Ching Chu Sun, Peter Hendrix, Harald Baayen and Michael Ramscar

2 The price of knowledge: A bilingual paired associate learning study

Abstract: Age effects in experimental psychology are typically interpreted as evidence for cognitive decline. Alternatively, age-related decreases in performance on cognitive tasks could be a result of increased linguistic experience (Ramscar et al. 2014). We present the results of a paired associate learning experiment in which we tested old and young German monolinguals and Chinese-German bilinguals. Younger participants performed similarly in L1 and L2. Older participants performed better in L2 than in L1. The current findings cannot be accounted for by cognitive decline, but follow straightforwardly from basic principles of learning.

Keywords: paired associate learning, aging, bilingual, NDL, discrimination learning

1 Introduction

As adults age, their performance on various measures of cognition – such as those testing reasoning, memory, and response speeds – changes, with scores on many tests declining. Two obvious explanations suggest themselves for this. The first takes these changes as evidence of a loss of cognitive function over time (Naveh-Benjamin and Old 2008; Deary et al. 2009; Salthouse 2009; Singh-Manoux et al. 2012) advocating the view that, "the sad truth is that even normal aging has a devastating effect on our ability to learn and remember, on the speed with which we process information, and on our ability to reason" (Epstein 2012). The other posits that these changes reflect the increased information processing demands that accompany greater learning from experience and the failure of cognitive measures to control for this (Ramscar et al. 2013; Ramscar et al. 2014).

It is clearly important to understand whether or how much cognitive abilities decline in adulthood. Attempts to discern which account is correct, however, are forced to struggle with a seemingly inevitable confound: older adults have more experience than younger adults in almost any cognitive domain. In what follows, we de-confound the association between age and experience by testing the

Ching Chu Sun, University of Tübingen, ching-chu.sun@uni-tuebingen.de
Peter Hendrix, Harald Baayen and Michael Ramscar, University of Tübingen

https://doi.org/10.1515/9783110610895-006

same cognitive ability (Paired Associate Learning; henceforth PAL) in native (L1) speakers of German and Chinese and second-language (L2) speakers of German. Whereas native speakers' experience is confounded with age, this is much less the case for L2 speakers' experience. If the cognitive abilities underlying PAL performance decline with age, then we ought to expect to see the same patterns of performance differences between older and younger participants in both L1 and L2. However, if PAL performance simply reflects experience with a language, we ought to see better performance for older speakers in L2 than in L1.

1.1 Paired associate learning

Paired Associate Learning is a common psychometric measure of people's ability to learn and recall new information. In standard verbal versions of the test such as the PAL subtest of Wechsler's Memory Scale (WMS) (des Rosiers and Ivison 1986), participants hear pairings between words that act as cue items (e.g., *baby*; *jury*) and words that are response items (*cries*; *eagle*). Participants listen to lists of the pairings, and then supply the response to each cue at test. Although performance on the individual pairs in the test varies, performance becomes progressively slower and less accurate as age increases.

These changes in PAL test performance can only be interpreted as evidence of declining cognitive performance if one assumes that the functioning of the cognitive processes engaged by the PAL task can be estimated from exposure to a uniform association rate for each test item (i.e., participants hear each w_1-w_2 pair the same number of times). However, convergent results from a huge body of empirical work (Miller et al. 1995; Siegel and Allan 1996; Ramscar et al. 2010), as well as human neuroscience findings (McDannald et al. 2014, d'Acremont et al. 2009; Schultz 2006; Schultz et al. 1997), animal studies (Yin and Knowlton 2006; Tremblay et al. 1998; Kamin 1969; Rescorla 1968) and computational and mathematical models (Dayan and Berridge 2014; Daw et al. 2008; Sutton and Barto 1981; Pearce and Hall 1980; Rescorla and Wagner 1972) show that association rates (the rate at which PAL items are encountered together) are insufficient to explain the patterns of behavior produced by associative learning.

Rather, what is traditionally called associative learning has been shown to be a discriminative process that detunes uninformative and reinforces informative dimensions within a system of inputs in order to minimize future prediction error (see Ramscar et al. 2010 for a review). As such, apart from association rates, at least two other quantitative factors must be taken into account in predicting and assessing the performance of this system. First, the background rates of cue words (see Rescorla 1968; Ramscar, Dye, and Klein 2013) co-determine the predictability of response words. The background rate of the cue word is the frequency with which it appears in the absence of a response word. When association rates are held constant, cue words with higher background rates are less informative for response words than cue words with lower background rates. The greater the frequency of the cue word, therefore, the harder it is to learn the association with the response word (Ramscar, Dye, and McCauley 2013). A second factor that needs to be taken into account is blocking (see Kamin 1969; Arnon and Ramscar 2012). Blocking refers to the principle that once a learner is able to accurately predict an outcome, the need to learn associations between additional cues and that outcome no longer exists (see Rescorla 1968; Ramscar, Dye, and Klein 2013). In the context of paired associate learning, greater predictability of an outcome word given a cue word based on prior learning thus makes it harder to learn a cue-outcome pair in the PAL task.

Moreover, the skewed distribution of language implies that the relative influence of these factors will inevitably change with experience (Ramscar et al. 2013; Ramscar et al. 2014). To demonstrate how this can be expected to influence PAL performance over time, Figure 1 simulates the development of associations between items in a very simple model lexicon comprising two easy (*North - South*; *Cat - Dog*) and two hard (*Banana - Dog*; *North - Dog*) PAL pairs using the Danks equilibrium equations (Danks 2003) for the Rescorla-Wagner model (Wagner & Rescorla 1972), as implemented in the ndl package for R (Arppe et al. 2014). To show that learnability cannot be predicted from a fixed association rate, the association rates of the hard items were held constant while those for the easy items were varied. As can be seen in Figure 1, as the frequencies of the easy phrases and the words that comprise them increases, the expected association between *North* and *Dog* after 1 exposure declines as the frequency of the easy items increases (see Ramscar et al. 2013).

One straightforward consequence of the nature of learning and the long tail of linguistic distributions is evident in even this very simple model: linguistic experience will tend to increase background rates relative to association rates, making PAL learning harder. It will also tend to disfavor the learning of PAL pairs comprising words that co-occur infrequently in language more as compared to those that co-occur frequently. Interestingly, this pattern is actually evident in empirical studies of PAL learning.

Figure 2 plots the mean performance by item for 100 older (age 40–49) and 100 younger (20–29) adults (50% females in each group) tested in a normative study of the WMS-PAL subtest (desRosiers and Ivison 1986). As can clearly be seen, the decline in the performance of the 40-something adults is far greater on the hard (low co-occurrence) items than on the easy (high co-occurrence) items. This is a direct result of increased discrimination between the uninformative hard pairings and the informative easy pairings, and occurs as a result of the actual function of the associative learning system, as opposed to a decline in its functionality.

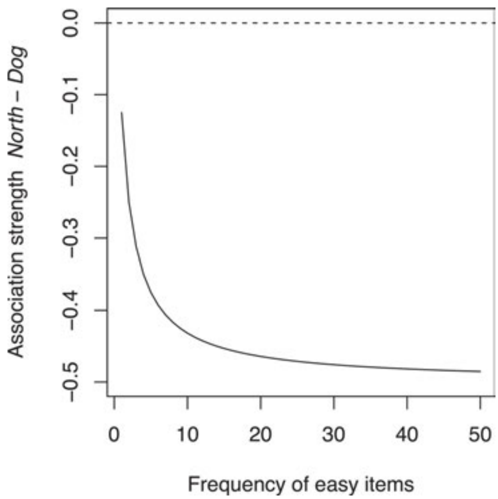

Figure 1: The association strength for *North* to *Dog* given one trial of training as a function of the frequency of two easy (common) associations *North* - *South* and *Cat* - *Dog*. The frequency of the two hard (uncommon) pairings *Banana* - *Dog* and *North* - *Dog* is always 1. When learning is simulated using the (Danks 2003) equilibrium equations for the (Wagner and Rescorla 1972) model, the association weight between *North* and *Dog* declines as the easy pairs' frequencies increase, even though both the structure of the lexicon and the association rate of *North* - *Dog* remain unchanged (adapted from Ramscar et al. 2013).

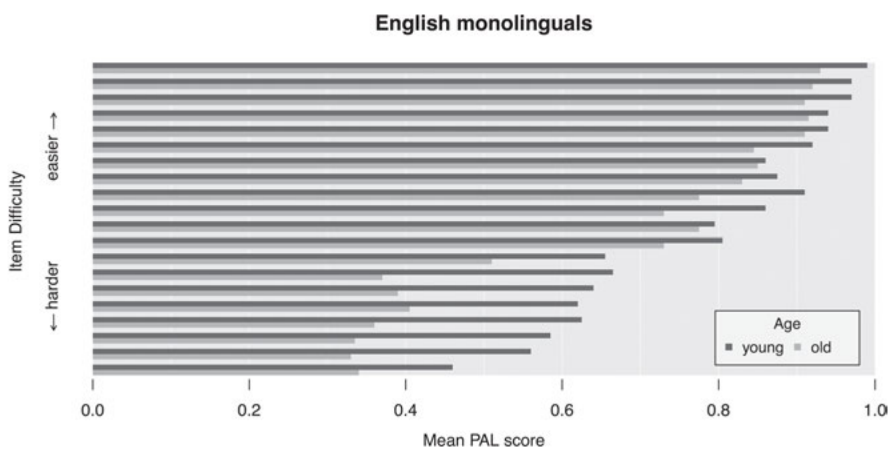

Figure 2: By-item performance on forms 1 and 2 of the WMS-PAL subtest for 100 younger (20–29) and 100 older (age 40–49) participants (desRosiers and Ivison 1986). The younger adults outperform older adults on all items, and performance differences are greater for the hard items than for the easy items.

A re-analysis of the full desRosiers and Ivison (1986) data set with 20–29, 30–39, 40–49, 50–59 and 60–69 year-old adults using a (beta regression) generalized additive mixed-effect model yields not only a random effect for Item (χ^2 = 470.577, p < 0.001) and a main effect of Gender (females perform better than males, z = −4.952, p < 0.001), but also significant interactions of Age by Co-occurrence Frequency (i.e., association rate: the number of times w_1 and w_2 appear next to each other in Google documents, χ^2 = 88.716, p < 0.001) and Age by Cue Frequency (i.e., the frequency of the cue word, χ^2 = 44.027, p < 0.001). These effects are presented in Figure 3.

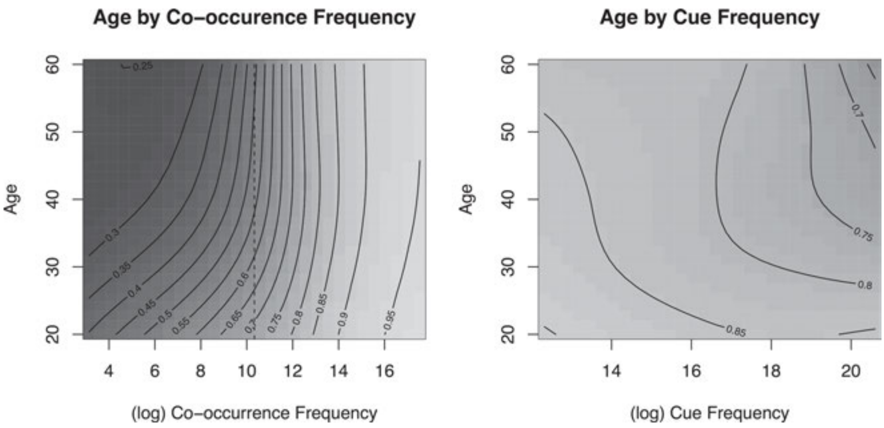

Figure 3: Effects of the (log-transformed) Co-occurrence Frequency by Age (left panel) and the (log-transformed) Cue Frequency by Age (right panel) on the Paired Associate Learning performance for English monolinguals. Color coding indicates proportion of correct responses in the paired associate learning test. The dotted line in the left panel indicates the mid-point of the (log) Co-occurrence Frequency range.

The left panel of Figure 3 shows the Age by Co-occurrence Frequency interaction. Consistent with the predictions of discrimination learning, lower co-occurrence frequencies lead to decreased performance in the PAL task. Furthermore, the sensitivity to the association rate between w_1 and w_2 increases as a function of age. Whereas the difference in performance for items with high co-occurrence frequencies is small, at the midpoint of the Co-occurrence Frequency range the predicted performance for the youngest participants is 72%, whereas the predicted performance for the oldest participants is 49% only (difference: 23%, see dotted line in the left panel of Figure 3).

The effect of Age by Cue Frequency (right panel of Figure 3) is more subtle in nature. Nonetheless, it is clear that PAL performance decreases as a function of the frequency of the cue word and does so to a greater extent for older participants than for younger participants. The re-analysis of the desRosiers and Ivison (1986) data therefore indicates that adults' PAL performance becomes ever more closely aligned with the informativity of w_1 and w_2 in their native language (i.e., the distributional information determining w_1-w_2 learnability) as their age increases.

1.2 Paired associate learning in L1 versus L2

Attempts to separate the effects of age and learning run into an obvious confound in that age and experience are highly correlated. Moreover, the nature of human and animal learning systems (in which, as we emphasized above, error plays a critical role; see Wagner and Rescorla 1972; Ramscar et al. 2010; Schultz 2006) means that this problem cannot be solved by designing novel tasks and stimuli, simply because it is impossible to design a task or stimulus that engages only novel sensory and behavioral dimensions.

PAL tests suggest an alternative approach to this problem because languages are structured, quantifiable systems that afford their speakers training in some of their aspects in ways that are very predictable. For example, although two English speakers of different ages with similar experience of German will differ in semantic experience (while both may have equal exposure to the German word *Brief* ('letter'), an older speaker will likely have more experience with letters and the word *letter*; see Arnon and Ramscar 2012), their experience with the distributional structure of the L2 (how often *Brief* occurs with other German words) will be very similar. More importantly, 45-year-old native English speakers with 10 years experience of German will have had less experience of the word *Brief* than the word *letter*, and their experience of the associative properties of the German lexicon will have resulted in them learning to dissociate the word *Brief* from *Herbst* less than their experience of the English lexicon will have resulted in them learning to dissociate the word *letter* from *autumn*. Accordingly, because there is less prior learning to overcome, we might expect that all other things being equal, learning the L2 PAL pair *Brief* - *Herbst* ought to be easier than learning the L1 pair *letter* - *autumn*.

However, all other things are not equal: underlining the point we made earlier about background rates in apparently novel items, because German and English are related languages, they share phonetic, orthographic, lexical and other features at various levels of systematic abstraction. Most obviously, *brief* is also a word in English (and similarly, although *Herbst* is not an English word, *herbs* is). Further, while *brief* and *letter* do not have a common orthography, *Butter* and *butter* do. And while *Milch* and *milk* are orthographically distinct, they barely differ phonetically and semantically.

It follows that our predictions about PAL learnability in L1 and L2 can better be tested in two languages in which the degrees of phonetic, orthographic, and lexical overlap between L1 and L2, and any systematicity therein, can be minimized. To this end, in the following experiment, we contrasted German (a non-tonal, West Germanic language that derives most of its lexicon from the Germanic branch of the Indo-European family of languages) with Mandarin Chinese (a tonal language that is a member of the Sino-Tibetan language family).

If PAL tests are a straightforward measure of learning and memory capacity (or, more opaquely, 'cognition', see Lindenberger 2014), either PAL performance for older participants should be similar in L1 and L2, or else we might find better performance in L1 than in L2 due to greater experience with the language. By contrast, basic principles of learning theory predict that older participants should perform better in L2 than in L1.

2 Methods

2.1 Participants

Four groups of participants took part in the experiment: young Chinese-German bilinguals, old Chinese-German bilinguals, young German monolinguals and old German monolinguals. Young participants were 18 to 28 years old, while old participants were 38 to 53 years old.[1] The age range for the older participants was set to 38 to 53 years old for two reasons. First, as can be seen in the left panel of Figure 3, the strongest age-related decline in performance in the normative PAL data desRosiers and Ivison (1986) takes place between the ages of 20 and 45. After the age of 45, the decline in performance is minimal. The age of the older participants in our study, therefore, is high enough to show the typical age effects observed in the PAL task. Second, increasing the age of the older participants in our study would increase the probability of including participants with undiagnosed mental diseases, such as dementia. While the effects of such diseases on cognitive functioning are interesting, they are outside the scope of the current study, which focuses on normal, healthy aging. After excluding participants with insufficient vocabularies in their second language (see below) and otherwise non-useable participants from the data,

1 Note that our use of the terms "old" and "older" deviates from the typical use of these words in the aging literature, which typically defines older adults as 65+ years old. The use of these terms throughout this paper refers to the relative age of the 38 to 53 years old participants as compared to the 18 to 28 year old participants, and is used for ease of reference.

Table 1: By-group age and vocabulary scores for each of the 4 groups of participants. Standard deviations are provided in brackets.

		Age	German Vocabulary	Chinese Vocabulary
Chinese-German bilinguals	young	24.55 (2.27)	31.75 (5.35)	67.65 (6.46)
	old	43.60 (4.66)	40.25 (7.86)	64.65 (7.09)
German monolinguals	young	23.45 (3.06)	81.95 (6.25)	– (–)
	old	44.90 (4.36)	84.10 (4.38)	– (–)

20 participants remained for each group. The mean age and the average vocabulary score in both languages for the 20 participants in each group is shown in Table 1.

The harder items in the PAL task contained relatively infrequent words. To ensure that participants knew all the words used in the test, we conducted vocabulary tests prior to the PAL task in both languages to assess the linguistic competence of participants in each language. For a more detailed description of these vocabulary tests, see the Materials section.

The performance on the vocabulary test in German was highly variable, particularly for the young Chinese-German bilinguals. Participants in this group were typically graduate students at the University of Tübingen. While some of these students were highly competent in their L2, others had considerably less experience. A number of these less experienced L2 learners explicitly stated that they did not know some of the words in the German PAL test.

To minimize the risk of young Chinese-German bilingual participants not knowing words in the PAL test, we selected from the group of 34 young Chinese-German bilinguals we were able to recruit in the Tübingen area the 20 participants with the best performance on the vocabulary pretest in German. Vocabulary test performance was calculated as a weighted sum of the number of correct answers, with item weights being the proportion of correct answers for the items across the young and old bilingual participants. This procedure has the added advantage that the difference in German vocabulary scores between old and young bilingual participants is smaller than it would otherwise be.[2]

On average, old Chinese-German bilinguals had somewhat more experience in their second language than did young Chinese-German bilinguals. The problem of insufficient L2 proficiency therefore proved much less prominent for old

2 The German PAL performance of the Chinese-German bilinguals with the worst vocabulary scores was worse than that of the 20 participants included in the analysis below (proportion correct: 0.64 versus 0.78). This suggests that at least some participants with poor vocabulary scores indeed did not know some of the items in the German PAL test.

Chinese-German bilinguals than for young Chinese-German bilinguals. Nonetheless, for consistency with the selection criterion used for young Chinese-German bilinguals, we excluded from the old Chinese-German bilinguals those participants with vocabulary scores lower than that of the 20th best young Chinese-German bilingual. This resulted in the exclusion of 4 old Chinese-German bilinguals.

Apart from insufficient proficiency in the second language, a total of 5 participants across the 4 groups were excluded for not meeting the requirements outlined in the experiment advertisement (4 participants did not attend university, 1 "German monolingual" was not a native speaker of German). Due to a shortage of old German participants that met the requirements outlined in the experiment advertisement, the final set of 20 old German participants contains one participant who did not attend university.

2.2 Materials

For both languages we administered a paired associate learning test as well as a vocabulary test. The vocabulary tests for both German and Chinese consisted of 100 multiple choice questions with 4 possible answers. The 3 incorrect answers were chosen from the same part-of-speech category as the correct answer. An example item for German, for instance, is the test word *Hemd* ('shirt'), with the four possible answers *Shirt* ('shirt'), *Jacke* ('jacket'), *Pullover* ('sweater') and *Weste* ('vest'). An example item for Chinese is *暮色* ('twilight'), with the four possible answers *浓雾* ('thick fog'), *黄昏* ('dusk'), *清晨* ('morning') and *月亮* ('moon').

The word frequency distributions for the German and Chinese vocabulary tests were matched. Test words on the German vocabulary test ranged in frequency from 10 to 0.001 per million in a 9 billion word corpus of German web pages. For Chinese, we did not have a similarly large corpus at our disposal. We therefore first selected words with a frequency ranging from 10 to 0.2 per million from the 5 million word Taiwan Sinica Corpus (CKIP 2014). For these words we also obtained Google search frequencies. We then used the median Google-to-Sinica frequency ratio and a list of Google frequencies for low frequency words in Chinese to complete the list of test items for the Chinese vocabulary test.

The PAL test in both languages consisted of three groups of 10 pairs. Pairs ranged in difficulty from easy (e.g., *Nord* - *Süd* ('north' - 'south') or *学校* - *读书* ('school' - 'study')) to hard (e.g., *Schiff* - *Puppe* ('ship' - 'doll') or *洋葱* - *手指* ('onion' - 'finger')). The anticipated difficulty of an item was gauged through the co-occurrence frequency of the words in a pair (see below for details). The first group contained the easiest items, the second group contained items with medium difficulty, and the third group contained the most difficult items. Words occurred no more than once in each of the paired associate learning tests. None of the words used in the paired associate learning tests were used as test words or answer alternatives in the vocabulary tests. The full list of the items used in the PAL tests in German and Chinese is presented in Table 2 (German) and Table 3 (Chinese) in Appendix 1.[3]

For each of the 30 pairs, we obtained the Co-occurrence Frequency of the words in that pair through the number of Google documents in which these words occurred together. Furthermore, we extracted the Google search unigram frequencies for all words in the PAL test as a measure of the background rates of both words. PAL items were designed such that the Cue Frequency, Response Frequency and Co-occurrence Frequency distributions for each language were approximately normal. An exact matching of the item difficulty in the German and Chinese PAL tests was impossible, given the fact that the exact size of Google in Chinese and German is hard to determine (or, more generally speaking, the unavailability of a sufficiently large corpus with a known corpus size for Chinese). Nonetheless, we matched items in the German and Chinese PAL tests for item difficulty as well as possible using average frequency conversion ratio’s based on Google frequency counts for sample words in both languages. All frequency measures were log-transformed prior to analysis to remove a rightward skew from the frequency distributions.

2.3 Design

The vocabulary pre-test for each language consisted of 100 multiple-choice questions with 4 possible answers. The order of the answers was randomized for each question, but consistent between participants. The paired associate learning task for each language consisted of three blocks of 10 pairs. The order of the blocks was held constant between participants, with the block of easy items being administered first and the block of hard items being administered last. Similarly, the order of the items was held constant between participants. The items appeared in the same order in the training phrase and in the test phase. Vocabulary pre-tests were administered prior to the paired associate learning test in each language. For the Chinese-German bilinguals the vocabulary pre-test and the paired associate learning test in Chinese preceded both tests in German.

3 Three semantic concepts, “dog”, “city” and “to swim”, occurred in both the Chinese PAL test and the German PAL test. These concepts appeared in different pairs in both languages, which should make pairs including these concepts somewhat harder the second time around. Given that for Chinese-German bilinguals the PAL test in Chinese preceded the PAL test in German, the repetition of these three semantic concepts therefore biases against our hypothesis that participants should perform better in the second language.

The dependent variable is the correctness of the response in the paired associate learning task. Paired associate responses were scored as correct when the response was either the target word or a member of the target word's morphological paradigm. For the target word *Blume* ('flower'), for instance, the plural form *Blumen* ('flowers') was considered a correct response. Similarly, for the target word *Wärme* ('warmth'), the response *warm* ('warm') was scored as correct. Furthermore, because of the differences between the phonology of Mandarin Chinese and German, some pronunciations for German target words by bilingual participants resembled phonological neighbors of the target word in German. For the item *Tanz* - *Feld*, for instance, it was hard to distinguish the target response/*fɛlt*/from pronunciations such as/*fɛrt*/or/*fert*/(pronunciation of *Pferd* ('horse') for native speakers of German in northern and central Germany) and/*felt/*(pronunciation of *fehlt* (3rd person singular of 'to lack')). For these types of phonologically ambivalent responses we followed the original scoring by a Chinese-German bilingual at the time of acquisition (i.e., the test phase of the experiment). We corrected the scoring at the time of acquisition from correct to incorrect for two responses to the word pair *Tanz* - *Feld* (one young bilingual, one old bilingual). In these cases the initial "pf" cluster in the pronunciation/*pfert*/indicated that a participant intended to pronounce *Pferd* rather than *Feld*.[4]

A number of predictors were included as independent variables in the design. For each participant the Age (numerical), Gender (categorical: male, female) and Education (categorical: non-PhD, PhD), as well as the result of the vocabulary pre-test in each language (numerical, 0–100) was included as a predictor. We also included a binary variable, In Second Language, which was set to 0 if an item was administered in the native language and to 1 if the item was administered in the second language. In addition, the Cue Frequency, the Response Frequency and the Co-occurrence Frequency as described in the materials section above were included as predictors. Finally, we included the order of an item in a list (i.e., in a block of 10 pairs) and the order of an item within the experiment as a whole as experimental control variables.[5]

4 Note that one could argue that even if a participant pronounces *Pferd* this is indicative of successful learning, because it suggests that a participant misperceived *Feld* as *Pferd*, but correctly remembered the perceived pair *Tanz* - *Pferd*. The difficulty of this pair is similar to that of *Tanz* - *Feld* (log co-occurrence frequency: 8.23 versus 7.04). Nonetheless, we decided to be conservative and score as incorrect the two responses mentioned above for which the initial "pf" cluster in the pronunciation indicated that the intended pronunciation was *Pferd* rather than *Feld*.

5 The correlation between the order of an item in a list and the order of the item in an experiment is $r = 0.33$, and is therefore unlikely to result in suppression in statistical models (see Wurm and Fisicaro 2014).

2.4 Procedure

The experiment was presented through a web page interface on a 15 inch MacBook Pro laptop. Items in both the vocabulary pretest and the paired associate learning test were presented auditorily through the laptop speakers. The items for the vocabulary and paired associate learning tests in both languages were recorded from native female speakers of Mandarin Chinese and German in a sound booth using professional recording equipment.

For each item in the vocabulary pre-test, participants were auditorally presented with the test word and the 4 possible answers. Participants were asked to select the answer that was most similar in meaning to the test word by clicking one of four buttons labelled 1 through 4 on the screen. Participants were asked to guess if they did not know the correct answer to a question.

In the paired associate learning task each block of 10 pairs consisted of a training phase and a test phase. Participants were asked to memorize the pairs of words presented in the training phase. In the test phase, participants were asked to produce the word that formed a pair with the auditorily presented word. The order of the words in a pair was consistent between the training and the test phase, such that the first word of a pair that was presented during the training phase was the auditorily presented "cue" word during the test phase. The test phase was self-paced: participants were asked to press the next button on the screen to move on to the next test word after verbally responding to a test word.

The average time required to complete the vocabulary pre-test for each language was about 30 minutes. The 3 blocks of paired associate learning took about 25 minutes per participant in each language, including a short break between each block. Including instructions and breaks, the duration of the experiment was about 1 hour for German monolinguals and 2 hours for Chinese-German bilinguals.

3 Analysis

The performance in the paired associate learning test was evaluated using a logistic generalized additive mixed-effect model (GAMM), as implemented in version 1.8–3 of the mgcv package for R (Wood 2006). The use of GAMMs allowed us to model non-linear predictor effects while also accounting for random effects.

We included a random intercept for Item, as well as random by-participant smooths for the order of an item in a list and in the experiment as a whole in the model to control for subject-, item- and task-related variance (cf., Baayen et al. 2015). The effects of categorical predictors were modeled through simple parametric terms, whereas the effects of numerical predictors were modeled through predictor smooths to allow for non-linear effects. R code for the reported GAMM is presented in Appendix 2.

4 Results

The GAMM analysis of the data revealed a significant random effect of Item ($\chi^2 = 310.673$, $p < 0.001$), as well as significant by-participant smooths for the Item Position in a list ($\chi^2 = 47.584$, $p < 0.001$) and the Item Position in the experiment as a whole ($\chi^2 = 138.523$, $p < 0.001$). The by-participant smooths for the Item Position in a list (left panel) and in the experiment (right panel) are presented in Figure 4.

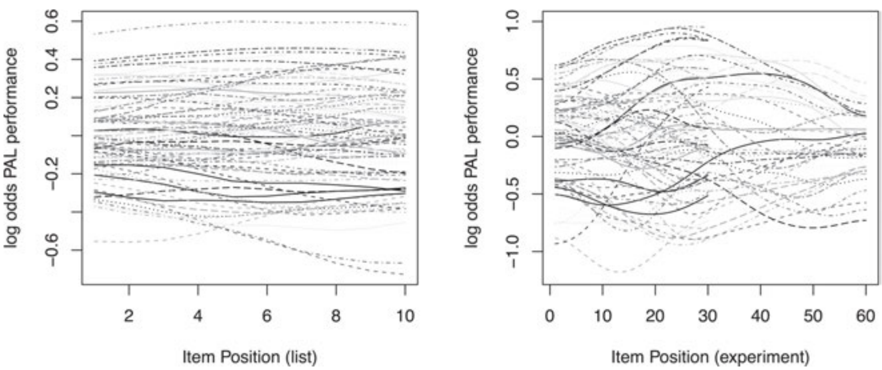

Figure 4: Penalized participant factor smooths for the position of an item in a block (left panel) and in the experiment as a whole (right panel). The effects of item position within a list and within the experiment show considerable variation between participants.

The left panel of Figure 4 demonstrates there is some between-participant variance for the effect of the order of an item in a list, with primacy effects for some participants and recency effects for others. The by-participant smooths for the position of an item within the experiment show more substantial variation. Some participants improve over the course of the experiment, whereas others become worse.

In addition, the GAMM analysis revealed a marginally significant main effect of Education ($z = -1.758$, $p = 0.079$), with PhDs performing somewhat worse than non-PhDs. This main effect of Education Level was significant in a post-hoc analysis in which we included the data for the old participants only (z = –2.073, p = 0.038). The poor performance of highly educated older participants follows straightforwardly from the principles of discrimination learning: the greater the experience of a person with the language, the worse that person should be at learning arbitrary associations in the paired associate learning task.

Finally, while the main effect of InSecondLanguage was not significant (z = –0.483, p = 0.629), we observed a significant interaction (as modelled through a te() smooth) between Age and Co-occurrence Frequency (χ^2 = 38.687, $p < 0.001$). This interaction was significantly different between the First Language and the Second Language (χ^2 = 9.122, p = 0.028).[6]

The contrast surface for InSecondLanguage for the tensor product interaction between Age and Co-occurrence Frequency is presented in Figure 5. Figure 5 shows that performance on the paired associate learning task is highly similar in L1 and L2 for the young participants. A post-hoc analysis in which we included the data for the young participants only showed an Age by Co-occurrence Frequency interaction that was significant (χ^2 = 19.658, p = 0.001), but that did not differ between the First Language and the Second Language (χ^2 = 1.357, p = 0.961). Furthermore, this post-hoc analysis revealed no main effect of InSecondLanguage (z = 0.250, p = 0.803).

6 Note that we used (log-transformed) unaltered Google co-occurrence frequencies for both German and Chinese, in spite of the fact that the size of Google in Chinese and German is different. We decided not to further calibrate the scales of the Co-occurrence Frequencies in German and Chinese for two reasons. First, the (log-transformed) Co-occurrence Frequency distribution was relatively similar for German and Chinese without any further adjustments (German: mean = 9.54; sd = 2.77, Chinese: mean = 11.45; sd = 2.88). Second, given the fact that the size of Google in German and Chinese is hard to determine it would be unclear how exactly to calibrate the Co-occurrence Frequency ranges in German and Chinese (see Materials section).

To verify that potential scale differences between Co-occurrence Frequency in German and Chinese did not influence our interpretation of the results, however, we carried out a post-hoc analysis in which the Co-occurrence Frequency of all items in Chinese was divided by the ratio of the mean Co-occurrence Frequency in Chinese and the mean Co-occurrence Frequency in German prior to the log-transformation (correlation with the original logged Co-occurrence Frequency measure: r = 0.984). In this analysis the tensor product interaction between Age and Co-occurrence Frequency remained significant (χ^2 = 43.458, $p < 0.0001$) and this tensor product interaction remained significantly different between the First Language and the Second Language (χ^2 = 8.220, p = 0.042). In addition, the main effect of InSecondLanguage reached significance: (z = 2.354, p = 0.019), with a better overall performance in the second language than in the first language.

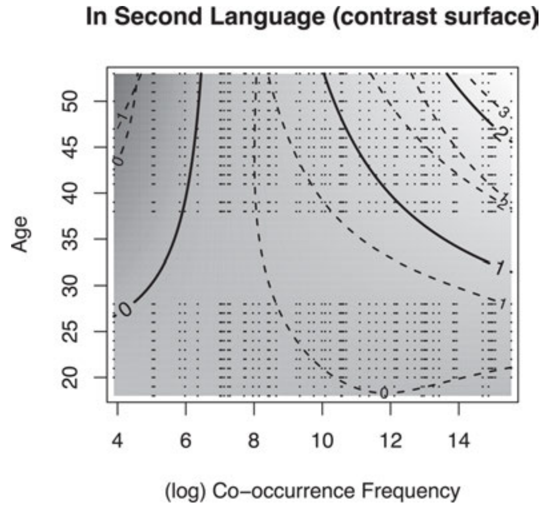

Figure 5: Effect of the (log-transformed) Co-occurrence Frequency and Age on the Paired Associate Learning performance: contrast between First Language and Second Language. The z-axis represents performance in the paired associate learning test on the logit scale. While young participants perform similarly in L1 and L2, old participants show an advantage of performing the PAL task in their L2 for most of the Co-occurrence Frequency range.

Older participants, however, show an advantage of performing the paired associate learning task in L2. As can be seen in Figure 5, older participants perform better in the second language for all but the pairs with the lowest co-occurrence frequencies, showing improved performance in L2 for co-occurrence frequencies of 6.5 or higher (i.e., for 90% of the word pairs). The better performance of older participants in the second language was confirmed by a post-hoc analysis on the data for the old participants only, which showed an Age by Co-occurrence Frequency interaction ($\chi^2 = 36.335$, $p < 0.001$) that was significantly different between the First Language and the Second Language ($\chi^2 = 14.959$, $p = 0.002$). In addition, this post-hoc analysis showed a main effect of InSecondLanguage ($z = 2.113$, $p = 0.035$), with a better performance in the second language than in the first language.

Figure 6 presents the by-item averages for the older monolinguals and the older bilinguals in German. Items are ranked from easy to hard. We re-defined item difficulty as the mean paired associate learning performance across the old and the young participants (as opposed to our a priori estimation of item difficulty on the basis of the co-occurrence frequency of the words in a pair that was used to construct the materials for this study). Consistent with the results from the GAMM, the performance of the Chinese-German bilinguals in their L2 is much better than that of the German monolinguals in their L1 for most of the item difficulty range.

In total, the older bilinguals outperformed the older monolinguals on 19 of the 30 items, while the older monolinguals performed better for 7 items only. For 13 items the German proportion of correct responses is at least 0.15 lower for the monolinguals than for the bilinguals (versus 5 items for which the proportion of correct responses is at least 0.15 lower for bilinguals than for monolinguals).

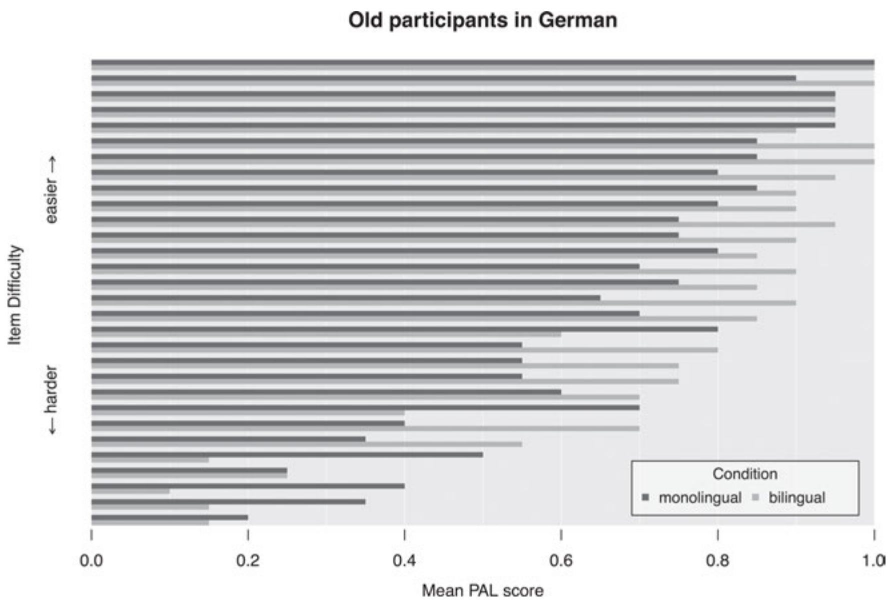

Figure 6: By-item performance in the Paired Associate Learning task for old monolinguals and bilinguals in German. For all but the hardest items, old Chinese-German bilinguals perform better in German than do age-matched German monolinguals.

The difference between the monolinguals and the bilinguals in German presented in Figure 6 was confirmed in an additional post-hoc analysis of older participants in German only, which once more showed an Age by Co-occurrence Frequency interaction ($\chi^2 = 24.828$, $p < 0.001$) that was significantly different between the First Language and the Second Language ($\chi^2 = 10.244$, $p = 0.017$).[7]

As can be seen at the bottom of Figure 6, the performance of older participants in their second language is particularly poor for the hardest pairs. This decrease in performance is substantially reduced for younger participants (see Figure 5). The experience of older bilingual participants in their first language is much greater

7 Note that the main effect of InSecondLanguage did not reach significance for this subset of the data ($z = 0.246$, $p = 0.806$).

than that of young bilinguals. Not only were the old bilinguals older than the young bilinguals, they also moved to Germany at a later age (mean age of learning the second language for old bilinguals: 27.45, young bilinguals: 19.90). Participants in the paired associate learning task are not restricted to thinking in the language the task is presented in. One potential explanation for the poor performance of older adults on the hardest items in their second language, therefore, is that older bilinguals may have recruited their native language when asked to learn arbitrary associations between relatively infrequent words in their second language.

In addition to the improved performance of older participants in the second language, we observed an attenuation of the Age effect in the second language. Figure 7 shows the predicted performance (Proportion Correct) in the First Language and in the Second Language as a function of Age and Co-occurrence Frequency. As can be seen in Figure 7, there is a clear age effect in the first language, which is qualitatively similar to the age effects reported in monolingual paired associate learning studies (see the GAMM for the English PAL data reported above). Throughout the left panel of Figure 7, the performance of the old participants is worse than that of the young participants. The difference is small for items with high association rates, but increases as Co-occurrence Frequency decreases. At the mid-point of the co-occurrence frequency range (as indicated by the dashed lines in Figure 7), for instance, the estimated performance of the oldest participants is 59% correct, whereas the performance of the youngest participants is 76% correct (difference: 17%). In the second language this age effect is substantially reduced. A clear age effect in the second language is present for the hardest pairs only, albeit for different reasons than in the first language (see above). At the mid-point of the co-occurrence frequency range the estimated performance is between 73% and 76% across the age range. For the easiest pairs old participants even perform somewhat better than young participants, although the performance of both groups is close to ceiling.

An inspection of the by-item averages for the German monolinguals in German, the Chinese-German bilinguals in Chinese and the Chinese-German bilinguals in German sheds further light on the reduced age effect for bilinguals in their second language. First, consider Figure 8, which shows the by-item performance of young and old German monolinguals in German. Young monolingual German participants performed better than old monolingual German participants across the item difficulty range. In total, the young monolingual German participants outperformed the old monolingual German participants on 23 of the 30 items, while the older participants performed better for 3 items only. For no less than 14 of the 30 items in the German paired associate learning test the mean item score (i.e., proportion correct) is at least 0.15 lower for the old participants than for the young participants.

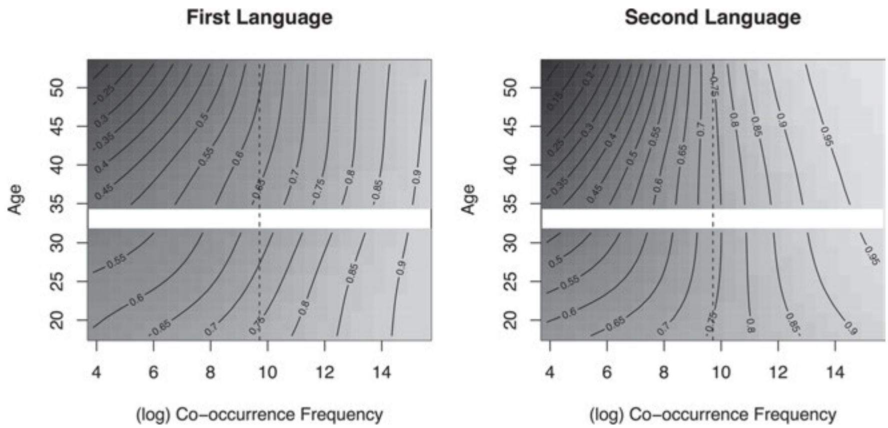

Figure 7: Effect of the (log-transformed) Co-occurrence Frequency and Age on the Paired Associate Learning performance in the first and second language. The z-axis represents proportion correct in the paired associate learning test. Dotted lines indicate the mid-point of the (log) Co-occurrence Frequency range. Older participants, but not younger participants show improved performance in their second language. The typical age effect for monolinguals is present in L1; this age effect is substantially reduced in L2.

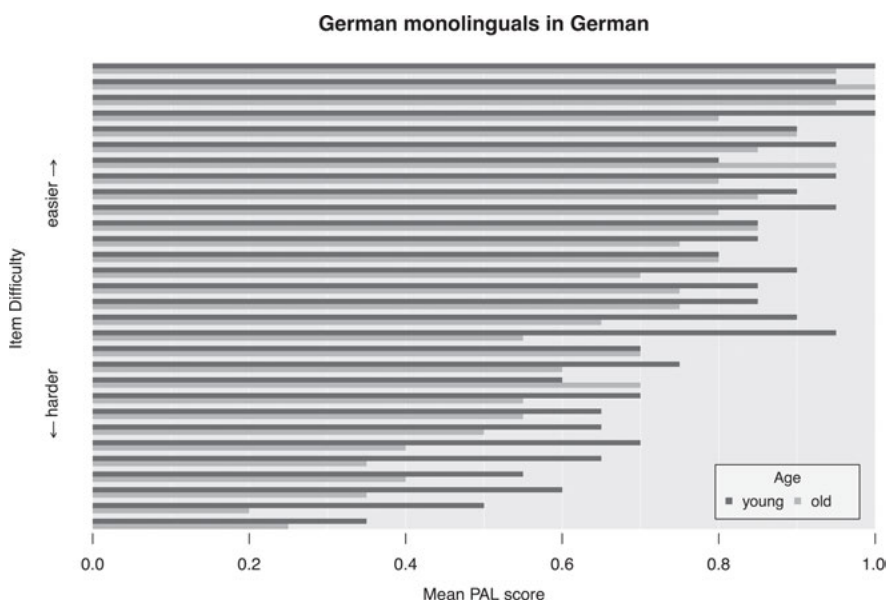

Figure 8: By-item performance in the Paired Associate Learning task for German monolinguals in German. Young participants perform better than old participants across the Item Difficulty range.

Figure 9 shows the by-item performance of the Chinese-German bilinguals in Chinese. For the easiest one third of the items the difference between old and young participants is negligible. For the hardest two thirds of the items, however, clear age differences are present. The young bilinguals outperformed the old bilinguals for 18 of the 30 items, while the old bilinguals did better for 4 items. For 9 out of the 30 items in the Chinese paired associate learning test the mean item score is at least 0.15 lower for the old participants than for the young participants. For 2 items the older Chinese-German bilinguals have an item score that is 0.15 higher than that of the young Chinese-German bilinguals for 2 items. Interestingly, these two items consist of pairs of words that may well be more associated for older participants than for younger participants: *喝水* - *吃药* ('drink water' - 'take medicine') and *房子* - *漏水* ('house' - 'water leakage').

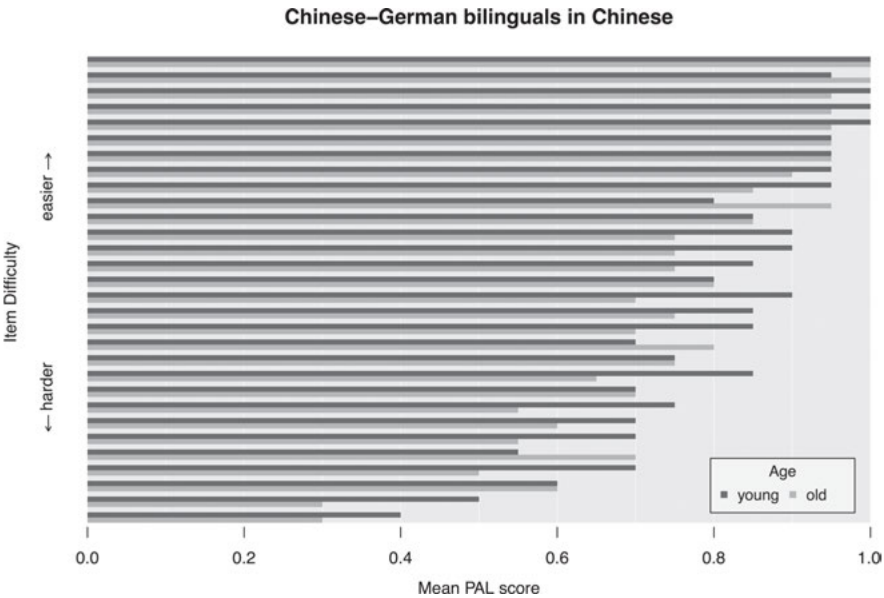

Figure 9: By-item performance in the Paired Associate Learning task for Chinese-German bilinguals in Chinese. As a result of less divergent experience with the language, the age effect is reduced as compared to German monolinguals in German. Nonetheless young participants still perform substantially better than old participants.

Similarly, the fact that the age effect is reduced for Chinese-German bilinguals as compared to German monolinguals follows from the reduced experience of Chinese-German bilinguals in their native language as compared to German monolinguals. German monolinguals typically lived in Germany for the entire duration of their life. The experience of young and old German monolinguals in German, therefore, is more-or-less a linear function of their age. By contrast, many of the old Chinese-German bilinguals have lived and worked in a German speaking environment for years and communicate in their first language less frequently. As a result, the difference in linguistic experience between young and old Chinese-German bilinguals in Chinese is much smaller than that between young and old German monolinguals in German – which results in a reduced age effect.

Despite the fact that the old Chinese-German bilinguals typically have somewhat more linguistic experience in German than the young Chinese-German bilinguals, the experience of young and old Chinese-German bilinguals in German should be more similar than that of young and old Chinese-German bilinguals in Chinese and that of young and old German monolinguals in German. We therefore expected to see a further reduction of the age effect for Chinese-German bilinguals in German. As can be seen in Figure 10 this prediction is borne out.

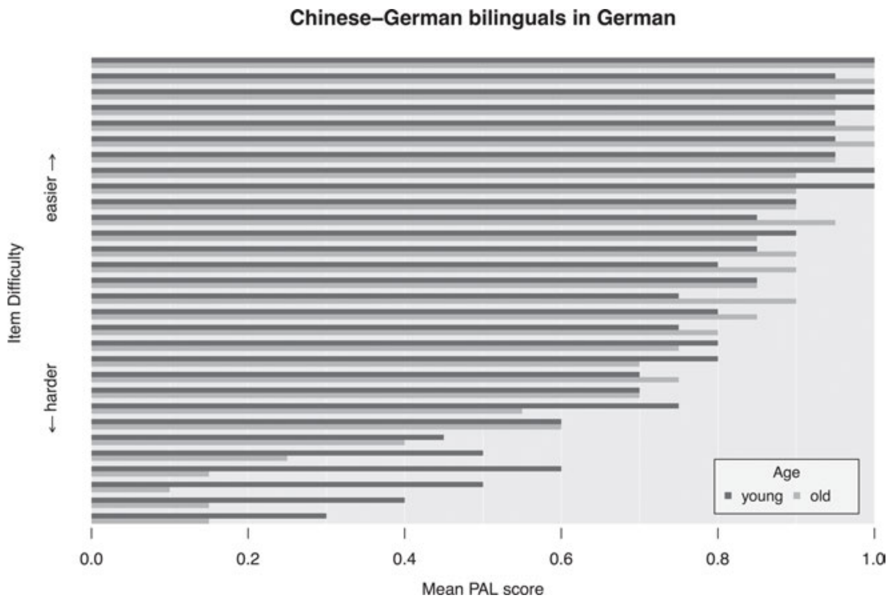

Figure 10: By-item performance in the Paired Associate Learning task for Chinese-German bilinguals in German. Young and old participants perform comparably for most of the items. Old participants are outperformed by young participants on a few of the hardest items only (see discussion in the text).

The difference between the old and young participants is negligible for most of the item difficulty range. In total, the young participants performed better than the old participants at 14 items, whereas the old participants outperformed the young participants on 10 items. A difference of at least 0.15 between the mean item scores for young and old participants is present for 6 of the hardest items only.[8]

For 1 item, older participants performed substantially better than younger participants: *waschen* - *Seife* ('to wash' - 'bar soap'). Although this effect may be due to variance, it could also be the case that the advantage for older participants may reflect an interesting aspect of discrimination learning: the association between a cue word and a response word is determined not only by the co-occurrence frequency of the cue word and the target word, but also by the co-occurrence frequency of the cue word with other words. The more words the cue occurs with and the higher the co-occurrence frequencies with other words, the weaker the association strength between the cue word and the target word.

Figure 11 illustrates this point for the item 'to wash' - 'bar soap', by showing the frequency for the word *bar soap*, as well as the frequency of two other words that are expected to frequently co-occur with *to wash*: *shower gel* and *body wash* in the Google n-gram viewer from 1988 to 2008. The frequency of *bar soap* remained fairly constant over the last 20 years. Simultaneously, however, the frequency of other words that are expected to co-occur with *to wash* substantially increased. By 2008, words like *shower gel* and *body wash* were more frequent than *bar soap*.

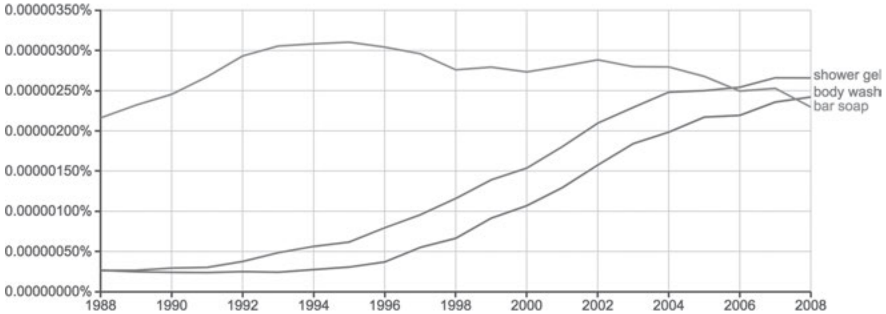

Figure 11: Frequency of *bar soap*, *shower gel* and *body wash* in the Google n-gram viewer from 1988 to 2008. The frequency of *bar soap* remains fairly constant, whereas the frequency of *shower gel* and *body wash* increases over time. As a result, the association strength between *to wash* and *bar soap* is weaker for recent L2 learns of German, which makes the PAL item 'to wash' - 'bar soap' harder for young Chinese-German bilinguals than for old Chinese-German bilinguals.

8 As mentioned above, this difference may be due to older participants recruiting their first language for the hardest pairs.

The old bilingual participants in our study have been in Germany for an average duration of 16.15 years, whereas the young bilinguals have been in Germany for an average duration of 4.65 years. When it comes to words that frequently co-occur with *to wash*, therefore, old participants have years of experience teaching them to almost exclusively expect *bar soap*, while young participants have a much more diverse experience that includes words like *shower gel* and *body wash*. From this perspective, the fact that old bilingual participants outperform young bilingual participants for the item 'to wash' - 'bar soap' is unsurprising.

While the advantage for the item 'to wash' - 'bar soap' for older bilingual participants as compared to younger bilingual participants is a non-significant effect for a single item only, it illustrates an important point with regard to discrimination learning in the context of paired associate learning. The expected performance in the paired associate learning test is determined not only by the association rates of the words in a pair, but by the distributional properties of the language as a whole, as well as by the linguistic experience of each individual participant.

The plots of the by-item performance in the paired associate learning test presented in Figures 8 through 10 demonstrate that the differences in performance on the paired associate learning test between old and young participants are most pronounced when the difference in linguistic experience is the greatest. By contrast, the differences between old and young participants are limited when the difference in experience is limited as well. More simply put: greater linguistic experience leads to worse performance.

5 Discussion

A well-established finding in the field of experimental psychology is that performance in a range of cognitive tasks decreases with age. The decreased performance of older participants in these tasks is typically interpreted as cognitive decline: a general and inevitable decline in cognitive function in the later stages of life (see e.g., Naveh-Benjamin and Old 2008; Deary et al. 2009; Salthouse 2009; Singh-Manoux et al. 2012). Recently, however, we have argued that age-related changes in the performance on cognitive tasks may be a result of increased experience with the stimuli used in these tasks (see Ramscar et al. 2013; Ramscar et al. 2014).

Paired associate learning is a classic example of a task in which older participants perform worse than younger participants (see e.g., desRosiers and Ivison 1986). The age effect in paired associate learning is small for "easy" word pairs (i.e., *north* - *south*) and large for "hard" word pairs (i.e., *jury* - *eagle*). This interaction between age and item difficulty does not readily fit with a view in which age-related changes in performance are due to a general decline in cognitive function.

Instead, from such a perspective, we would expect a main effect of age that is independent of item difficulty.

By contrast, from a discrimination learning perspective, the interaction between age and item difficulty is a natural consequence of experience. Under this view, the crucial distinction between hard and easy items in the paired associate learning task is the association between the cue word and the response word. The words *north* and *south* often occur together and therefore are highly associated. The words *jury* and *eagle* rarely co-occur and thus are weakly associated. Due to increased experience with the language, older participants are more sensitive to the association rates of words in the language. A lifetime of learning has taught them that *jury* and *eagle* do not occur together, and makes it hard to learn to respond *eagle* when presented with the word *jury*.

Despite the fact that the age by item difficulty interaction fits better with a discrimination learning approach than with cognitive decline, it is not straightforward to provide compelling evidence against the idea of cognitive decline. Experience with linguistic stimuli and age are highly correlated. Indeed, it is impossible to separate the effects of age and linguistic experience for paired associate learning when testing monolinguals.

Comparing the performance of monolinguals with the performance of bilinguals, however, allows us to tease apart the effects of age and linguistic experience. Experience in L1 is much greater for older participants than for younger participants. Experience in L2, however, varies as a function of the time spent learning L2, rather than as a function of age per se. The "cognitive decline" perspective on age-related changes in cognitive behavior predicts that older participants should show similar performance in L1 and L2, or better performance in L1 than in L2. Discrimination learning, however, predicts that older participants should perform better in L2 than in L1.

Here, we reported the results of a paired associate learning task for monolingual speakers of German, as well as for Chinese-German bilinguals. The results are consistent with the predictions of the discrimination learning perspective on the decreased performance of older participants in paired associate learning. Overall, older participants performed better in L2 than in L1, whereas younger participants showed similar performance in L1 and L2. For L2 word pairs with medium-to-high association rates older participants performed equally well or slightly better than younger participants. It was only for the items with the lowest association rates that we found an adverse age effect in L2. Possibly, this age effect is due to older participants recruiting their first language for the hardest L2 items (see our discussion above).

Interestingly, we furthermore found an effect of education level. Older participants with a PhD degree performed significantly worse than older participants without a PhD degree. Under the assumption that participants with a PhD degree have increased experience with the language, this is another example of a counter-intuitive effect that readily fits with the predictions of discrimination learning: the greater the experience with a language, the harder it is to learn new associations.

The results reported here are in line with findings from a number of recent studies that have documented the influence of linguistic experience on the performance in cognitive tasks (see Ramscar et al. 2013; Ramscar et al. 2014). Collectively, these studies suggest that a careful re-evaluation of experimental findings results in a more balanced view on age-related decreases in performance, in which surprisingly little evidence for a decline in cognitive function at an older age remains once the behavioral consequences of learning are accounted for.

Appendix 1

Table 2: Items for the German PAL test.

	Cue	Response	Cue (English)	Response (English)
Group 1	*Stadt*	*Köln*	'city'	'Cologne'
	Nord	*Süd*	'north'	'south'
	Reise	*Urlaub*	'travel'	'holiday'
	Liebe	*Herz*	'love'	'heart'
	Vater	*Sohn*	'father'	'son'
	Katze	*Hund*	'cat'	'dog'
	Spielzeug	*Kinder*	'toys'	'kids'
	Metall	*Eisen*	'metal'	'iron'
	Nacht	*Sterne*	'night'	'stars'
	schlafen	*träumen*	'to sleep'	'to dream'
Group 2	*Schnee*	*Ski*	'snow'	'ski'
	Sonne	*Wärme*	'sun'	'warmth'
	Bett	*Kissen*	'bed'	'pillow'
	Garten	*Blume*	'garden'	'flowers'
	waschen	*Seife*	'to wash'	'soap'
	Strom	*Lampe*	'power'	'light'
	Auto	*Ziel*	'car'	'destination'
	Schrank	*Hose*	'cabinet'	'pants'
	schwimmen	*Insel*	'to swim'	'island'
	Fuß	*Stein*	'foot'	'stone'
Group 3	*Kerze*	*Teller*	'candle'	'plate'
	denken	*malen*	'to think'	'to paint'
	Mauer	*Student*	'wall'	'student'
	Brief	*Herbst*	'letter'	'autumn'
	Salz	*Pflanze*	'salt'	"plant'

Table 2: (Continued)

	Cue	Response	Cue (English)	Response (English)
	Mensch	*Flasche*	'person'	'bottle'
	Tanz	*Feld*	'dance'	'field'
	Schiff	*Puppe*	'ship'	'puppet'
	Banane	*See*	'banana'	'lake'
	Schlüssel	*Zigarette*	'key'	'cigarette'

Table 3: Items for the Chinese PAL test. For many words parts-of-speech tagging out of context is ambivalent between noun and verb. For these words the translation to English is listed as '(to) [word]'.

	Cue	Response	Cue (English)	Response (English)
Group 1	*运动*	*健康*	'exercise'	'health'
	日本	*国家*	'Japan'	'country'
	医生	*疾病*	'doctor'	'disease'
	吃饭	*旅行*	'to eat'	'(to) travel'
	股票	*买卖*	'stock'	'(to) trade'
	妈妈	*家庭*	'mother'	'family'
	朋友	*帮助*	'friend'	'(to) help'
	杀人	*犯罪*	'to murder'	'to commit a crime'
	学校	*读书*	'school'	'(to) study'
	美丽	*眼睛*	'beautiful'	'eyes'
Group 2	*赚钱*	*银行*	'to make money'	'bank'
	房子	*漏水*	'house'	'to leak / leakage'
	天空	*飞机*	'sky'	'airplane'
	喝水	*吃药*	'to drink water'	'to take medicine'
	头发	*乌黑*	'hair'	'black'
	城市	*拥挤*	'city'	'crowded'
	小狗	*咬人*	'dog'	'to bite'
	金钱	*工作*	'money'	'job'
	小鸟	*鸣叫*	'bird'	'(to) chirp'
	马路	*冬天*	'road'	'winter'
Group 3	*车祸*	*黑暗*	'car accident'	'darkness'
	桌子	*树木*	'table'	'tree'
	公园	*气球*	'park'	'balloon'
	眼镜	*走路*	'glasses'	'(to) walk'
	游泳	*天气*	'(to) swim'	'weather'
	糖果	*胡椒*	'candy'	'pepper'
	铅笔	*书包*	'pencil'	'backpack'
	领带	*饼干*	'necktie'	'cookie'
	法官	*老鹰*	'judge'	'eagle'
	洋葱	*手指*	'onion'	'finger'

Appendix 2

```
# Load library (version 1.8-3)
library(mgcv)

# Load data
load("data.rda")
# Model
model = bam(Correct ~ s(Item, bs="re") +
                        s(PositionExperiment, Participant,
                        bs="fs",m=1) +
                        s(PositionList, Participant,
                        bs="fs", m=1) +
                        PhD + InSecondLanguage +
                        te(CooccurrenceFrequency, Age) +
                        te(CooccurrenceFrequency, Age,
                        by=InSecondLanguage),
              data=data, family="binomial")
```

References

Arnon, Inbal & Michael Ramscar. 2012. Granularity and the acquisition of grammatical gender: How order of acquisition affects what gets learned. *Cognition*, 122 (3), 292–305.

Arppe, Antti, Peter Hendrix, Petar Milin, Rolf Harald Baayen & Cyrus Shaoul. 2014. ndl: Naive discriminative learning [Computer software manual]. Retrieved from http://CRAN.R-project.org/package=ndl (R package version 0.2.16)

Baayen, Rolf Harald, Shravan Vasishth, Douglas Bates & Reinhold Kliegl. 2015. Out of the Cage of Shadows. arXiv: 0707.3168.

CKIP. 2014. Academia Sinica balanced corpus. Published at http://asbc.iis.sinica.edu.tw. Taipei: Chinese Knowledge Information Processing Group, Academia Sinica.

d'Acremont, Mathieu, Zhong-Li Lu., Xiangrui Li, Martial van der Linden & Antoine Bechara. 2009. Neural correlates of risk prediction error during reinforcement learning in humans. *NeuroImage*, 47 (4), 1929–1939.

Danks, David. 2003. Equilibria of the Rescorla-Wagner model. *Journal of Mathematical Psychology*, 47 (2), 109–121.

Daw, Nathaniel D., Aaron Courville & Peter Dayan. 2008. Semi-rational models of conditioning: The case of trial order. In Nick Chater & Mike Oaksford (eds.), *The probabilistic mind*. Oxford: Oxford University Press.

Dayan, Peter & Kent C. Berridge. 2014. Model-based and model-free pavlovian reward learning: Revaluation, revision, and revelation. *Cognitive, Affective, & Behavioral Neuroscience*, 1–20.

Deary, Ian J., Janie Corley, Alan J. Gow, Sarah E. Harris, Lorna M. Houlihan, Riccardo E. Marioni, Lars Penke, Snorri B. Rafnsson & John M. Starr. 2009. Age-associated cognitive decline. *British Medical Bulletin*, 92, 135–152.

desRosiers, Gabriel & David Ivison. 1986. Paired-associate learning: normative data for differences between high and low associate word pairs. *Journal of Clinical Experimental Neuropsychology*, 8, 637–642.

Epstein, Robert. 2012. Brutal truths about the aging brain. *Discover*, 33, 48–50.

Kamin, Leon J. 1969. Predictability, surprise, attention, and conditioning. In Byron A. Campbell & Russell M. Church (eds.), *Punishment and aversive behavior*, 279–296. New York: Appleton-Century-Crofts.

Lindenberger, Ulman. 2014. Human cognitive aging: Corriger la fortune? *Science*, 346, 572–578.

McDannald, Michael A., Joshua L. Jones, Yuji Takahashi & Geoffrey Schoenbaum. 2014. Learning theory: A driving force in understanding orbitofrontal function. *Neurobiology of Learning and Memory*, 108, 22–27.

Miller, Ralph R., Robert C. Barnet, & Nicholas J. Grahame. 1995. Assessment of the Rescorla-Wagner model. *Psychological Bulletin*, 117 (3), 363–386.

Naveh-Benjamin, Moshe, & Old, Susan R. 2008. Aging and memory. In John H. Byrne, H. Eichenbaum, Randolf Menzel, Henry L. Roediger, & David Sweatt (eds.), *Learning and memory: A comprehensive reference*, 787–808. Oxford: Elsevier.

Pearce, John M., & Geoffrey Hall. 1980. A model for pavlovian learning: variations in the effectiveness of conditioned but not of unconditioned stimuli. *Psychological Review*, 87, 532–552.

Ramscar, Michael, Melody Dye & Joseph Klein. 2013. Children value informativity over logic in word learning. *Psychological Science*, 24 (6), 1017–1023.

Ramscar, Michael, Melody Dye & Stewart McCauley. 2013. Error and expectation in language learning: The curious absence of ‘mouses’ in adult speech. *Language*, 89 (4), 760–793.

Ramscar, Michael, Peter Hendrix., Bradley Love, & Harald Baayen. 2013. Learning is not decline: The mental lexicon as a window into cognition across the lifespan. *Mental Lexicon*, 8 (3), 450–481.

Ramscar, Michael, Peter Hendrix, Cyrus Shaoul, Petar Milin & Rolf Harald Baayen. 2014. The myth of cognitive decline: Non-linear dynamics of lifelong learning. *Topics in Cognitive Science*, 6, 5–42.

Ramscar, Michael, Daniel Yarlett, Melody Dye, Katie Denny & Kirsten Thorpe. 2010. The effects of feature-label-order and their implications for symbolic learning. *Cognitive Science*, 34 (6), 909–957.

Rescorla, Robert. 1968. Probability of shock in the presence and absence of CS in fear conditioning. *Journal of Comparative and Phsyiological Psychology*, 66 (1), 1–5.

Rescorla, Robert. 1988. Pavlovian conditioning: It’s not what you think it is. *American Psychologist*, 43, 151–160.

Salthouse, Timothy. 2009. When does age-related cognitive decline begin? *Neurobiology of Aging*, 30, 507–514.

Schultz, Wolfram. 2006. Behavioral theories and the neurophysiology of reward. *Annual Review of Psychology*, 57, 87–115.

Schultz, Wolfram, Peter Dayan, & P. Read Montague. 1997. A neural substrate of prediction and reward. *Science*, 275, 1593–1599.

Siegel, Shepard, & Lorraine G. Allan. 1996. The widespread influence of the rescorla-wagner model. *Psychonomic Bulletin & Review*, 3 (3), 314–321.

Singh-Manoux, Archana, Mika Kivimaki, M. Maria Glymour, Alexis Elbaz, Claudine Berr, Klaus P. Ebmeier, Jane E. Ferrie & Aline Dugravot. 2012. Timing of onset of cognitive decline: results from whitehall ii prospective cohort study. *British Medical Journal*, 344.
Sutton, Richard S., & Andrew G. Barto. 1981. Toward a modern theory of adaptive networks: Expectation and prediction. *Psychological Review*, 88, 135–170.
Tremblay, Léon, Jeffrey R. Hollerman & Wolfram Schultz. 1998. Modifications of reward expectation-related neuronal activity during learning in primate striatum. *Journal of Neurophysiology*, 80 (2), 964–977.
Wagner, Allan & Robert Rescorla. 1972. A theory of Pavlovian conditioning: Variations in the effectiveness of reinforcement and nonreinforcement. In Abraham H. Black & William F. Prokasy (eds.), *Classical conditioning II*, 64–99. New York: Appleton-Century-Crofts.
Wood, Simon N. 2006. *Generalized additive models*. New York: Chapman & Hall/CRC.
Wurm, Lee H., & Sebastiano A. Fisicaro. 2014. What residualizing predictors in regression analysis does (and what it does not do). *Journal of Memory and Language*, 72, 37–48.
Yin, Henry H., & Barbara J. Knowlton. 2006. The role of the basal ganglia in habit formation. *Nature Reviews Neuroscience*, 7 (6), 464–476.