in the infant brain. *Neuroreport, 19*(3), 283-286. doi:10.1097/WNR.0b013e3282f5105a

Friedman, N. P., & Miyake, A. (2004). The relations among inhibition and interference control functions: A latent-variable analysis. *Journal of Experimental Psychology: General, 133*(1), 101-135. doi:10.1037/0096-3445.133.1.101

Frith, U., & Vargha-Khadem, F. (2001). Are there sex differences in the brain basis of literacy related skills? Evidence from reading and spelling impairments after early unilateral brain damage. *Neuropsychologia, 39*(13), 1485-1488. doi:10.1016/S0028-3932(01)00063-X

Gathercole, V. C. (2015). Are we at a socio-political and scientific crisis? *Cortex: A Journal Devoted to the Study of the Nervous System and Behavior, 73*, 345-346. doi:10.1016/j.cortex.2015.07.022

Güvendir, E. (2013). Sex-related structural differences in language areas of the human brain and their implications for intergroup relations in ancestral groups. *Language Sciences, 40*, 174-182. doi:10.1016/j.langsci.2013.05.003

Hernandez, A. E., & Reyes, I. (2002). Within- and between-language priming differ: Evidence from repetition of pictures in Spanish-English bilinguals. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 28*(4), 726-734. doi:10.1037/0278-7393.28.4.726

Horst, J. S. (2013). Context and repetition in word learning. *Frontiers in Psychology, 4*, article 149. doi:10.3389/fpsyg.2013.00149

Horst, J. S., Parsons, K. L., & Bryan, N. M. (2011). Get the story straight: Contextual repetition promotes word learning from storybooks. *Frontiers in Psychology, 2*, article 17. doi:10.3389/fpsyg.2011.00017

Hull, R., & Vaid, J. (2007). Bilingual language lateralization: A meta-analytic tale of two hemispheres. *Neuropsychologia, 45*(9), 1987-2008. doi:10.1016/j.neuropsychologia.2007.03.002

Hulme, C., Thomson, N., Muir, C., & Lawrence, A. (1984). Speech rate and the development of short-term memory span. *Journal of Experimental Child Psychology, 38*(2), 241-253. doi:10.1016/0022-0965(84)90124-3

Kail, R. (2002). Developmental change in proactive interference. *Child Development, 73*(6), 1703-1714. doi:10.1111/1467-8624.00500

Kane, M. J., Conway, A. R. A., Miura, T. K., & Colflesh, G. J. H. (2007). Working memory, attention control, and the N-back task: a question of construct validity. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 33*(3), 615-622. doi:10.1037/0278-7393.33.3.615

Keppel, G., & Underwood, B. J. (1962). Proactive inhibition in short-term retention of single items. *Journal of Verbal Learning and Verbal Behavior, 1*(3), 153-161. doi:10.1016/S0022-5371(62)80023-1

Lange-Küttner, C. (2005). Word structure effects in German and British reading beginners. *Zeitschrift für Pädagogische Psychologie, 19*(4), 207-218. doi:10.1024/1010-0652.19.4.207

Lange-Küttner, C. (2012). The importance of reaction times for developmental science: What a difference milliseconds make. *International Journal of Developmental Science, 6*(1-2), 51-55.

Lange-Küttner, C. (2013). Array effects, spatial concepts, or information processing speed: What is the crucial variable for place learning? *Swiss Journal of Psychology, 72*(4), 197-217. doi:10.1024/1421-0185/a000113

Lange-Küttner, C., & Krappmann, L. (2011). Ein modalitätsspezifisches Wortgedächtnis ist ausreichend: Wortgedächtnismodalität bei Leseanfängern und Neuronalen Netzen (engl. Abstract, one modality-specific word memory is sufficient – word memory modality in reading beginners and neural networks). *Zeitschrift für Pädagogische Psychologie, 25*(1), 19-37. doi:10.1024/1010-0652/a000027

Lange-Küttner, C., & Küttner, E. (2015). How to learn places without spatial concepts: Does the what-and-where reaction time system in children regulate learning during stimulus repetition? *Brain and Cognition, 97*, 59-73. doi:10.1016/j.bandc.2015.04.008

Lange-Küttner, C., Küttner, E., & Chromekova, M. (2014). Deterioration and recovery of DAP IQ scores in the repeated assessment of the Naglieri Draw-a-person (DAP) test in 6- to 12-year-old children. *Psychological Assessment, 26*(1), 297-306. doi:10.1037/a0034581

Lange-Küttner, C., Puiu, A.-A., Nylund, M., Cardona, S., & Garnes, S. (2013). Speech preparation and articulation time in bilinguals and men. *International Journal of Speech & Language Pathology and Audiology, 1*(1), 37-42. doi:10.12970/2311-1917.2013.01.01.5

Lange-Küttner, C., & Sykorova, E. (2015). Mojibake – the rehearsal of word fragments in verbal recall. *Frontiers in Psychology, 6,* article 350. doi: 10.3389/fpsyg.2015.00350

Lavie, N. (2005). Distracted and confused? Selective attention under load. *Trends in Cognitive Sciences, 9*(2), 75-82. doi:10.1016/j.tics.2004.12.004

Lavie, N., Hirst, A., de Fockert, J. W., & Viding, E. (2004). Load theory of selective attention and cognitive control. *Journal of Experimental Psychology: General, 133*(3), 339-354. doi:10.1037/0096-3445.133.3.339

LeBel, E. P., Campbell, L., & Loving, T. J. (2017). Benefits of open and high-powered research outweigh costs. *Journal of Personality and Social Psychology, 113*(2), 230-243. doi:10.1037/pspi0000049

Leech, G., Rayson, P., & Wilson, A. (2001). *Word frequencies in written and spoken English: Based on the British National Corpus*. London: Longman.

Lehmann, M. (2015). Rehearsal development as development of iterative recall processes. *Frontiers in Psychology, 6*, article 308. doi:10.3389/fpsyg.2015.00308

Loess, H., & Waugh, N. C. (1967). Short-term memory and intertrial interval. *Journal of Verbal Learning & Verbal Behavior, 6*(4), 455-460. doi:10.1016/S0022-5371(67)80001-X

Mann, V. A., Sasanuma, S., Sakuma, N., & Masaki, S. (1990). Sex differences in cognitive abilities: A cross-cultural perspective. *Neuropsychologia, 28*(10), 1063-1077. doi:10.1016/0028-3932(90)90141-A

McKelvie, S. J. (1987). Learning and awareness in the Hebb digits task. *Journal of General Psychology, 114*(1), 75. doi:10.1080/00221309.1987.9711057

Mechelli, A., Crinion, J. T., Noppeney, U., O'Doherty, J., Ashburner, J., Frackowiak, R. S., & Price, C. J. (2004). Neurolinguistics: Structural plasticity in the bilingual brain. *Nature, 431*(7010), 757-757. doi:10.1038/431757a

Merrill, E. C., Yang, Y., Roskos, B., & Steele, S. (2016). Sex differences in using spatial and verbal abilities influence route learning performance in a virtual environment: A comparison of 6- to 12-year old boys and girls. *Frontiers in Psychology, 7*, article 258. doi:10.3389/fpsyg.2016.00258

Meyer, J. S., Ingersoll, C. G., McDonald, L. L., & Boyce, M. S. (1986). Estimating uncertainty in population growth rates: Jackknife vs. Bootstrap techniques. *Ecology, 67*, 1156-1166.

Miller, R. G. (1974). The jackknife - a review. *Biometrika, 61*, 1-15.

Morra, S., & Epidendio, V. (2015). Speeded probed recall is affected by grouping. *Experimental Psychology, 62*(6), 385-394. doi:10.1027/1618-3169/a000302

Mosse, E. K., & Jarrold, C. (2008). Hebb learning, verbal short-term memory, and the acquisition of phonological forms in children. *The Quarterly Journal of Experimental Psychology, 61*(4), 505-514. doi:10.1080/17470210701680779

Nilsson, L.-G., & Bäckman, L. (1991). Encoding dimensions of subject-performed tasks. *Psychological Research, 53*(3), 212-218. doi:10.1007/BF00941389

Normand, M. P. (2016). Less is more: Psychologists can learn more by studying fewer people. *Frontiers in Psychology, 7*, article 934. doi:10.3389/fpsyg.2016.00934

Oberauer, K., Awh, E., & Sutterer, D. W. (2016). The role of long-term memory in a test of visual working memory: Proactive facilitation but no proactive interference. *Journal of Experimental Psychology: Learning, Memory, and Cognition*. doi:10.1037/xlm0000302

Oller, D. K., & Jarmulowicz, L. (2007). Language and literacy in bilingual children in the early school years. In E. Hoff & M. Shatz (Eds.), *Blackwell handbook of language development* (pp. 368-386). Malden: Blackwell Publishing.

Olson, R. K., Davidson, B. J., Kliegl, R., & Davies, S. E. (1984). Development of phonetic memory in disabled and normal readers. *Journal of Experimental Child Psychology, 37*, 187-206.

Paap, K. R. (2015). Do many hones dull the bilingual whetstone? *Bilingualism: Language and Cognition, 18*(01), 41-42. doi:10.1017/S1366728914000431

Paap, K. R., Johnson, H. A., & Sawi, O. (2014). Are bilingual advantages dependent upon specific tasks or specific bilingual experiences? *Journal of Cognitive Psychology, 26*(6), 615-639. doi:10.1080/20445911.2014.944914

Paap, K. R., Johnson, H. A., & Sawi, O. (2015). Bilingual advantages in executive functioning either do not exist or are restricted to very specific and undetermined circumstances. *Cortex: A Journal Devoted to the Study of the Nervous System and Behavior, 69*, 265-278. doi:10.1016/j.cortex.2015.04.014

Paap, K. R., Johnson, H. A., & Sawi, O. (2016). Should the search for bilingual advantages in executive functioning continue? *Cortex, 74*, 305-314. doi:10.1016/j.cortex.2015.09.010

Paap, K. R., Myuz, H. A., Anders, R. T., Bockelman, M. F., Mikulinsky, R., & Sawi, O. M. (2017). No compelling evidence for a bilingual advantage in switching or that frequent language switching reduces switch cost. *Journal of Cognitive Psychology, 29*(2), 89-112. doi:10.1080/20445911.2016.1248436

Paap, K. R., & Sawi, O. (2014). Bilingual advantages in executive functioning: Problems in convergent validity, discriminant validity, and the identification of the theoretical constructs. *Frontiers in Psychology, 5*(962), article 2. doi:10.3389/fpsyg.2014.00962

Page, M. P. A., & Norris, D. (2009). Is there a common mechanism underlying word-form learning and the hebb repetition effect? Experimental data and a modelling framework. In A. Thorn & M. Page (Eds.), *Interactions between short-term and long-term memory in the verbal domain* (pp. 136-156). Hove, UK: Psychology Press.

Pashler, H., & Harris, C. R. (2012). Is the replicability crisis overblown? Three arguments examined. *Perspectives on Psychological Science, 7*(6), 531-536. doi:10.1177/1745691612463401

Pashler, H., & Wagenmakers, E. J. (2012). Editors' introduction to the special section on replicability in psychological science. *Perspectives on Psychological Science, 7*(6), 528-530. doi:10.1177/1745691612465253

Persson, J., & Reuter-Lorenz, P. A. (2008). Gaining control: Training executive function and far transfer of the ability to resolve interference. *Psychological Science, 19*(9), 881-888. doi:10.1111/j.1467-9280.2008.02172.x

Pfeiffer, U. J., Vogeley, K., & Schilbach, L. (2013). From gaze cueing to dual eye-tracking: Novel approaches to investigate the neural correlates of gaze in social interaction. *Neuroscience & Biobehavioral Reviews, 37*(10, Part 2), 2516-2528. doi:10.1016/j.neubiorev.2013.07.017

Rahimi-Golkhandan, S., Maruff, P., Darby, D., & Wilson, P. (2012). Barriers to repeated assessment of verbal learning and memory: A comparison of international shopping list task and Rey auditory verbal learning test on build-up of proactive interference. *Archives of Clinical Neuropsychology, 27*(7), 790-795. doi:10.1093/arclin/acs074

Reutener, D. B. (1972). Background, symbolic, and class shift in short-term verbal memory. *Journal of Experimental Psychology, 93*(1), 90-94. doi:10.1037/h0032504

Roodenrys, S., Hulme, C., & Brown, G. (1993). The development of short-term memory span: Separable effects of speech rate and long-term memory. *Journal of Experimental Child Psychology, 56*(3), 431-442. doi:10.1006/jecp.1993.1043

Schaadt, G., Hesse, V., & Friederici, A. D. (2015). Sex hormones in early infancy seem to predict aspects of later language development. *Brain and Language, 141*, 70-76. doi:10.1016/j.bandl.2014.11.015

Schendel, J. D. (1976). Analysis of the rehearsal processes in the release from proactive interference. *Journal of Experimental Psychology: Human Learning and Memory, 2*(1), 76-82. doi:10.1037/0278-7393.2.1.76

Schleepen, T. M. J., & Jonkman, L. M. (2010). The development of non-spatial working memory capacity during childhood and adolescence and the role of interference control: An n-back task study. *Developmental Neuropsychology, 35*(1), 37-56. doi:10.1080/87565640903325733

Stroebe, W., & Strack, F. (2014). The alleged crisis and the illusion of exact replication. *Perspectives on Psychological Science, 9*(1), 59-71. doi:10.1177/1745691613514450

Szmalec, A., Duyck, W., Vandierendonck, A., Mata, A. B., & Page, M. P. A. (2009). The Hebb repetition effect as a laboratory analogue of novel word learning. *The Quarterly Journal of Experimental Psychology, 62*(3), 435-443. doi:10.1080/17470210802386375

Szpunar, K. K., McDermott, K. B., & Roediger, H. L., III. (2008). Testing during study insulates against the buildup of proactive interference. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 34*(6), 1392-1399. doi:10.1037/a0013082

Szpunar, K. K., McDermott, K. B., & Roediger, H. L., III. (2009). 'Testing during study insulates against the buildup of proactive interference': Correction. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 35*(1), 156-156. doi:10.1037/a0014896

Thomas, D. R., Burr, D. S., & Vogt, M. B. (1982). Proactive interference in the retention of 'attentional sets' in pigeons. *Animal Learning & Behavior, 10*(1), 15-21. doi:10.3758/BF03212041

Turvey, M. T., Cremins, J. J., & Lombardo, T. (1969). Taxonomic categories and proactive interference in short-term memory. *Psychonomic Science, 15*(6), 307-308.

Turvey, M. T., & Egan, J. (1969). Contextual change and release from proactive interference in short-term verbal memory. *Journal of Experimental Psychology, 81*(2), 396-397. doi:10.1037/h0021177

Turvey, M. T., Fertig, J., & Kravetz, S. (1969). Connotative classification and proactive interference in short-term memory. *Psychonomic Science, 16*(4), 223-224. doi:10.3758/BF03336391

Underwood, B. J. (1945). The effect of successive interpolations on retroactive and proactive inhibition. *Psychological Monographs, 59*(3), i-33. doi:10.1037/h0093547

Underwood, B. J. (1982). Proactive interference and the simultaneous acquisition retention phenomenon. *Journal of Verbal Learning & Verbal Behavior, 21*(2), 142-149. doi:10.1016/S0022-5371(82)90511-4

Underwood, B. J. (1983). 'Conceptual' similarity and cumulative proactive inhibition. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 9*(3), 456-461. doi:10.1037/0278-7393.9.3.456

Underwood, B. J., Broder, P. K., & Zimmerman, J. (1973). Retention of verbal discrimination lists as a function of number of prior lists, word frequency, and type of list. *Journal of Experimental Psychology, 100*(1), 101-105. doi:10.1037/h0035515

Underwood, B. J., & Ekstrand, B. R. (1967). Word frequency and accumulative proactive inhibition. *Journal of Experimental Psychology, 74*(2, Pt.1), 193-198. doi:10.1037/h0024589

Vaid, J., & Genesee, F. (1980). Neuropsychological approaches to bilingualism: A critical review. *Canadian Journal of Psychology, 34*(4), 417-445. doi:10.1037/h0081108

Valian, V. (2015). Bilingualism and cognition. *Bilingualism: Language and Cognition, 18*(01), 3-24. doi:10.1017/S1366728914000522

Wahlheim, C. N. (2015). Testing can counteract proactive interference by integrating competing information. *Memory & Cognition, 43*(1), 27-38. doi:10.3758/s13421-014-0455-5

Wahlheim, C. N., & Jacoby, L. L. (2011). Experience with proactive interference diminishes its effects: Mechanisms of change. *Memory & Cognition, 39*(2), 185-195. doi:10.3758/s13421-010-0017-4

Watkins, O. C., & Watkins, M. J. (1975). Buildup of proactive inhibition as a cue-overload effect. *Journal of Experimental Psychology: Human Learning and Memory, 1*(4), 442-452. doi:10.1037/0278-7393.1.4.442

Wickens, D. D. (1970). Encoding categories of words: An empirical approach to meaning. *Psychological Review, 77*(1), 1-15. doi:10.1037/h0028569

Wickens, D. D., Born, D. G., & Allen, C. K. (1963). Proactive inhibition and item similarity in short-term memory. *Journal of Verbal Learning and Verbal Behavior, 2*(5), 440-445. doi:10.1016/S0022-5371(63)80045-6

Wilkie, D. M. (1986). Pigeons' spatial memory: V. Proactive interference in the delayed matching of key location paradigm occurs only under restricted conditions. *Animal Learning & Behavior, 14*(3), 257-266. doi:10.3758/BF03200066

Woumans, E., Santens, P., Sieben, A., Versijpt, J., Stevens, M., & Duyck, W. (2015). Bilingualism delays clinical manifestation of Alzheimer's disease. *Bilingualism: Language and Cognition, 18*(3), 568-574. doi:10.1017/S136672891400087X