Running head: LEARNING IS NOT DECLINE

Learning is not decline: The mental lexicon as a window into cognition across the lifespan

Michael Ramscar[1], Peter Hendrix[1], Bradley Love[2] and Harald Baayen[1]

[1]Eberhard Karls Universität Tübingen

[2]University College London

The Mental Lexicon, *in press*

Contact information:

Michael Ramscar
Eberhard Karls Universität Tübingen
Wilhelmstrasse 19
72074 Tübingen

Email: *michael.ramscar@uni-tuebingen.de*

Abstract

As otherwise healthy adults age, their performance on cognitive tests tends to decline. This change is traditionally taken as evidence that cognitive processing is subject to significant declines in healthy aging. We examine this claim, showing current theories over-estimate the evidence in support of it, and demonstrating that when properly evaluated, the empirical record often indicates that the opposite is true.

To explain the disparity between the evidence and current theories, we show how the models of learning assumed in aging research are incapable of capturing even the most basic of empirical facts of "associative" learning, and lend themselves to spurious discoveries of "cognitive decline." Once a more accurate model of learning is introduced, we demonstrate that far from declining, the accuracy of older adults lexical processing appears to improve continuously across the lifespan. We further identify other measures on which performance does not decline with age, and show how these different patterns of performance fit within an overall framework of learning.

Finally, we consider the implications of our demonstrations of continuous and consistent learning performance throughout adulthood for our understanding of the changes in underlying brain morphology that occur during the course of cognitive development across the lifespan.

Learning is not decline: The mental lexicon as a window into cognition across the lifespan

As otherwise healthy adults age, their performance on cognitive tests tends to change. These performance changes have traditionally been taken as showing that the functionality of underlying cognitive processes is subject to significant declines even in healthy aging (Deary et al, 2009; Salthouse, 2009, 2011; Singh-Manoux et al., 2012). However, in a recent article Ramscar et al (2014) pointed out that it is impossible to determine whether cognitive processes actually decline across the lifespan in the absence of models of processing, and without an understanding of the way that learning changes the processing demands imposed on the cognitive system. In a series of case studies, Ramscar et al showed that, over a range of cognitive tasks, once a proper measure of processing load is taken into account, the pattern of performance change typically seen across the lifespan can be accounted for without having to invoke "declines" in otherwise undefined "processes." Once learning processes were formally defined, the performance of older and younger adults on cognitive tests could be more straightforwardly modeled in terms of a set of relatively consistent capacities faced with processing loads of increasing size and complexity.

The central thesis put forward by Ramscar et al is that the evidence for cognitive decline in healthy minds is weak and that the methods used to argue that our cognitive abilities decline critically fail to account for the growing information processing loads that experience brings. Since this article was published, many researchers have questioned these claims (see e.g., Rabbitt, 2014; Carey, 2014; Brink, 2014). The message coming from these specialists on cognitive aging is that there is good evidence that the minds and brains of healthy adults do decline, and that these declines occur in ways that belie Ramscar et al's information processing concerns.

In what follows, we examine these objections, and show that researchers massively over-estimate the extent to which the available empirical evidence actually supports claims to the effect that cognitive processes decline in the course of healthy aging. Indeed, we present evidence that, when properly evaluated, in many case the empirical record indicates that the opposite is true:

First, we show how the models of learning that are tacitly accepted across the aging literature are not only simplistic, but that they ignore all of the progress that has been made in understanding learning over the past half century. We show that the models of learning assumed in aging research are incapable of capturing the most basic of established empirical facts relating to simple "associative" learning processes.

Second, we show how this faulty model lends itself naturally to spurious discoveries of "cognitive decline." We demonstrate how, given the naïve assumptions about learning made by psychometricians, analyses of a large, normative Paired Associate Learning (PAL) data set appears to reveal that PAL performance in otherwise healthy adults is subject to significant declines between 39 and 49 years of age. We show how this conclusion is unwarranted, revealing that once a more accurate model of learning is combined with a more faithful representation of lexical information, it would seem that far from declining, the accuracy of older adults' representation of the lexicon improves continuously across the lifespan.

Third, we consider the confirmation bias that prevails throughout the aging literature: we suggest that researchers find "declines" because they expect to find them, and because their research programs are designed to confirm the "cognitive decline" hypothesis. We identify several measures on which performance does not decline with age, and show how more accurate modeling can make sense of these different patterns of performance within an overall framework of learning.

Finally, we consider the implications of our findings, and of our demonstrations of continuous and consistent learning performance across the lifespan for our understanding of the changes in underlying brain morphology that occur during the course of cognitive development across the lifespan.

The nature of lexical learning across the lifespan

A central part of the argument put forward in Ramscar et al (2014) is that lexical learning continues throughout the lifespan. This raises a question, where is the evidence of this continued learning? As Rabbitt (2014) puts it:

> "Ramscar et al insist that vocabulary tests cannot be appropriate measures because they are biased towards [sic] low frequency words and so do not accurately assess older people who know more rare words that are not tested. It is questionable whether most older people actually do know more rare words than most young adults, but scores on vocabulary tests are not the only, or the best comparison. … Perhaps Ramscar et al elide this point because of their need to counter a quite different objection that old people generally have only equal or even lower scores on vocabulary tests than the young."

Ramscar et al (2004) show how some straightforward facts about sampling and the statistical nature of lexical distributions (Baayen, 2001) guarantee that vocabulary tests will become increasingly less accurate as people get older. If we disregard vocabulary tests as a useful tool for assessing cognitive decline, we are left with Rabbitt's suggestion that older people may not actually know more rare words than young people. Does this actually make sense?

Consider life as a continuous process of sampling the world. In infancy, the part of the world sampled is highly restricted to the cot, the high-chair, and the family (Pereira, Smith, & Yu, 2014). During the school years, pupils are trained to absorb selected samples of the world at a rate far beyond individual experience would allow. In their twenties and thirties, speakers marry, and may have children of their own. They move to other places, travel more widely, and experience an ever-increasing array of technological innovations. In their sixties, speakers may become grandparents, start a new hobby and become expert bridge players, or captains of industry. It seems likely that as their experiences of the world accumulate, speakers will need a more diverse and more specialized vocabulary to communicate their experiences to other speakers. In other words, given how experience is sampled over the lifetime, it is extremely unlikely that the limited vocabulary acquired by the end of puberty would remain unchanged and sufficient for the remainder of life.

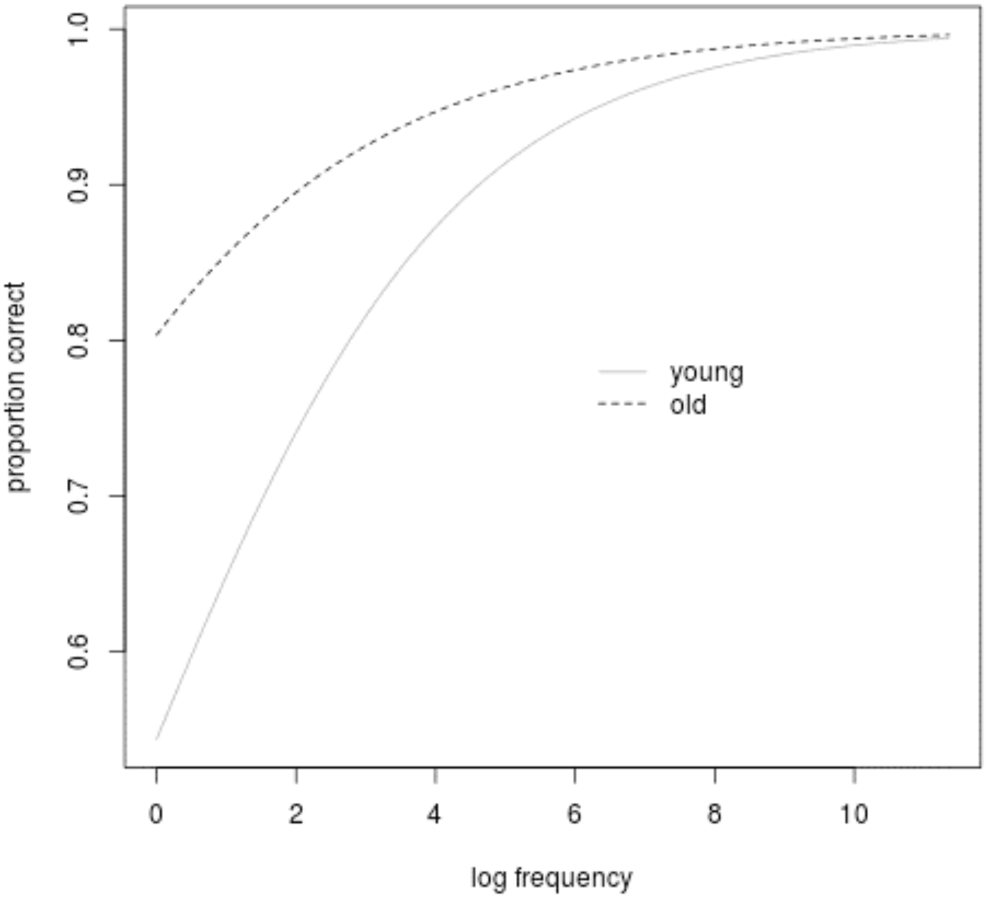

Figure 1: Proportion correct responses in visual lexical decision for the young and old subjects in Balota et al. (1999), plotted as a function of log word frequency in a logistic linear mixed model. The old subjects notably outperform the young subjects on low frequency words.

We can supplement this intuitive line of reasoning by an empirical fact. Figure 1 presents the accuracy of young (mean age 21.1 years) and old (mean age 73.6 years) adults in a lexical decision task with 2284 words (Balota et al. 1999) as a function of these words' (log-transformed) frequencies of occurrence in the CELEX lexical database (Baayen et al., 1995; see

Table A1 in the appendix for details of the corresponding statistical model). For the highest-frequency words, both groups perform with comparable accuracy. As frequency decreases, accuracy plummets for the young subjects to approximately chance performance. And while the old participants also make more errors on low frequency than high frequency words, they still outperform the young participants by a wide margin. For the lowest-frequency words, where the young subjects are at chance, the older subjects still get 80% of their responses correct. The markedly different accuracy levels of older adults as compared to the younger adults for the lowest-frequency words are completely incompatible with Rabbitt's suggestion that the former would not know more rare words than the latter.

Given that vocabulary tests clearly misrepresent what older speakers really know, this raises a question: to what extent are current interpretations of scores on other psychometric tests equally guilty of distorting the true extent of lexical knowledge in older adults? As we will now show, because learning in the lexicon involves more than simply adding new items to a list, current interpretations of scores for other psychometric tests are also guilty of underestimating older adults' cognitive abilities.

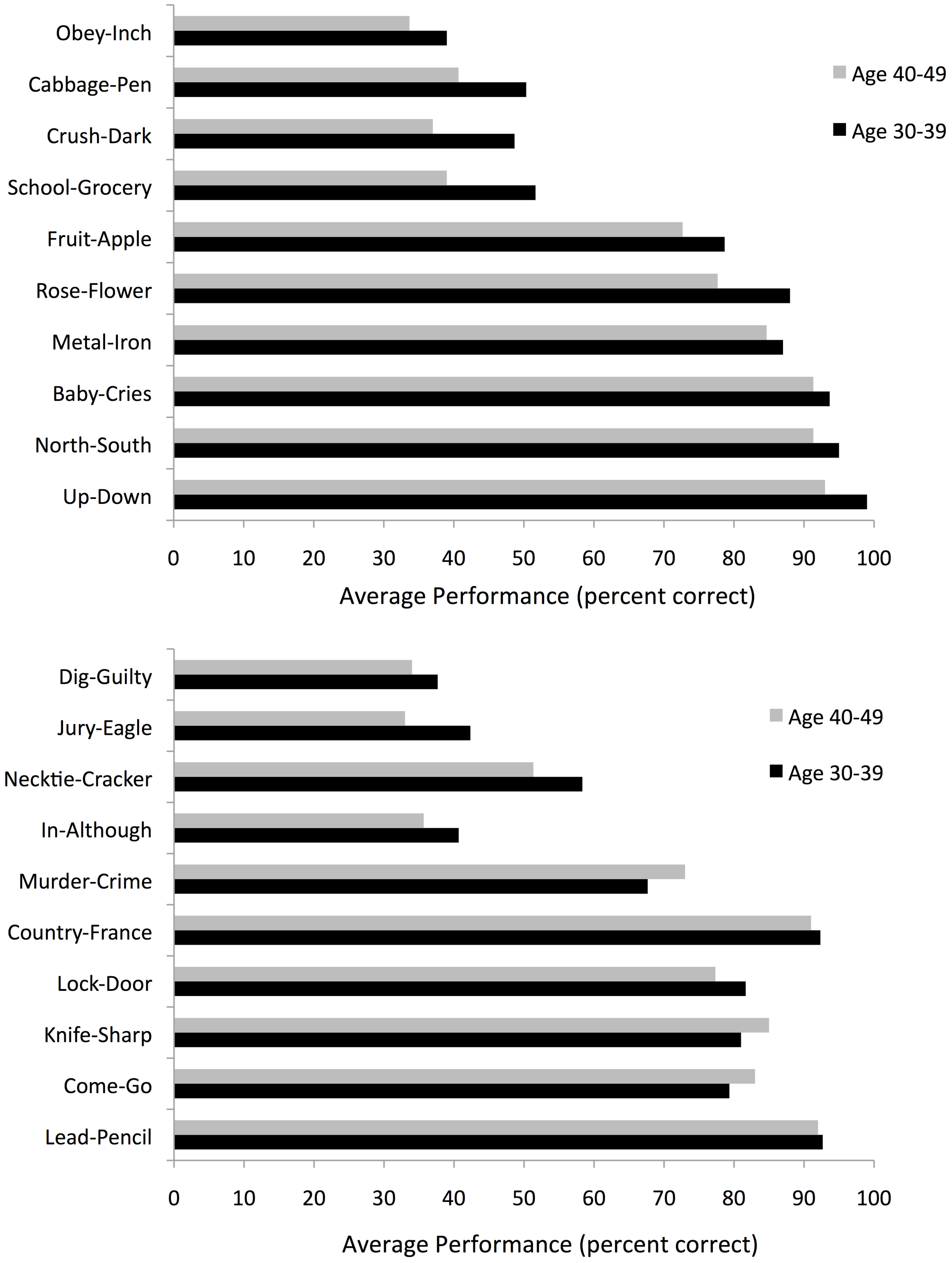

Figure 2: Average by-item performance for adults aged 30-39 and aged 40-49 (50% females in each group), tested on forms 1 (Top Panel, N=200) and 2 (Bottom Panel, N=200) of the WMS-PAL subtest (desRosiers & Ivison, 1988). Performance changes systematically: on average, performance differences are greater for harder items than the easier items.

To demonstrate this point, we take as an example “Paired Associate Learning” (PAL), a common cognitive performance measure in which people are required to memorize associations between two words (e.g., *dig-guilty*, or *lead-pencil*). The test is popular as a clinical measure, and often used as a means for evaluate learning and memory processes in experimental settings. Further, in comparison with other memory measures, researchers consider that, “performance on PAL may be of greater prognostic relevance for day to day functioning where the same associative abilities are required” (desRosiers & Ivison, 1988). In a typical test, participants hear a list of cue (w_1) and response (w_2) words (e.g., *dig-guilty*, *lead-pencil…*) and are then required to produce w_2 when given w_1 as a cue. Figure 2 plots the performance of 200 30-39 year olds and 200 40-49 year olds, who provided normative data for forms 1 and 2 of the PAL subtest of Wechsler’s Memory Scale (WMS; desRosiers & Ivison, 1988).

If we were simply to focus on the changes between the items in the performance of the thirty and forty year-olds shown in Figure 2, then these data would appear to provide evidence that PAL learning capacities decline significantly between ages 39 (M PAL Accuracy = 70%), and 49 (M Accuracy = 66%; $t(39)=4.793$, $p<0.0001$). This finding, though perhaps surprising, would seem to support other claims to the effect that age-related cognitive declines are clearly visible after adults reach their mid-forties (Singh-Manoux, et al., 2012).

However, in addition to declining performance, these data also clearly show that PAL performance changes between the two groups in a systematic fashion: “hard” PAL items appear to become proportionally harder to learn over time. This suggests that the initial interpretation we suggested, that these data provide evidence of declining learning abilities in early middle-age, may be premature: All other things being equal, we might expect that, given the association rate in each test is consistent (participants hear each w_1 - w_2 pair once, repeated across three list trials), then if we assume that PAL tests are a straightforward measure of participants' ability to learn associations, "declines" in this ability ought to be consistent across items.

However, empirically, it has been shown that *association rates* (the frequencies at which items are encountered together) are insufficient to explain the systematic patterns of behavior associated with associative learning. In particular, two additional frequency factors have been shown to exert a significant influence on learning in associative tasks: cue *background rates* (Rescorla, 1968; Ramscar et al, 2013a; in the case of PAL, the frequency with which a cue word appears absent a response word), and *blocking* (the predictability of a response in the learning context based on prior learning, Kamin, 1969; Arnon & Ramscar, 2013, which in the case of PAL, is the predictability of the response word given the cue).

Further, the skewed distribution of language means that the relative influence of the factors that either inhibit (*blocking* and *background rates*) or promote learning (*association rates*) are likely to change as learners sample more and more words over time. This is important, given that the *co-occurrence-rates* participants are exposed to in training in a PAL learning study remain constant: a pair is heard, and participants have to learn to associate it. This means that unless the effects that sampling and prior learning can be expected to have on PAL learning are controlled for, it is impossible to know whether changes in PAL performance shown in Figure 2 result from increased experience or cognitive declines.

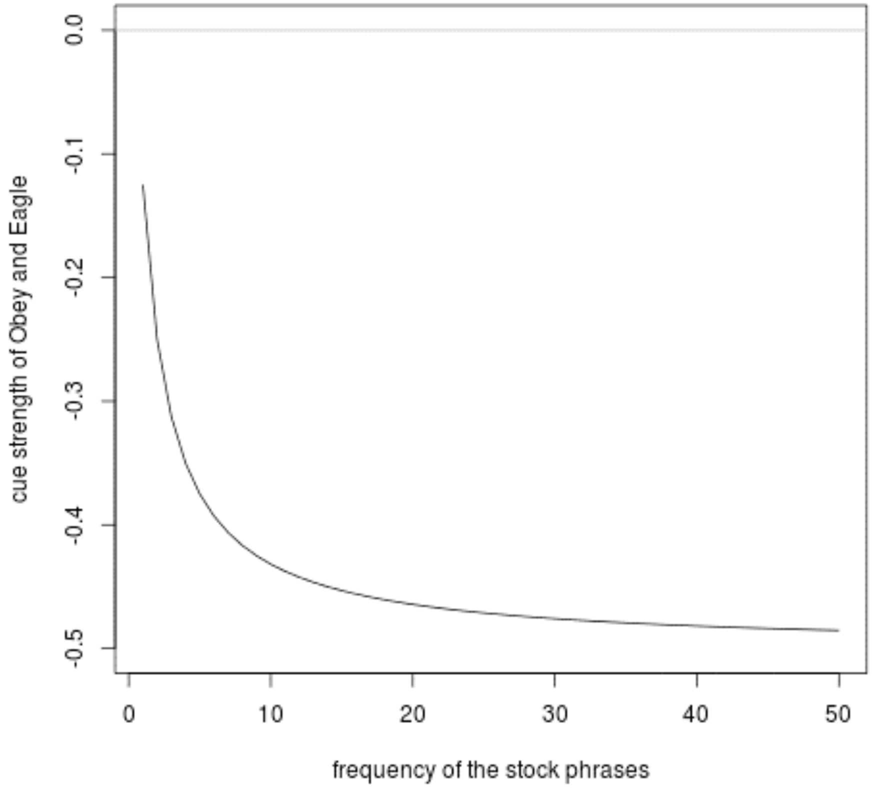

Figure 3: Cue strength for *Obey* to *Eagle* as a function of the frequency of two 'stock phrases' *American - Eagle* and *Obey - Rules*, in a small lexicon with the two stock phrases and two 'novel' pairings *Legal - Eagle* and *Obey - Eagle*. The frequency of the novel pairings is always 1. The plot shows how when learning is simulated using Danks (2003) equations for the Rescorla-Wagner (1972) model, the association weight between *Obey* and *Eagle* declines as the stock phrases' frequencies increase, even though both the structure of the lexicon and the association rate of *Obey - Eagle* remain constant (at 1).

To illustrate this point, Figure 3 shows what happens when all these factors come into play during associative learning in a very simple model of a lexicon. It depicts a small sample lexicon containing the stock phrases *American – Eagle* and *Obey – Rules*, and the “novel” pairings *Legal – Eagle* and *Obey – Eagle.* The plot shows the outcome of learning simulated using Danks (2003) equations for the Rescorla-Wagner model (a simple model that still embodies sufficient system complexity to account for the basic facts of associative learning; Rescorla & Wagner, 1972). As can be seen, as the frequencies of the stock phrases increase, the association weight between *Obey* and *Eagle* declines. This is despite the fact that the structure of the lexicon and the association rate of *Obey* - *Eagle* both remain constant (at a frequency of 1; see the appendix for full details, and the R code for this simulation).

Figure 3 illustrates the consequences of a basic fact about associative learning that has been known for around half a century: That the outcome of any single learning trial cannot be predicted by considering the association rate that a learner is exposed to on a single trial in isolation (Rescorla & Wagner, 1972; Ramscar et al, 2010). This also explains why it is wrong to think that by-item declines in performance in PAL reveal declines in associative learning capacities: This faulty inference depends on the assumption that PAL performance is determined by association rates alone. This erroneous assumption, which characterized classical behaviorism – and which, disturbingly, still lies at the heart of many contemporary researchers’ understanding of learning – has been rejected by all modern learning theories (see Rescorla, 1988, for discussion of both of these points).

To try to get a better estimate of the systematic role that the factors that have actually been shown to determine the outcome of associative learning are playing in this instance, we estimated values for three critical parameters (*background, blocking* and *association rates*; Kruschke, 2003; Nelson, Dyrdal, & Goodmon, 2005) that can be expected to influence the learning of PAL w_1-w_s word pairs (these parameters were: log transformed, w_1 word frequencies, differences in word frequencies expressed as frequency(w_2)/frequence(w_1), taken for the period 1980-1990 from the Corpus of Historical American English, COHA, Davies, 2012 and w_1-w_2 co-occurrence rates, taken from Google). When these were entered into a linear regression in order to predict the relative performance of 30-39 and 40-49 year olds tested in 1988 (plotted in Figure 2) for each word pair in the normative PAL data, they accounted for over 65% of the observed variance in performance ($r=.82$, $F(4)=19.385$, $p<0.01$). Also, as predicted, our estimated *background* and *blocking* rates were associated with lower scores, whereas *association* rates were associated with higher scores (all, $p<.01$).

Once the parameters we estimated from the corpus data are entered into the picture, it would appear that most, if not all of the difference in PAL performance seen between adults in their thirties and forties is due to learning. In particular, it appears that over time, learning hard PAL w_1-w_2 word pairs gets harder as language experience increases. As Figure 3 helps illustrate, this is because as learners master the informative details of the lexicon, the learning of a nonsensical link between two unconnected words must increasingly compete with prior learning to the effect that this link *is* nonsensical. It is worth noting here that one reason for this is that the learning of these kinds of *dissociations* is an important part of *discrimination learning*. Virtually all of the models of associative learning that have been developed in the past half-century actually implement discriminative learning principles, because it is these principles that actually appear to govern the processes that we still colloquially refer to as, “associative learning” (see Rescorla, 1988; Ramscar et al, 2010).

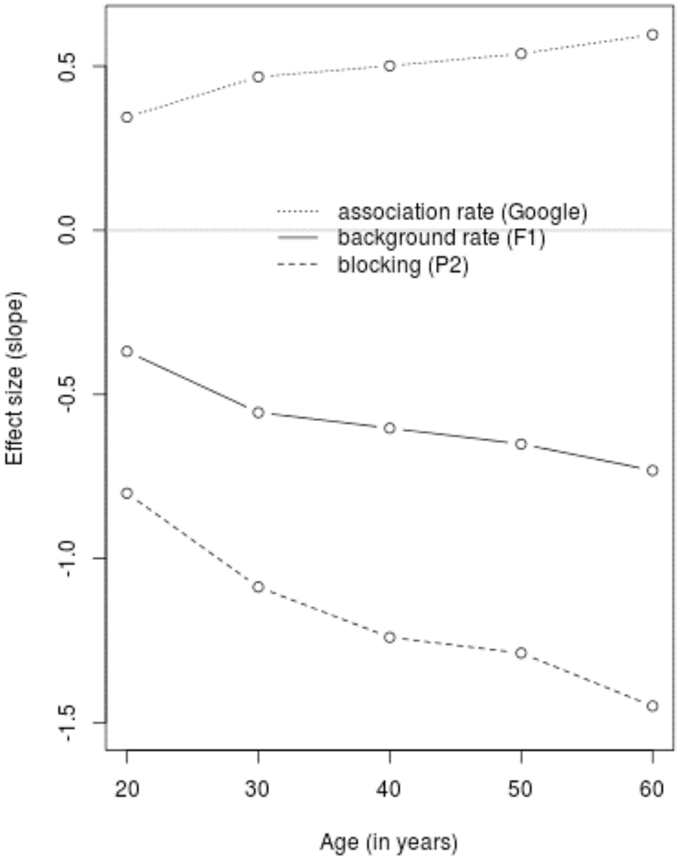

Figure 4: Mixed-effects slope estimates for the three parameters that estimate learnability constraints on the by-item PAL performance of the full set of 60-69, 50-59, 30-39 and 20-29 year-old adults in the normative data set (desRosiers & Ivison, 1988). Larger slope values indicate a greater degree of alignment with the structure of the language. All predictor effects and interactions in the model are significant, and all slopes (except the slope for blocking (P2) for the youngest age group) are significantly different from 0 (see Table 2 in the appendix). There is no significant main effect of age in the model. This analysis shows how lifelong PAL performance patterns reveal an ever-growing understanding of the systematic structure of the English lexicon as adult age increases, rather than any decline in learning capacity.

Our finding that many of the changes in PAL learning performance observed in early-middle-age are better attributed to learning than cognitive decline raises a further question: To what extent are the changes seen in PAL learning across the rest of the adult lifespan the product of the same, systematic learning factors?

To begin to address this question, we analyzed the full set of normative PAL data collected by desRosiers & Ivison (1988). The data was collected from 1000 adults, 50% of whom were men and 50% women. These were equally divided into the age groups 20-29, 30-39, 40-49, 50-59 and 60-69, and each participant completed either form 1 or form 2 of WMS PAL, such that each of data point represents the average score of 100 tests. The participants in desRosiers & Ivison's study were patients who had been hospitalized for non-neuropsychiatric conditions, and thus were tested in the same, relatively age-neutral context.

Along with our three corpus based learning parameters, these data were entered into a linear mixed effects model with word pair as random-effect factor. The details of this model are reported in the appendix (Table A2). As can be seen from the plot of mixed-effects slope estimates in Figure 4, the impact of the factors that basic learning theory predicts will inhibit the learning of an association – blocking and background rates – grow systematically across the lifespan, as does the influence of the factor that basic learning theory predicts will promote the learning of an association (the *association rate*). *In other words, our analysis of Paired Associate Learning indicates that the changing performance patterns that have been observed across the lifetime in PAL tasks are evidence of an ever-growing understanding of the systematic structure of the English lexicon that develops as age and experience grow.* Not only do these patterns not support the idea that older adults' learning capacities are in decline, they are also clear evidence that, in fact, learning capacities are both retained and fully engaged across adulthood.

Why decline appears to be in the eye of the beholder. And why a cognitive account of lifelong cognitive development matters

Ramscar et al (2014; see also Ramscar, 2014) point to another serious shortcoming in current approaches to the study of cognitive development in adulthood: whether older participants' performance on even the simplest of cognitive tests improves or declines appears to be a function of the context in which participants are tested. Consider, for example, the FAS task, in which people are asked to generate as many words beginning with F as they can in 60 seconds, followed by as many words beginning with A in 60 seconds, followed by as many words that begin with S. (A couple of rules govern the words that are allowed as responses in the test: Proper names like *Steve* or *France* are not allowed, nor are different versions of the "same" word, i.e., *friends*, *friendly.*)

In a meta-analysis of 134 studies, Ramscar et al found that while older participants outperformed younger adults at FAS recall in smaller studies, in very large surveys of the elderly population, older participants' performance declined as the total number of people tested in a study increased. Moreover, this effect was not due to regression to the mean (the analysis presented in Ramscar et al, 2014, controlled for this); Instead, it appears that in the data reported in the literature, there is a clear relationship between the FAS test scores of older adults and the number of older adults tested.

Test performance is both influenced by context, and can vary widely across cohorts (see e.g., Lynn, 1982; Flynn, 1987; Teasdale & Owen, 2005). However, although there is a large historical literature devoted to these very real problems (Schaie, 1959; 1973; 1975; 1977; 1988), and despite the fact that the researchers who employ the current generation of psychometric methods are careful to acknowledge that they cannot conclude anything about causality from the results of any given test, it appears that in practice whenever changes are observed on the correlated results of these tests, the temptation for researchers to interpret them causally – as evidence of decline – invariably proves too strong to resist. Although the cognitive aging literature contains a huge body of work that purports to chart the supposed declines in "cognitive abilities" that are assumed to undermine the minds of adults as they age, the fact is that this literature contains little more than a correlational record showing that scores change on tests that are incapable in themselves of supporting causal inferences about the reasons for change (Naveh-Benjamin & Old, 2008; Deary et al, 2009; Salthouse, 2009; Salthouse, 2011; Singh-Manoux et al., 2012).

Our analysis of PAL learning highlights the worrying shortcomings of these current methodologies. The introduction of even the most elementary discriminative learning model into an analysis of Paired Associate Learning <u>reverses</u> the standard, purely correlational interpretation of PAL test scores. Systematically lower scores do not reveal decline, but rather, they simply show a pattern of changes that any cognitively plausible model of lexical learning would predict. Indeed, the same patterns of change among lexical associations with experience revealed by our analysis can be observed in two-year olds (Ramscar et al, 2013a).

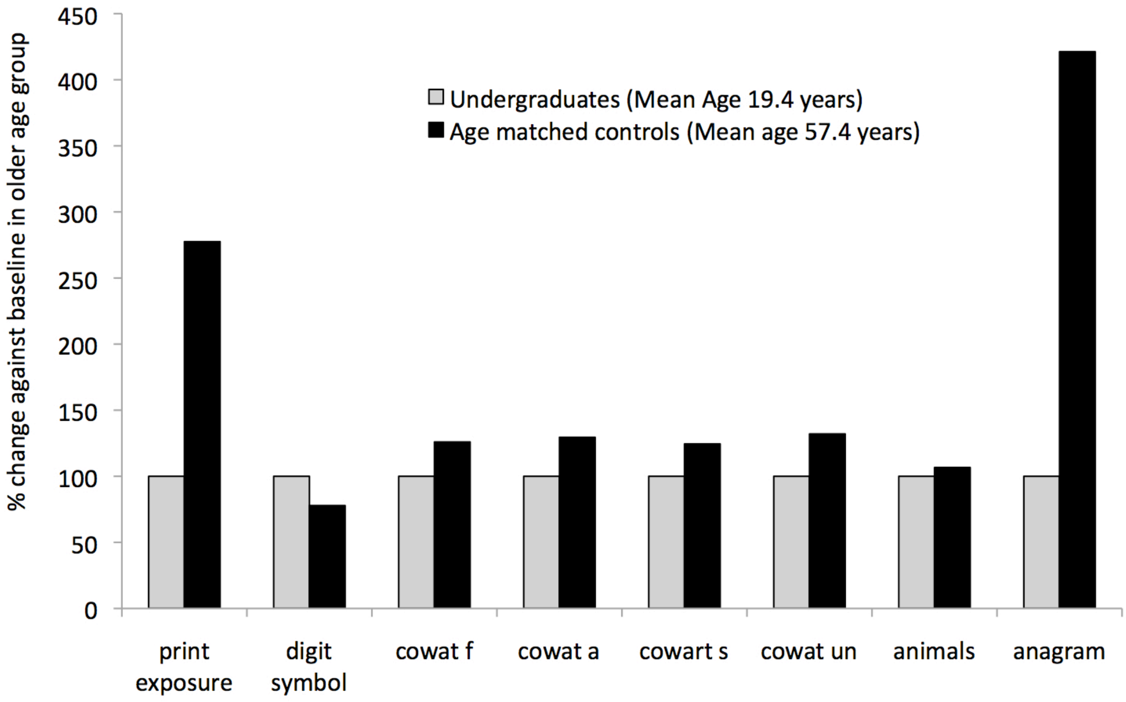

Figure 5: Performance of two control groups (group 1 N=23; M age 19.4 years; group 2 N=23; M age 57.4 years) in Hargreaves et al (2012). Young adult performance is plotted as a reference, and the performance of the older adults reflects the change in their performance against this baseline. The leftmost two bars represent estimates of the relative amount of print exposure each group has experienced, and the remaining bars plot performance in tests of Digit Symbol recall (recalling strings of alphanumeric characters), generating words beginning with F, A, S and UN, animal naming, and anagram solving.

These points matter a great deal. Figure 5 plots the performance of a group of 23 nineteen year olds and a group of 23 fifty-seven years olds on a range of measures of "cognitive performance" (from Hargreaves, Pexman, Zdrazilova, & Sargious, 2012). There are several things to note about this data:

First, it was an experimental study in which participants were carefully matched on a range of control variables, including the number of participants tested in each group.

Second, it was study of expertise rather than a study of “aging”. There are many good reasons to believe that this matters, most notably that making elderly participants aware of their age and of the stereotypes associated with aging, can inhibit their performance on tests (Steele, 1997; Hess et al, 2003; Hess & Hinson, 2006; Hess, Hinson, & Hodges, 2009).

Third, while, perhaps unsurprisingly, older adults’ print exposure is greater than that of the younger adults, it is notable that of the 7 cognitive measures tested, the older group out–perform the younger group on 6 of them; And even if we collapse all of the FAS related tasks into one, the fact is that the older adults outperform the young on 3 out of 4 of these measures. Further, while the younger adults performance on digit symbol is slightly better than that of the older adults, the younger adults performance on the anagram task is *vastly* inferior to that of the older adults.

Further evidence for the inferior sensitivity of younger adults to the distributional properties of the language can be gleaned from the lexical decision data discussed above, which indicated that adults perform the task more accurately (Figure 1). Figure 6 presents some central partial effects in a generalized additive mixed model fitted to the response latencies, with younger subjects in the left panels and older subjects in the right panels. The top panels present the effect of frequency. As expected, reaction times are longer for lower-frequency words. Older subjects slow down slightly more than younger subjects for the lower-frequency words.

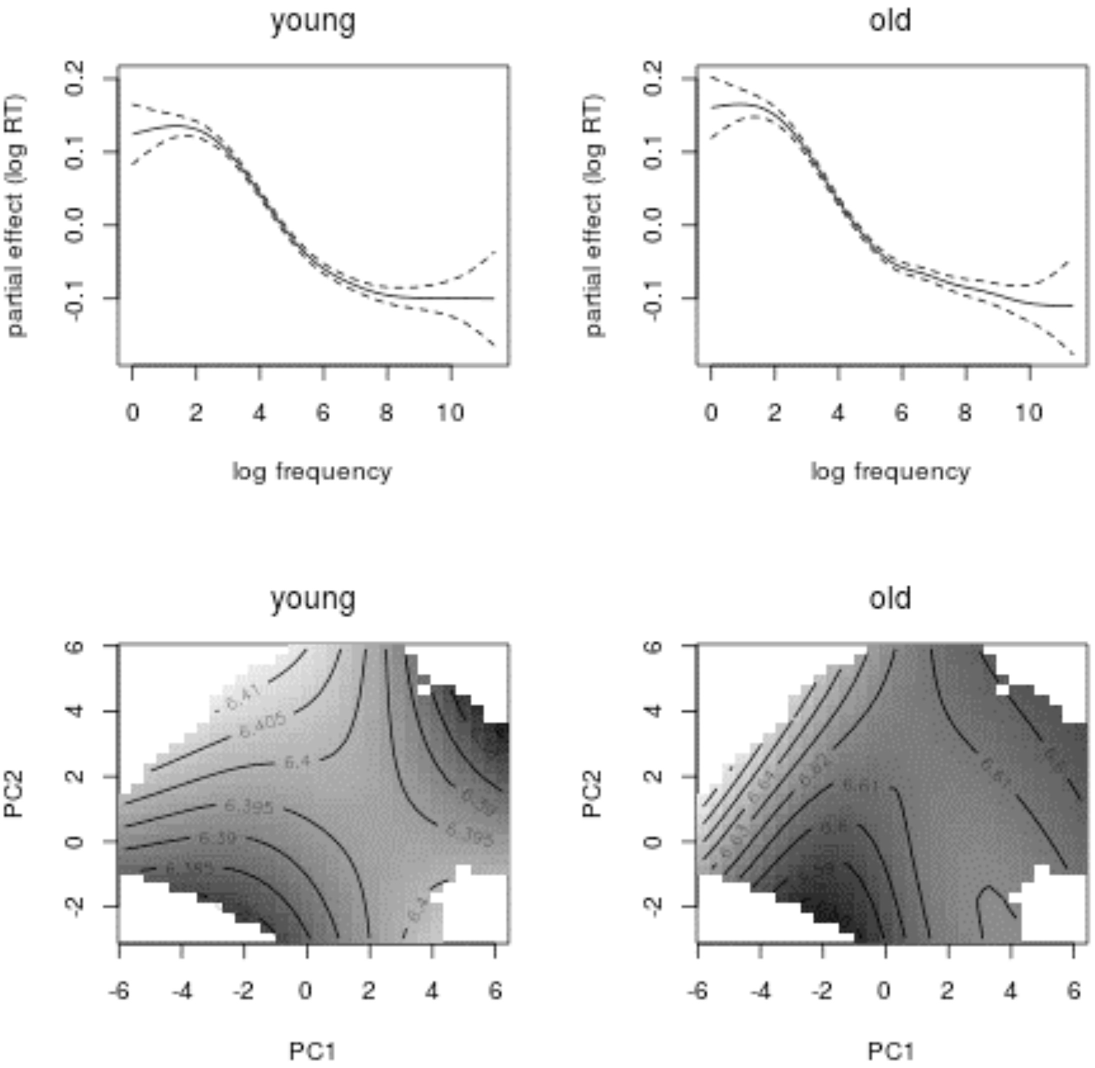

Figure 6. Upper panels: the partial effect of log frequency on log response latency. Younger adults (left) have shorter latencies for the lower-frequency words compared to older adults (right). Lower panels: the nonlinear interaction of two latent variables for orthographic consistency. The interaction does not reach significance for the younger adults, but is highly significant for the older adults, which reveal enhanced sensitivity to the distributional properties of sound-meaning mappings in English.

Before we attribute this to cognitive decline, we should first consider the lower panels, which present contour plots for the interaction of two latent variables, PC1 and PC2. PC1 contrasted forward enemies with phonological neighbors (and accounted for 43.5% of the variance in the 10 consistency measures), while PC2 contrasted friends with backward enemies (accounting for 22.2% of the variance; or further details, see Baayen, Feldman & Schreuder, 2006). (for details, see Baayen, Feldman, & Schreuder, 2006). The non-linear interaction of these two (orthogonal) latent variables was modeled with a tensor product smooth (Wood 2006, see the appendix Table A3 for the full model). Importantly, the interaction for the younger adults fails to reach significance, in contrast to the interaction for the older adults. Note that the contour lines for the younger adults are 0.005 log RT units apart, whereas those for the older adults are 0.01 log RT units apart. In other words, the younger adults are beginning to show, albeit very weakly, the pattern that characterizes the performance of the older adults. It is the older adults which evidence clear fine-grained sensitivity to the consonances and dissonances between spelling and sound in English.

A difference between older and younger subjects not shown in Figure 6 is that older subjects have longer RTs compared to younger subjects. Older participants' judgments are thus slower and more accurate, a phenomenon that might also be described as a "speed accuracy trade-off". This is consistent with applications of the drift-diffusion model to speed and accuracy data, which also indicates that the quality of information processing is not impaired in healthy aging, and suggests instead that slow downs in aging are largely attributable to other factors (Ratcliff, Thapar, & McKoon, 2011; 2010), such as non-decision time (e.g., motor movement) and boundary separation (i.e., the degree of cautiousness in responding).

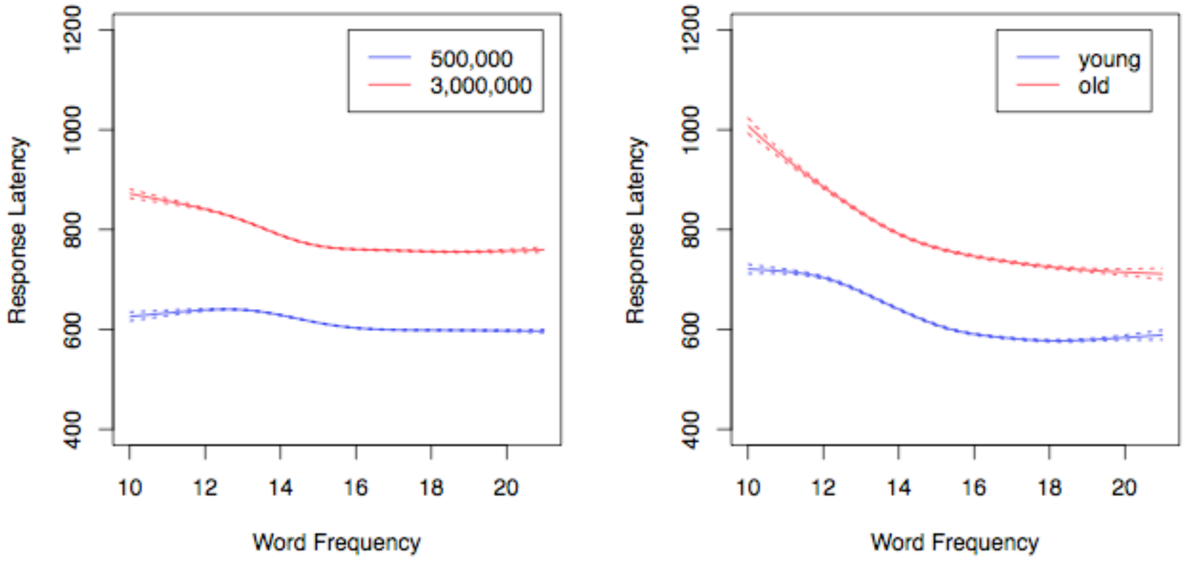

Figure 7. Left panel: fit of a generalized additive model to the simulated response latencies for the 2284 words tested by Balota et al. (1999) taken from the old and young models presented in Ramscar et al (2014). Right panel: fit of a generalized additive model to the empirical response latencies for the same word taken from young (mean age: 21.1) and old (73.6) adults (Balota et al, 1999). It is well established that lexical decision responses are slower for lower- frequency words (e.g., “whelp”) than higher-frequency words (“where”). This overall effect of frequency is present for both young and old adults and in the models. However, while frequency effects asymptote at higher frequencies in both models, they also level off again at the lowest frequencies in the younger model, a pattern also observed in the empirical data (see Figure 6 for more detail).

What is worth noting here is that while the characterization of a lexicality response as accurate or inaccurate in cross-generational lexical decision is made against the background of the lexicon of the total speech community, a young participant classifying a low-frequency word that she doesn't know as a non-word is still making what is to her an accurate and correct response. This has subtle consequences for this speed-accuracy trade-off, because this means that young adults are confronted with materials that contain – for them – a higher proportion of non-words than words, and this will make real words more noticeable and easier to respond to.

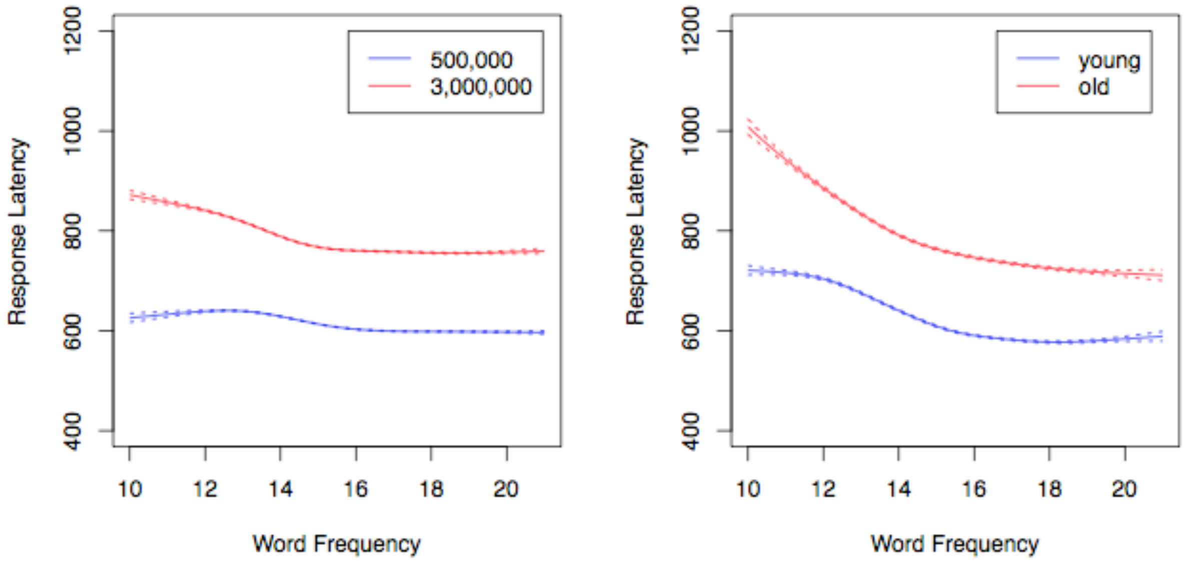

Figure 7. Left panel: fit of a generalized additive model to the simulated response latencies for the 2284 words tested by Balota et al. (1999) taken from the old and young models presented in Ramscar et al (2014). Right panel: fit of a generalized additive model to the empirical response latencies for the same word taken from young (mean age: 21.1) and old (73.6) adults (Balota et al, 1999). It is well established that lexical decision responses are slower for lower- frequency words (e.g., "whelp") than higher-frequency words ("where"). This overall effect of frequency is present for both young and old adults and in the models. However, while frequency effects asymptote at higher frequencies in both models, they also level off again at the lowest frequencies in the younger model, a pattern also observed in the empirical data (see Figure 6 for more detail).

While this suggests yet another reason to doubt that the slower responses of older adults present compelling evidence for degraded performance, importantly it also indicates that older and younger adults may be processing *different* information in completing this task. Ramscar et al. (2014) presented a large-scale simulation study using the Rescorla-Wagner equations, which correctly predicted the pattern of slower responses to lower frequency words observed in older adults. Figure 7 plots the simulated and empirical reaction times reported in Ramscar et al. (2014) in a way that allows the models' predictions – as well as the empirical effects just described – to be easily apprehended.

These models allow us to develop causal hypotheses about the nature of information in this task, and to explain *why* these particular patterns of response latencies emerge. (Note, simply saying, "because frequency," hardly counts as a causal hypothesis: saying that lower frequency words are read slower because they are lower in frequency – and we know more frequent words get read faster – is a re-description of the data, not an explanation.) First, note that the variance in model's predicted simulated RTs for younger and older adults in the lower frequency-range (Figure 7, left panel) is entirely a function of the weights the models learned from the training sets for the two age groups, which are set without free parameters (save for the selection of the size of the training samples themselves). These weights connect letter conjunctions (*n-gram* cues) in the words the model reads with the lexemes that humans and the models have to discriminate in reading (i.e. they represent the learned connection between the letters *d o* and *g* and the lexeme *dog*).

In the models, the *n-gram* cues are initially undifferentiated, and their learned values are set competitively, as the models seek to predict words from the letters they 'read.' The logic underlying this process can be intuitively grasped by considering the different cue values of letters in scrabble: if you have Q and A but no U, QA is an excellent cue for the legal scrabble words *qaid*, *qanat* and *qat*; however, D and I appear in many words, so DI only weakly supports individual words like *oxidize*, *dim* and *odium*.

Table 1. The 20 lowest frequency items in the set used to train the models and test older and young adults; BLASH has the lowest frequency of these items, and SKULK the highest. As can be seen, many of the letter bigrams in this set of words are fairly rare in English (see also, Nusbaum, 1985).

BLASH	SOUSE	CROME	VELDT
SCHNOOK	WHIG	GIBE	SLOE
LETCH	FILCH	LISLE	CONK
ZOUNDS	RHEUM	FLAYS	FRAPPE
JAPE	PARCH	SPLOTCH	SKULK

As Table 1 shows, when it comes to the test (and hence the training) set that produced the data plotted in Figure 6, the lower frequency words contain significantly more uncommon (low frequency) *n-grams* than the higher frequency words (Ramscar et al., 2014). This means that, while in a smaller vocabulary, these low frequency *n-grams* tend to be very good cues to a given lexeme, as a vocabulary grows, the probability that new words will also contain these *n-grams* increases (suppose someone who knows *qaid*, learns *qanat* playing scrabble, or hears that *qantas* is the name of an airline), then cue competition will increase (this is the same process that produces the changes in the weights of the w_1 cue *obey* to the w_2 response *eagle* in Figure 3).

Thus, in the model, the slower reaction times for lower frequency words for older speakers are the result of learning, and reflect the increasing demands imposed by having to discriminate between more and more lexemes composed of the same set of alphanumeric cues, which is in turn reflected in the increased accuracy of older participants in the lexical decision task. Or, to put it another way, *while we may learn more and more words across our lifetimes, we do not learn any more letters*; And although we can arrange letters in ways that yield more cues, even this process is not infinite. This means that over time, learning more vocabulary items must inevitably increase the information processing demands associated with letter cues. (This point also raises developmental issues, which we return to below.)

We should acknowledge that the models are hypotheses, and that compared to the complexities of actual human experience, their training is absurdly sparse. However, against this we should note that the learning algorithm in the model at least approximates our best understanding of how brains learn (Schultz et al, 1997; Schultz, 2006; Daw, Courville & Dayan, 2008; Daw et al, 2011), and that in practice, even noting their limitations, these models do seem capable of providing insights into human learning that are useful, if not exactly "right" (Box & Draper, 1987; see Ramscar & Yarlett, 2007; Gureckis & Love, 2010; Ramscar et al, 2010; Ramscar et al, 2011; Ramscar et al, 2013a; Ramscar et al, 2013b; Ramscar et al, 2013c; Arnon & Ramscar, 2013; Baayen et al, 2011; Baayen, Hendrix & Ramscar, 2013).

Moreover, some systematic insights into the complexities of learning and aging can only be gotten from a useful hypothetical model. To return to the empirical data plotted in Figure 5, it clearly suggests that there may be an age-related interaction between digit symbol performance and anagram solving: a model provides a framework in which to examine whether a causal relationship between these facts actually exists. For example, the model we just described suggests that the slowing in older adult's lexical decision responses is the product of loading more and more lexical outcomes on a relatively finite set of combinations of letter cues (see also Anderson, 1974).

We might expect that this will lead to (at least) two outcomes: First, just as virtually any PAL w_1 - w_2 pair will become more dissociated as lexical learning increases (Figure 3), we can expect that the association between virtually any two letter unigrams will decrease in exactly the same way for very similar reasons. This means that increased lexical learning will inevitably make the digit symbol task harder in the same way that it makes learning specific PAL pairings harder. However, given that the dissociations in letter *n-grams* are driven at least to some extent by their being learned as cues to more and more words, it follows that increased lexical learning will mean that any given *n-gram* is likely to cue more and more lexemes. Given the role of prefrontal cortex (PFC) in filtering response behavior (Shimamura, 2000; Chrysikou, Weber, & Thompson-Schill, in press), and the way that learning alters the dynamics of the responses that PFC serves to filter (Ramscar & Gitcho, 2007; Thompson-Schill, Ramscar & Chrysikou, 2009) and helps modulate perseverative behavior (Ramscar et al, 2013b), it is hardly surprising that learning to associate more lexical outcomes with *n-grams* serves to improve people's ability to solve anagrams, or that older scrabble experts should be particularly adept at this (Hargreaves et al, 2012).

Thus, just as a model can help us understand why there is an interaction between speed and accuracy in lexical decision with age, and why slower lexical decisions speed do not necessarily betoken "decline," so it can also help us understand why there is an interaction between digit symbol accuracy and anagram solving with age, and why performance on the two measures in the latter is as inextricably linked as speed and accuracy "performance" in the former.

Why better functional models of behavior across the lifespan are necessary to understanding the aging brain

An improved functional understanding of the behavioral changes we see in healthy adults as they age is necessary to determining which (if any) of these changes can be considered "decline." Model-based analyses (Davis, Love, & Preston, 2012a; 2012b; Turner et al., 2013) of brain imaging data and model selection methods (Kriegeskorte & Kievit, 2013; Mack, Preston & Love, 2013) have proven invaluable in interpreting brain activity for younger adults. If we are to understand the aging brain, better functional models will also need to be developed, applied, and evaluated. Thus, for example, studies employing a range of imaging techniques and a variety of tasks have revealed a posterior to anterior shift in patterns of task-related activation as adults age, and a concomitant decease in the degree to which tasks trigger activation patterns that are lateralized to one (in particular the left) hemisphere (Reuter-Lorenz et al. 2000; Cabeza et al. 1997, Grady et al. 2006; Cabeza 2002; Cabeza et al 2002; Cabeza et al 2004; see Park & Reuter-Lorenz, 2009 for a review). In the cognitive neuroscience literature this shifting pattern of activations is thought to reflect the results of a "scaffolding" process, in which contra-lateral and frontal areas step in to pick up the slack in the processing capacity of left-lateral and posterior areas that occurs in "direct response to the magnitude of neural insults that occur with age," (Park & Reuter-Lorenz, 2009).

Although this idea is widely accepted, the literature offers no details (nor even detailed speculations) as to quite how it is that frontal areas come to be aware of the plight of their insulted brethren in posterior regions (it seems reasonable to assume that dead cells don't talk), nor are accounts forthcoming of how it is that ensembles of neurons in functionally distinct areas of cortex are able to acquire the tunings that enable them to replicate the functions of circuits in other regions that have succumbed to the ravages of time. Finally, nor does the literature offer a suggestion, let alone a convincing explanation, of why it is that, for example, posterior regions are systematically more susceptible to insults than anterior regions across the entire human population as it ages.

By contrast, consider the model of the relationship between digit span performance and anagram solving we just described. It outlines a (clearly over-) simple feed-forward network in which the perception of letter cues in turn activate lexemes, and, in the case of anagram solving, any activated lexemes are then filtered along with other task relevant information in prefrontal cortex in order to generate a response. Over the lifetime, the model thus predicts a systematic pattern of changes will occur as a learner's lexical experience grows: First, in posterior regions associated with reading letter forms, such as left posterior occipitotemporal sulcus (pOTS; Mano et al, 2013) the model predicts that more experience will lead to less neural activation, because learning and cue-competition will increasingly serve to tune *n-gram* representations (this is consistent with findings from imaging studies, which show stronger activation of the left pOTS to pseudowords and low frequency words as compared to high frequency words; Kronbichler et al. 2004, 2007; Bruno et al. 2008; Schurz et al. 2010; Mano et al, 2013); Second, since these increasingly tuned *n-gram* representations will become associated with more and more lexemes, which will activate when they are activated, the model predicts that greater experience will result in more activation of the anterior regions associated with lexical processing itself, such as the superior temporal gyrus (STS), inferior frontal gyrus (IFG) and PFC (see e.g., Friederici, 2011), because activating increasingly specialized *n-gram* representations will result in the activation of increasingly larger sets of lexemes.

That is, a simple functional model of reading that incorporates learning can predict and explain the interaction between digit span performance and anagram solving and provide an account of the posterior to anterior shift in activation during lexical processing without having to make the many dubious assumptions involved in theories of "neural scaffolding." Further, the model we have outlined does so in ways that are highly compatible with our best understanding of the functions of the brain regions involved. Moreover, given that this model does not automatically assume that changing patterns of activation are evidence of "neural insults," it is more consistent with biological models of brain aging, which have revealed that the brains of healthy adults do not experience significant cell loss as they age, nor do they undergo dramatic changes in neuronal morphology (in a recent review, Burke & Barnes, 2006, describe the widely held beliefs to the contrary as "the myth of brain aging".)

The patterns of change in neuronal morphology over the lifespan are both more complex and more puzzling than the notion of "brain atrophy" embraced by the scaffolding hypothesis supposes. Most of the typical changes in brain morphology that are observed in healthy aging involve declines in the density and organization of neuronal dendrites and spines (gray matter) and axons (white matter). Although the typical pattern of change that is usually observed in many areas of the dorsal, frontal, and parietal lobes in adulthood involve reductions in grey matter density, in some brain areas, such as the cingulate gyrus, the density of grey matter appears to remain consistent across the lifespan in healthy adults (Sowell et al, 2003). Moreover, in some brain areas, such as the parahippocampal gyrus, there is evidence of significant dendritic growth in normal human aging (but not in senile dementia, Buell et al, 1979; 1981).

While it goes without saying that the complex and systematic pattern of changes that are actually seen in neural morphology are not going to be explained without the development of functional models of what brain systems actually do, the need for good functional models of cognitive processes is particularly acute because given the changes that typify aging in healthy brains, it is extremely difficult to disentangle “declines” in brain function from the effects of learning. This is because learning is itself reflected in neuronal morphology as changes in the density and composition of grey and white matter (see e.g., Merrill et al, 2001; Zuo et al, 2005; Rapp et al, 1996; Flood et al, 1991, 1993; Burke & Barnes, 2006; Zatorre, Fields & Johansen-Berg, 2012).

It follows from this that in order to be sure that all of the changes in neural morphology that ones sees in a healthy brain are insults, rather than signs of learning, one first needs a functional model of “normal” learning and processing. For example, studies of 11 – 17 year-olds have revealed patterns of changes in gray and white matter densities that are remarkably similar to those associated with aging (Alemán-Gómez et al, 2013). Should these findings be interpreted as a marker for the (extremely) early onset of age-related declines in neural plasticity, or as ordinary, business-as-usual learning?

Similarly, consider that more extensive age-related reductions in grey matter density are typically observed in the posterior temporal cortex in the left (as compared to right) hemisphere (Sowell et al, 2003). Are these differences, which are particularly evident in posterior language areas, really just the result of simple (and presumably random) insults, and atrophy caused by “brain aging?” While it is, of course, possible that they are, it seems highly likely that, given that language is one of the most extensive functional systems any brain ever learns, at least some of the systematic changes in neuronal morphology seen in posterior temporal cortex reflect the effect of learning this system. If so, then this means that, again, the process of distinguishing learning from decline – if indeed there is decline in healthy brain aging – is likely to be far more subtle and complex than the literature currently acknowledges.

Understanding lexical processing and understanding lexical learning

Rabbitt (2014) points out that our assumption that greater information processing loads result in slower processing appears to be confounded by empirical results that, in comparing across people of the same age, as opposed to between people of different ages, reveal that people who have larger vocabularies often perform better than people with smaller vocabularies.

> "people of any age whose brains are so stuffed with words that they can produce more names of animals within a fixed time also produce words in other categories correspondingly faster and more accurately. This does not support the Ramscar hypothesis that words are retrieved more slowly from a large vocabulary." (Rabbitt, 2014).

The reason these findings do not contradict our central hypothesis, which is, of course, that the mind and brain are not fixed systems:

> plasticity is not an occasional state of the nervous system; instead, it is the normal ongoing state of the nervous system throughout the lifespan. A full, coherent account of any sensory or cognitive theory has to build into its framework the fact that the nervous system, and particularly the brain, undergoes continuous changes in response to modifications in its input afferents and output targets. (Pascual-Leone et al, 2005)

It goes without saying that learning changes the brain. Just 7 days training in something as inconsequential as juggling is sufficient to produce visible changes in gray matter density and to the organization of white matter pathways in the occipito-temporal areas associated with the processing of complex visual motion (Draganski et al, 2004; Driemeyer et al, 2008); notably, these patterns of change are even visible in elderly participants (Boyke et al, 2008; albeit that the elderly learn less well on average over the same time frame).

Because learning changes the brain, prior learning always impacts subsequent learning. As our case study of PAL learning demonstrated, there is no such thing as "learning" in a vacuum. This in turn means that a full, coherent account of lexicial processing across the lifetime can't simply consider the effects of having a large vocabulary in a vacuum: In order to understand the interaction between experience, vocabulary size and processing, one has to consider how people end up with different sized vocabularies, and how this might affect learning and processing at different stages of linguistic development.

Studies of children clearly show that, consistent with Rabbitt's observation, larger vocabulary scores actually predict faster lexical processing in childhood. Children with larger vocabularies process words faster than children with smaller vocabularies (Fernald & Marchman, 2012; Bion, Borovsky & Fernald, 2013). Perhaps unsurprisingly, these studies have also shown that vocabulary scores and processing speeds are highly correlated with the amount of language a child is exposed to (Fernald & Marchman, 2012; Bion, Borovsky & Fernald, 2013; Weisleder & Fernald, 2013; Fernald, Marchman, & Weisleder, 2013). Moreover, as Hart & Risely (1995) revealed in their landmark studies, depending on the social environment a child grows up in, the amount of language she hears can differ dramatically.

These points are of particular importance when we are dealing with human brains, because in children, not only will learning be having an impact on the local morphology of areas processing the various factors that contribute to behavior, but also because the maturation and development of the overall structure of the human brain is occurring throughout childhood

(Ramscar & Gitcho, 2007; Thompson-Schill, Ramscar & Chrysikou, 2009). Given what we know about the way brains learn and develop, it seems at least reasonable to suppose that children who are exposed to large amounts of language develop much richer neural networks in the areas involved in lexical processing than children whose linguistic experience is impoverished. In developing the model we used to predict lexical processing speeds above, we considered the relation between network density and processing speed, and, in theory at least (in practice, we should acknowledge that this is a free parameter), the model predicts that dedicating more processing hardware to a task in the brain will lead to faster processing speeds.

Ultimately, we want to be able to do is integrate the many strands that influence the development of neural networks in the maturing mind, and the way processing in these networks responds to information gains in mature minds. We don't pretend for a second that our models are even close to doing all this. Yet consider the complexity involved in the task we just described, and in the interactions between experience, weights and learning in the analyses and simulations we described above. However simple and flawed the various models presented above are, they at least offer insight.

This observation highlights an important point in this debate: Researchers in the brain and cognitive sciences are engaged in a tortuous process of trying to reverse engineer a complex physical information processing device. Yet the simple fact is that very few researchers in the field have any training in information processing systems, and of the few that do, most have training at the software rather than the hardware end. Most researchers have only the dimmest idea how increases in data and task complexity impact information processing in the physical systems that actually do the processing. While it is clear that the brain is not a computer in a straightforward sense, the fact is that machine information processing still represents our best model of neural information processing (indeed, formally, it is the only model we have) and it seems highly unlikely that a “no model” approach will lead to meaningful progress in our search to understand the mind, and the effects of age on it.

Conclusion: Learning is not decline

Do the ravages of time inevitably result in declines in cognitive processing in otherwise healthy brains? We remain open to the possibility that the answer to this question is, “yes.” What we have sought to show above is that, when it comes to aspects of lexical learning where we now can quantify the environment in which learning takes place (in large part thanks to the development of large corpora), once one controls for the effects of learning on performance, there is precious little variance left to be described in terms of “decline.” Not only are the changing patterns of performance observed in paired-associate learning better accounted for by learning models than by vague notions of “cognitive decline,” but detailed comparison of the specifics of those performance patterns suggests that instead of declining with age, older adults’ lexical knowledge in fact becomes more and more attuned to the information structure of the lexicon.

Similarly, we have shown that if one simply attends to speed in lexical decision tasks, one will inevitably find evidence of decline. Whereas if one integrates a measure of accuracy into one’s analysis, a different picture emerges: a picture in which an improvement in one dimension – accuracy – is shown to come at a cost in another, speed.

In one sense, the findings we report are hardly surprising: they simply suggest that one rarely gets something from nothing, and that actions have reactions, and these are hardly new ideas. However, we would suggest what our findings reveal about our folk theories of the mind, and of the effect of aging on the mind ought to surprise us: It would appear that many of the implicit assumptions that serve to underpin received ideas about cognitive declines involve the idea that some things – especially learning – ought to come for free, and that some actions – especially the acquisition of knowledge – ought not to have a reaction. Our findings do not only undermine these implicit assumptions, they also undermine the muddled thinking that is embodied in received ideas about cognitive decline, both in the literature, and in society more generally.

We have sought to show how many of the tacit, over-simplified assumptions about the nature of learning in the literature are leading researchers to seriously overestimate of the degree to which cognitive function declines with age. We would not wish to argue that this means that functionality does not change. For instance it may be that a side-effect of some kinds of prior-learning is that subsequent learning is inhibited in ways that, essentially, amount to functional losses, in much the same way that children learning of a native sound system functionally impedes the later learning of non-native phonetic contrasts (Werker & Tees, 1984). Rather, we would suggest that a better understanding of learning can do much to assist our understanding of cognitive functions themselves (see e.g., Baayen et al, 2011), and the way these functions develop across the lifetime.

Finally, in relation to this last point, we should reiterate why all this is important. In numerous studies, Carol Dweck and her colleagues have shown how people who believe that their abilities can be improved through hard work learn far better than those who think that their abilities are fixed (Dweck, 2006, 1999; Dweck & Leggett, 1988; Yeager & Dweck, 2012; see also Mangels et al, 2012; Rattan, Good & Dweck, 2012; Dweck et al, 1978). Since it is clear from the findings we present here that people's ability to learn stays with them at all ages, and from the work of Dweck and colleagues that thinking of ability as a fixed factor has an adverse influence on children and younger adults' ability to learn, we can only shudder to think what the pervasive mythology of “cognitive decline” is doing to older adults’ ability to adopt a positive mindset, or to believe that their efforts can lead to improvement.

References

Alemán-Gómez, Y., Janssen, J., Schnack, H., Balaban, E., Pina-Camacho, L., Alfaro-Almagro, F., ... & Desco, M. (2013). The Human Cerebral Cortex Flattens during Adolescence. *The Journal of Neuroscience,* 33(38), 15004-15010.

Anderson, J. R. (1974). Retrieval of propositional information from long-term memory. *Cognitive Psychology*, 6(4), 451-474.

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

Baayen, R. H., Davidson, D. J., & Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of memory and language*, 59(4), 390-412

Baayen. R.H. (2001). *Word Frequency Distributions*. Amsterdam: Kluwer

Baayen, R. H., Milin, P., Đurđević, D. F., Hendrix, P., & Marelli, M. (2011). An amorphous model for morphological processing in visual comprehension based on naive discriminative learning. *Psychological Review*, 118(3), 438-482

Baayen, R. H., Hendrix, P. and Ramscar, M. (2013). Sidestepping the combinatorial explosion: Towards a processing model based on discriminative learning. *Language and Speech* 56, 329–347

Baayen, R. H., Feldman, L. B., & Schreuder, R. (2006). Morphological influences on the recognition of monosyllabic monomorphemic words. *Journal of Memory and Language*, 55(2), 290-313

Bion, R. A., Borovsky, A., & Fernald, A. (2013). Fast mapping, slow learning: Disambiguation of novel word–object mappings in relation to vocabulary learning at 18, 24, and 30months. *Cognition*, 126(1), 39-53.

Box, G. E., & Draper, N. R. (1987). *Empirical model-building and response surfaces*. John Wiley & Sons.

Borovsky, A., Elman, J.L., & Fernald, A. (2012). Knowing a lot for one's age: Vocabulary skill and not age is associated with anticipatory incremental sentence interpretation in children and adults. *Journal of Experimental Child Psychology*, 112(4), 417-36

Boyke, J., Driemeyer, J., Gaser, C., Buchel, C. & May, A. (2008). Training-induced brain structure changes in the elderly. *J. Neurosci*. 28, 7031–7035

Bruno JL, Zumberge A, Manis FR, Lu Z-L, Goldman JG. (2008). Sensitivity to orthographic familiarity in the occipito-temporal region. *Neuroimage*. 39:1988--2001.

Brink, S. (2014, January 28). An Aging Brain Is Still Pretty Smart [Web log article]. Retrieved from *National Geographic*, http://news.nationalgeographic.com/news/2014/01/140128-aging-brain-memory-cognitive-decline-neurology/

Buell, S. J. & Coleman, P. D. (1981). Quantitative evidence for selective dendritic growth in normal human aging but not in senile dementia. *Brain Res*. 214, 23–41

Buell, S. J. & Coleman, P. D. (1979). Dendritic growth in the aged human brain and failure of growth in senile dementia. *Science* 206, 854–856

Burke, S. & Barnes, C. (2006). Neural plasticity in the aging brain. *Nature Reviews Neuroscience* 7 (1): 30–40.

Carey, B. (2014, January 27). The Older Mind May Just Be a Fuller Mind [Web log article]. Retrieved from *The New York Times*, http://newoldage.blogs.nytimes.com/2014/01/27/the-older-mind-may-just-be-a-fuller-mind

Chrysikou, E. G., Weber, M. J., & Thompson-Schill, S. L. (2013). A matched filter hypothesis for cognitive control. *Neuropsychologia*.

Danks, D. (2003). Equilibria of the rescorla-wagner model. *Journal of Mathematical Psychology*, 47(2), 109–121.

Daw, N.D., Courville, A.C., Dayan, P. (2008) "Semi-Rational Models of Conditioning: The Case of Trial Order." in *The Probabilistic Mind*; Chater, N., Oaksford, M., Eds.; Oxford University Press: Oxford.

Daw, N. D., Gershman, S. J., Seymour, B., Dayan, P., & Dolan, R. J. (2011). Model-based influences on humans' choices and striatal prediction errors. *Neuron*, 69(6), 1204-1215.

Davies, M. (2012) "Expanding Horizons in Historical Linguistics with the 400 million word Corpus of Historical American English". *Corpora* 7: 121-57.

Davies, M. (2009). The 385+ Million Word Corpus of Contemporary American English (1990-present). International *Journal of Corpus Linguistics* 14: 159-90.

Davis, T., Love, B.C., & Preston, A.R. (2012a). Learning the Exception to the Rule: Model-Based fMRI Reveals Specialized Representations for Surprising Category Members. *Cerebral Cortex*, 22, 260-273.

Davis, T., Love, B.C., & Preston, A.R. (2012b). Striatal and Hippocampal Entropy and Recognition Signals in Category Learning: Simultaneous Processes Revealed by Model-based fMRI. Journal of Experimental Psychology: Learning, Memory, and Cognition, 38, 821-839.

Deary, I.J. et al. (2009). Age-associated cognitive decline. *Brit. Med. Bull.* 92, 135-152 des Rosiers G & Ivison D. (1986). Paired-associate learning: normative data for differences between high and low associate word pairs. *Journal of Clinical Experimental Neuropsychology*; 8: 637–42

Draganski, B., Gaser, C., Busch, V., Schuierer, G., Bogdahn, U., & May, A. (2004). Neuroplasticity: changes in grey matter induced by training. *Nature*, 427(6972), 311-312.

Driemeyer, J., Boyke, J., Gaser, C., Büchel, C., & May, A. (2008). Changes in gray matter induced by learning—revisited. *PLoS One*, 3(7), e2669.

Dweck, C. S., Davidson, W., Nelson, S., & Enna, B. (1978). Sex differences in learned helplessness: II. The contingencies of evaluative feedback in the classroom and III. An experimental analysis. *Developmental psychology*, 14(3), 268

Dweck, C. S. (1999). *Self-theories: Their role in motivation, personality and development*. Philadelphia: Psychology Press.

Dweck, C. S., & Leggett, E. L. (1988). A social-cognitive approach to motivation and personality. *Psychological review*, 95(2), 256.

Dweck, C. S. (2006). Mindset: The new psychology of success. New York: Random House.

Fernald, A., Marchman, V. A. & Weisleder, A. (2012). SES differences in language processing skill and vocabulary are evident at 18 months. Developmental Science,16(2), 234-248.

Fernald, A., & Marchman, V. A. (2012). Individual Differences in Lexical Processing at 18 Months Predict Vocabulary Growth in Typically Developing and Late‐Talking Toddlers. *Child Development*, 83(1), 203-222.

Flynn J. R. (1987). Massive IQ gains in 14 nations: What IQ tests really measure. *Psychological Bulletin* 101: 171–191

Friederici, A. D. (2011). The brain basis of language processing: from structure to function. *Physiological Reviews*, 91(4), 1357-1392.

Flood, D. G. (1991). Region-specific stability of dendritic extent in normal human aging and regression in Alzheimer's disease. II. Subiculum. *Brain Res*. 540, 83–95

Flood, D. G. (1993). Critical issues in the analysis of dendritic extent in aging humans, primates, and rodents. *Neurobiol. Aging* 14, 649–654

Gureckis, T.M. and Love, B.C. (2010) Direct Associations or Transformations of an Internal State? Exploring the Mechanism Underlying Sequential Learning Behavior. *Cognitive Science*, 34, 10-50

Hart, B & Risley, T.T. (1995) *Meaningful Differences in the Everyday Experience of Young American Children*. Brookes Publishing.

Hess, T. M., Auman, C., Colcombe, S. J., & Rahhal, T. A. (2003). The impact of stereotype threat on age differences in memory performance. *The Journals of Gerontology Series B: Psychological Sciences and Social Sciences,* 58(1), P3-P11.

Hess, T. M., & Hinson, J. T. (2006). Age-related variation in the influences of aging stereotypes on memory in adulthood. *Psychology and Aging*, 21(3), 621.

Hess, T. M., Hinson, J. T., & Hodges, E. A. (2009). Moderators of and mechanisms underlying stereotype threat effects on older adults' memory performance. *Experimental Aging Research*, 35(2), 153-177.

Kriegeskorte, N. & Kievit, R.A. (2013). Representational geometry: integrating cognition, computation, and the brain. *Trends in Cognitive Science*, 17(8), 401-412.

Kronbichler M, Bergmann J, Hutzler F, Staffen W, Mair A, Ladurner G. (2007). Taxi vs. Taksi: orthographic word recognition in the left ventral occipitotemporal cortex. *J Cogn Neurosci*. 19:1584-1594.

Kronbichler M, Hutzler F, Wimmer H, Mair A, Staffen W, Ladurner G. (2004). The visual word form area and the frequency with which words are encountered: evidence from a parametric fMRI study. *Neuroimage*. 21:946-953.

Kruschke, J. K. (2003). Attention in learning. *Current Directions in Psychological Science*, 12, 171-175

Lynn, R. (1982). IQ in Japan and the United States shows a growing disparity. *Nature* 297 (5863): 222–223.

Mack, M.L., Preston, A.R. & Love, B.C. (2013). Decoding the Brain's Algorithm for Categorization from its Neural Implementation. *Current Biology*, 23, 2023-2027.

Mangels, J. A., Good, C., Whiteman, R. C., Maniscalco, B., & Dweck, C. S. (2012). Emotion blocks the path to learning under stereotype threat. *Social cognitive and affective neuroscience*, 7(2), 230-241

Mano, Q. R., Humphries, C., Desai, R. H., Seidenberg, M. S., Osmon, D. C., Stengel, B. C., & Binder, J. R. (2013). The role of left occipitotemporal cortex in reading: reconciling stimulus, task, and lexicality effects. *Cerebral Cortex*, 23(4), 988-1001.

Merrill, D. A., Chiba, A. A. & Tuszynski, M. H. (2001) Conservation of neuronal number and size in the entorhinal cortex of behaviorally characterized aged rats. *J. Comp. Neurol.* 438, 445–456

Morris RG, Moser EI, Riedel G, Martin SJ, Sandin J, Day M, & O'Carroll C (2003) Elements of a neurobiological theory of the hippocampus: the role of activity-dependent synaptic plasticity in memory. *Philos Trans R Soc Lond, B, Biol Sci* 358:773–786.

Nelson, D. L., Dyrdal, G. M., & Goodmon, L. B. (2005). What is preexisting strength? Predicting free association probabilities, similarity ratings, and cued recall probabilities. *Psychonomic Bulletin & Review*, 12(4), 711-719

Nusbaum, H.C. A Stochastic Account of the Relationship between Lexical Density and Word Frequency. *Indiana University, Research on Speech Perception*, Progress Report #11 (1985)

O'Donnell C, Nolan MF, & van Rossum MC. (2011) Dendritic spine dynamics regulate the long-term stability of synaptic plasticity." *The Journal of Neuroscience* 31.45: 16142-16156.

Park, D. C., & Reuter-Lorenz, P. (2009). The adaptive brain: aging and neurocognitive scaffolding. *Annual Review of Psychology*, 60, 173.

Pascual-Leone, A., Amedi, A., Fregni, F., & Merabet, L. B. (2005). The plastic human brain cortex. *Annu. Rev. Neurosci.*, 28, 377-401.

Pereira, A. F., Smith, L. B., & Yu, C. (2014). A bottom-up view of toddler word learning. *Psychonomic bulletin & review*, 21(1), 178-185

Rabbitt, P. (2014, February 1). *Age and the overstuffed mind* [Web log article]. Retrieved from Outlook from the Hutch: A blog on aging and the mind, http://pmar76.wordpress.com/2014/02/06/age-and-the-overstuffed-mind/

Ramscar, A. (2014, January 21). What happens to our minds and memories in healthy aging? [Web log article]. Retrieved from The Importance of Being Wrong, http://ramscar.wordpress.com/2014/01/19/what-happens-to-our-minds-and-memories-in-healthy-aging/

Ramscar, M., Hendrix, P., Shaoul, C., Milin, P., & Baayen, H. (2014). The Myth of Cognitive Decline: Non-Linear Dynamics of Lifelong Learning. *Topics in Cognitive Science* 6, 5-42.

Ramscar, M., & Gitcho, N. (2007). Developmental change and the nature of learning in childhood. *Trends in Cognitive Science*, 11(7), 274–279

Ramscar, M. & Yarlett, D. (2007) Linguistic self-correction in the absence of feedback: A new approach to the logical problem of language acquisition. *Cognitive Science,* 31, 927-960

Ramscar, M.J.A., D.G. Yarlett, M.W. Dye, K. Denny & K. Thorpe. (2010). The Effects of Feature-Label-Order and their implications for symbolic learning. *Cognitive Sci*. 34(6), 909-957

Ramscar M, Dye M, Popick HM, O'Donnell-McCarthy F (2011) The enigma of number: Why children find the meanings of even small number words hard to learn and how we can help them do better. *PLoS ONE* 6(7): e22501. doi:10.1371/journal.pone.0022501

Ramscar, M., Dye, M. & Klein, J. (2013a) Children value informativity over logic in word learning. *Psychological Science*, 24(6), 1017-1023.

Ramscar, M., Dye, M., Gustafson, J.W., & Klein, J. (2013b) Dual routes to cognitive flexibility: Learning and response conflict resolution in the dimensional change card sort task. *Child Development*, 84(4), 1308-23.

Ramscar, M., Dye, M. & McCauley, S. (2013c) Error and expectation in language learning: The curious absence of 'mouses' in adult speech. *Language*, 89(4), 760-793

Rapp, P. R. & Gallagher, M. (1996). Preserved neuron number in the hippocampus of aged rats with spatial learning deficits. *Proc. Natl Acad. Sci.* USA 93, 9926–9930

Ratcliff, R., Thapar, A. & McKoon, G. (2011). Effects of Aging and IQ on Item and Associative Memory. *Journal of Experimental Psychology: General*, 140(3), 464-487.

Ratcliff, R., Thapar, A. & McKoon, G. (2010). Individual differences, aging, and IQ in two-choice tasks. *Cognitive Psychology*, 60, 127-157.

Rattan, A., Good, C., & Dweck, C. S. (2012). “It's ok—Not everyone can be good at math”: Instructors with an entity theory comfort (and demotivate) students. *Journal of Experimental Social Psychology*, 48(3), 731-737.

Rescorla R.A. (1968). Probability of Shock in the Presence and Absence of CS in Fear Conditioning. *J. Comp. & Physiol. Psychol.*, 66, 1-5

Rescorla, R.A. & Wagner, A.R. (1972). A theory of Pavlovian conditioning: Variations in the effectiveness of reinforcement and nonreinforcement. In: Black & Prokasy (Eds.), *Classical Conditioning II: Current Research and Theory*. New York: Crofts

Rescorla, R.A. (1988). Pavlovian conditioning: It's not what you think it is. *American Psychologist*, 43, 151–160.

Rizzuto, D S., & Kahana, M J. "An autoassociative neural network model of paired-associate learning." *Neural Computation* 13.9 (2001): 2075-2092.

Salthouse, T.A. (2009). When does age-related cognitive decline begin? *Neurobiol. Aging* 30, 507-14

Salthouse. T.A. (2011). Consequences of age-related cognitive declines. *Ann. Rev. Psychol.* 63, 5.1–5.26

Schaie, K. W. (1977). Quasi-experimental research designs in the psychology of aging. *Handbook of the psychology of aging*, 1, 39-69

Schaie, K. W. (1974). Translations in gerontology--from lab to life. Intellectual functioning. *The American Psychologist*, 29(11), 802

Schaie, K. W. (1973). Methodological problems in descriptive developmental research on adulthood and aging. *Life-span developmental psychology: Methodological issues*, 253-280

Schaie, K. W. (1959). Cross-sectional methods in the study of psychological aspects of aging. *Journal of Gerontology*, 14(2), 208-215

Schaie, KW. (1988). Internal validity threats in studies of adult cognitive development. In: Howe, ML.; Brainerd, CJ., editors. *Cognitive development in adulthood: Progress in cognitive development research*. New York: Springer-Verlag, p. 241-272

Scholz, J., Klein, M.C., Behrens, T.E. & Johansen-Berg, H. (2009). Training induces changes in white-matter architecture. *Nat. Neurosci*. 12, 1370–1371

Schultz W (2006). Behavioral theories and the neurophysiology of reward. *Annual Review of Psychology*. 57, 87-115

Schultz, W., Dayan, P., & Montague, P. R. (1997). A neural substrate of prediction and reward. *Science*, 275(5306), 1593-1599.

Schurz, M., Sturm, D., Richlan, F., Kronbichler, M., Ladurner, G., & Wimmer H. (2010). A dual-route perspective on brain activation in response to visual words: evidence for a length by lexicality interaction in the visual word form area (VWFA). *Neuroimage*. 49:2649--2661

Shimamura, A. P. (2000). The role of the prefrontal cortex in dynamic filtering. *Psychobiology*, 28, 207-218.

Singh-Manoux, A. et al. (2012). Timing of onset of cognitive decline: Results from Whitehall II prospective cohort study. *British Medical Journal*. 344, d7622

Sowell ER, Peterson BS, Thompson PM, Welcome SE, Henkenius AL, Toga AW (March 2003). "Mapping cortical change across the human lifespan". *Nat. Neurosci*. 6 (3): 309–15

Steele, C. M. (1997). A threat in the air: How stereotypes shape intellectual identity and performance. *American Psychologist* 52 (6): 613–629

Teasdale TW & Owen DR (2005). A long-term rise and recent decline in intelligence test performance: The Flynn Effect in reverse. *Personality and Individual Differences* 39 (4): 837–843.

Turner, B. M., Forstmann, B. U., Wagenmakers, E. J., Brown, S. D., Sederberg, P. B., and Steyvers, M. (2013). A Bayesian framework for simultaneously modeling neural and behavioral data. *NeuroImage*. 72, 193-206.

Weisleder, A. & Fernald, A. (2013). Talking to children matters: Early language experience strengthens processing and builds vocabulary. *Psychological Science*, 24(11), 2143-2152.

Werker, J. & R. Tees. (1984). Cross-language speech perception: Evidence for perceptual reorganization during the first year of life. *Infant Behavior and Development*. 7, 49-63

Yeager, D. S., & Dweck, C. S. (2012). Mindsets that promote resilience: When students believe that personal characteristics can be developed. *Educational Psychologist*, 47(4), 302-314.

Zatorre, R. J., Fields, R. D., & Johansen-Berg, H. (2012). Plasticity in gray and white: neuroimaging changes in brain structure during learning. Nature Neuroscience, 15(4), 528-536.

Zucker RS (1999) Calcium-andactivity-dependentsynapticplasticity. *Curr Opin Neurobiol* 9:305–313.

Zuo Y, Lin A, Chang P, Gan WB (2005) Development of long-term dendritic spine stability in diverse regions of cerebral cortex. *Neuron* 46:181– 189.

Author Note

This work was supported by an Alexander von Humboldt research award to R. Harald Baayen.

Appendix

Table A1: Fixed-effects estimates in a mixed-effects logistic regression model with word as random-effect factor fitted to the visual lexical decision latencies of old and young participants. Data available in the languageR package (Baayen, 2008), dataset "english". Model fitted with the glmer function from the lme4 package (Bates et al, 2013).

Fixed effects:

	Estimate	Std. Error	z value	Pr(>|z|)
Intercept	1.40760	0.07432	18.940	< 2e-16
WrittenFrequency	0.36847	0.01515	24.320	< 2e-16
AgeSubject=young	-1.23280	0.05348	-23.053	< 2e-16
WrittenFrequency:AgeSubject=young	0.06914 0	.01220	5.666	1.46e-08

R code for the simulation of the PAL task:

```
lex = read.table("obeyRulesLex.txt", T)


lex

         Cues                  Outcomes       Frequency

1        Context_American      Eagle          1
2        Context_Obey          Rules          1
3        Context_Obey           Eagle         1
4        Context_Legal         Eagle          1


ibrary(ndl)
wmlist = list()
for (i in 1:50) {
  cuesOutcomes = lex
  cuesOutcomes$Frequency[1:2] = rep(i, 2)
  wmlist[[i]] = estimateWeights(cuesOutcomes)
  }

frequencies = 1:50
strength = sapply(wmlist, FUN=function(m)return(m[4,1]))
plot(frequencies, strength, ylim=c(-0.5, 0),type="l",
  xlab="frequency of the stock phrases",
  ylab="cue strength of Obey and Eagle")
  abline(h=0, col="darkgray")
```

Table A2: Coefficients in a linear mixed effects model (with Item as random-effect factor) for the results plotted in Figure 4. Slope coefficients estimate the slopes for each age group separately. The t-tests evaluate whether a slope is zero. A separate analysis using treatment contrasts (not shown) indicated that all interactions were highly significant. F1: frequency of the first word (background rate); P2: ratio of the frequency of the first and second word (blocking); Google: google frequency of the first and second word pair (association rate).

```
                  Estimate Std. Error t value
Intercept          2.55064    0.68220   3.739
Sex=Male          -0.10740    0.02211  -4.858
F1:AgeGroup20     -0.36897    0.14189  -2.600
F1:AgeGroup30     -0.55493    0.14189  -3.911
F1:AgeGroup40     -0.60279    0.14189  -4.248
F1:AgeGroup50     -0.65100    0.14189  -4.588
F1:AgeGroup60     -0.73176    0.14189  -5.157
P2:AgeGroup20     -0.80059    0.41103  -1.948
P2:AgeGroup30     -1.08672    0.41103  -2.644
P2:AgeGroup40     -1.23984    0.41103  -3.016
P2:AgeGroup50     -1.28778    0.41103  -3.133
P2:AgeGroup60     -1.44944    0.41103  -3.526
Google:AgeGroup20  0.34482    0.06214   5.549
Google:AgeGroup30  0.46774    0.06214   7.528
Google:AgeGroup40  0.50119    0.06214   8.066
Google:AgeGroup50  0.53815    0.06214   8.661
Google:AgeGroup60  0.59641    0.06214   9.599
```

Table A3: Specification of the generalized additive mixed model predicting log response latency in visual lexical decision from age, log frequency, and two latent variables for orthographic consistency.

Parametric coefficients:

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	6.661213	0.001773	3757.7	<2e-16
age=young	-0.221721	0.001862	-119.1	<2e-16

Approximate significance of smooth terms:

	edf	Ref.df	F	p-value
smooth frequency for age = old	6.555	7.290	218.335	< 2e-16
smooth frequency for age = young	6.245	6.975	229.127	< 2e-16
tensor product PC1 x PC2 for age = old	7.081	8.147	6.804	5.62e-09
tensor product PC1 x PC2 for age = young	3.829	4.147	2.122	0.073
random intercepts word	341.173	2192	1.601	< 2e-16