The myth of cognitive decline

Michael Ramscar, Peter Hendrix & Harald Baayen
Seminar für Sprachwissenschaft, Wilhelmstraße 19
Universität Tübingen, Tübingen, Germany

Abstract

Across a range of psychometric tests, reaction times slow as adult age increases. These changes have been widely taken to show that cognitive-processing capacities decline across the lifespan. Contrary to this, we suggest that slower responses are not a sign of processing deficits, but instead reflect a growing search problem, which escalates as learning increases the amount of information in memory. A series of computational simulations show how age-related slowing emerges naturally in learning models, as a result of the statistical properties of human experience and the increased information-processing load that a lifetime of learning inevitably brings. Once the cost of processing this extra information is controlled for, findings taken to indicate declines in cognitive capacity support little more than the unsurprising idea that choosing between or recalling items becomes more difficult as their numbers increase. We review the implications of this for scientific and cultural understanding of aging.

Keywords: Learning; Language; Memory; Psychometric Testing;

The Age of Tithonus

More and more people live longer and longer lives. Outside of 18 countries the UN describes as 'outliers' (Watkins et al, 2005), increased life expectancy and declining birth rates are raising median ages in populations across the globe. By 2030, 72 million Americans will be aged 65 or older, a twofold increase from 2000. The world's population is more aged than ever before in history, and its rate of aging is increasing.

People are clearly living longer; it is less clear that this is a blessing. In Greek mythology, Tithonus was the mortal lover of Eos, goddess of the dawn. While asking Zeus to make Tithonus immortal, Eos forgot to mention "eternal youth," dooming Tithonus to an eternity of decrepit babbling. The psychological and brain-sciences endorse the Tithonean view of aging, portraying adulthood as an extended period of mental decline: memories dim; thoughts slow; problem-solving abilities diminish (Naveh-Benjamin & Old, 2008; Deary et al, 2009); and each year, the onset of cognitive decrepitude is set ever younger (Salthouse, 2009; Singh-Manoux et al., 2012). One crumb of comfort is that older adults are, on average, happier (Charles & Carstensen, 2010), although in the circumstances, this might be taken as further evidence of their declining mental prowess.

In what follows, we show how the slowing response speeds that are taken as evidence of "cognitive decline" in adults emerge naturally in learning models (Baayen et al, 2011) as knowledge increases. These models, which are supported by a wealth of psychological (Ramscar et al, 2010) and neuroscientific (Schultz, 2006) evidence, correctly identify greater variation in the cognitive performance of older adults, successfully predicting that older adults will show more sensitivity to fine-grained differences in test items than younger adults. The models run (and can be rerun) on computers, eliminating the possibility that aging hardware influences their performance, which instead reflects the information-processing costs incurred as knowledge increases. Once the demands of processing this extra information are taken into account, it becomes clear that much of the evidence for age-related declines in cognitive capacity better supports the idea that information processing costs rise as the amount of information in a system increases.

The problem with "processing speed"

Findings from a range of psychometric tests suggest that the rates at which the mind processes information increase from infancy to young adulthood, and decline steadily thereafter (Salthouse, 2011). Increasing reaction times are a primary marker for age-related cognitive decline (Deary et al, 2010), and are even considered its cause (Salthouse, 1996), yet they are puzzling. Practice improves speed and performance on individual cognitive tasks at all ages (Dew & Giovanello, 2011). Since we get more practice using our cognitive capacities as we age, why does our performance on tests of them decline?

The answer lies in the way that psychometric tests neglect learning, and its relationship to the statistical patterns characteristic of human life. Learning is a discriminative process that serves to locally reduce the information processing costs associated with various aspects of knowledge and skill (Rescorla & Wagner, 1972). However, age increases the <u>range</u> of knowledge and skills individuals possess, which increases the overall <u>amount</u> of information processed in their cognitive systems. This extra processing has a cost.

Learning and the long tail of linguistic experience

Statistically, the distribution of human experience is highly skewed: Much of our day-to-day life is fairly repetitive, involving a small repertoire of common occurrences, such as reading the newspaper and going to work. At the same time, we experience a far more diverse repertoire of infrequent or even unique occurrences (we rarely read the exact same newspaper twice). When data is distributed like this, comparisons of means are often meaningless. Consider the problem of remembering birthdays: We are reminded of the birthdays of family members on an annual basis, and this usually makes us expert at remembering them. However, as we move through life, we also learn about other birthdays. Sometimes we hear these dates only once, such as when we attend a party for someone we barely know. As each new birthday is learned, our mean exposure to all the birthdays we know will decline, and the task of recalling a particular birthday will become more complex. Recalling six hundred birthdays with 95% accuracy need not imply a worse memory than recalling six with 99% accuracy.

Standard psychometric tests do not take account of the statistical skew of human experience, or the way knowledge increases with experience. As a result, when used to compare age groups, they paint a misleading picture of mental development. This can be demonstrated most clearly in relation to language. Language is a central aspect of cognition, its statistics are more readily quantified than other aspects of human experience, and all psychometric tests involve some linguistic information processing.

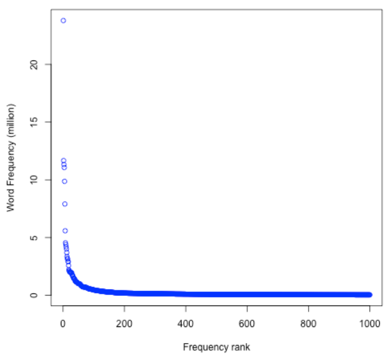

Figure 1. The frequencies of the 1000 most common words in the Corpus of Contemporary American English plotted by rank.

Importantly, linguistic distributions are skewed at every level of description (Baayen, 2001). Consider the relationship between word types (e.g., *dog*) and tokens (how often "*dog*" occurs; Figure 1). In any large sample of English, a few words occur very frequently (*the*, *and*), such that half of a typical sample comprises tokens of only 100 or so high-frequency types. The relative frequency of these types decreases rapidly (the most-frequent word may be twice as frequent as the second-most), while frequency differences between types decrease as their relative frequency declines. This means that the other half of a typical sample is composed of ever-fewer tokens of a very large number of types, with ever-smaller frequency differences between them. Typically, around half of these types occur just once.

This is a very long-tailed distribution: 49% of the 425 million tokens in the Corpus of Contemporary American English (COCA; Davies, 2009) come from the 100 most-frequent word types; the remaining 51% of tokens represent over 2.8 million word types. Although individual low-frequency types are, by definition, rare, their distribution makes the chances of encountering a low-frequency token in any given sentence extremely high (Möbius, 2003). This distribution ensures that any English speaker learns only a fraction of its total vocabulary, and that vocabularies grow steadily across the lifespan. However, the tests used to measure cognitive decline assume that vocabulary size is age-invariant in adults (Spearman, 1927; Carroll, 1993; Bowles & Salthouse, 2008), an assumption seemingly confirmed by psychometric vocabulary measures, which suggest that vocabulary growth in adulthood is marginal (such that slight increases are only reliably detected in meta-analyses; Verhaeghen, 2003).

Psychometric vocabulary measures are virtually guaranteed to register these results, because they attempt to extrapolate vocabulary size from sets of test words. These tests, which are "normed" on the knowledge of schoolchildren, are heavily biased towards frequent word-types (Raven, 1965; Heim, 1970; Wechsler, 1997). Unfortunately, while extrapolation is feasible for frequent words, for the millions of low-frequency word-types, knowledge of one randomly sampled word does not predict knowledge of another. Since the distribution of types ensures that adult vocabularies overwhelmingly (and increasingly) comprise low-frequency words, it follows that reliably extrapolating their size or growth from a small test sample is mathematically impossible (Baayen, 2001).

Simulating the effects of vocabulary learning on information processing

Most infants are sensitive to all the fine-grained phonetic discriminations made by the world's languages. As they learn a native vocabulary, this sensitivity to non-native phonetic distinctions diminishes (Werker & Tees, 1984). Rather than indicating that cognitive decline begins in infancy, this loss in sensitivity can be seen as an inevitable result of learning. In discriminative learning models, the values of initially undifferentiated sets of cues are shaped by experience, which drives the discovery of cue values that best predict a learning environment (Rescorla, 1988). Because this process involves learning to ignore uninformative cues, it can explain why decreasing sensitivity to uninformative phonetic information goes hand in hand with increasing knowledge about informative phonetic distinctions (Ramscar et al, 2010).

The learning component of the model we use to simulate the effects of experience on reading works in precisely this way. It is an extension of the Naive Discriminative Reader (NDR; Baayen et al, 2011), a two-layer network in which letter unigrams and bigrams serve as input cues, and lexical items serve as outcomes. The values of these cues are initially undifferentiated, and are set competitively as the model learns to predict words from the letters it 'reads.' In the NDR, every predictive cue is linked to each lexical outcome to form a set of subnets. The cue-weights in these subnets are set by the equilibrium equations of the Rescorla-Wagner learning rule (Danks, 2003), and are completely determined by the distributional properties of the model's training corpus. Simulated latencies derived from these weights accurately capture a wide variety of empirical effects in reading (Baayen et al, 2011).

To model the influence of experience on different populations, the NDR was modified to make it sensitive to the physical and informational consequences of knowledge growth. Given that the amount of activation a given cue receives from the perceptual system remains constant over time (Attwell & Laughlin, 2001), this modification keeps the total amount of activation spreading from cues to outcomes equal to the amount of activation arriving at them. Analogous to the principle of conservation of electric charge, this means that as vocabulary increases, and each cue becomes connected to an increasing number of outcomes, the amount of activation arriving at any given outcome decreases. Given a vocabulary of size V, the network support for any item i is proportional to a_i/V where a_i is the activation an item receives from the cues in the input.

This modification also accounts for the effects that an increased number of outputs have on information processing in neural systems (Hentchel & Barlow, 1991; supplementary materials). Shannon's *source coding theorem* shows that the smallest coding scheme for V words requires, on average, $H(V)$ bits. Since V determines the length of a message in a given code, the effective channel capacity C of an ensemble of neurons decreases as code complexity increases and the amount of redundancy in signals across the network decreases (Hentchel & Barlow, 1991). We denote these information costs by $f(V)$, where f is an unknown non-decreasing function expressing the coding and signaling costs in a vocabulary of size V.

The response latency (RT) associated with reading (operationalized as reaction times to speeded judgments on written words) is modeled as a weighted sum of these components:

$$RT_i = w_1 V / a_i + w_2 f(V) + c$$

with c a constant denoting the time required for response execution.

To simulate the effects of vocabulary-growth on adult reading, two NDR networks were trained on data drawn from the Google Trigrams Corpus (a large, naturalistic data set). The first network 'read' 500,000 word-trigram tokens, simulating reading to age 21, the typical age for "young adult" participants in studies; the second 'read' 3,000,000 word-trigram tokens,[3] simulating reading to age 70 (the typical age for "old adults"). Consistent with our analysis of the way linguistic distributions influence vocabulary growth, the old model acquired a much larger vocabulary: 32,536 word types, compared to the young model's 21,307 (Figure 2). These growth estimates are very conservative: the Trigram Corpus excludes trigrams with less than 40 occurrences, thereby omitting around 50% of the word types in the complete Google Corpus. Even with this constrained input, vocabulary expansion was far from asymptote, even after 5 million trigram tokens.

To examine the models' ability to simulate age-related reading differences, we compared their projected reading times for 2,904 English words to empirical latencies from older (mean age 73.6) and younger (21.1) readers for the same items (Balota et al, 1999). The empirical data exhibit the expected effect of age: mean reaction times are 163 ms shorter for younger than older adults. Simulated reaction times mirror this difference, with an average difference of 167 ms.

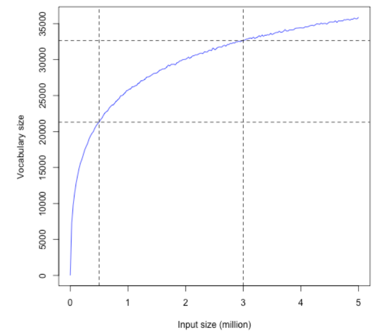

Figure 2. Empirically observed vocabulary growth after sampling from the Google Trigrams Corpus.

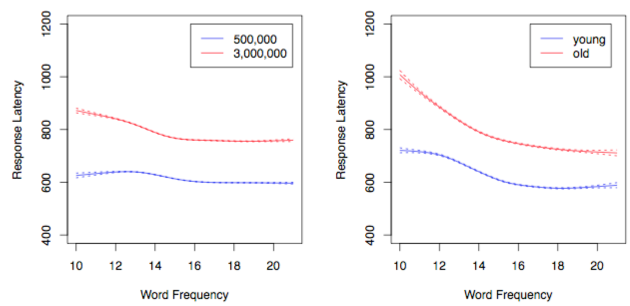

Figure 3. Left panel: fit of a generalized additive model to the simulated response latencies from the old and young models. Right panel: fit of a generalized additive model to the empirical response latencies from young (mean age: 21.1) and old (73.6) adults (Balota et al, 1999).

The models also correctly predict an important qualitative difference in the empirical word-frequency effect. It is well established that lexical decision responses are slower for lower- (e.g., "*whelp*") than higher-frequency words ("*where*"). This overall effect of frequency is present for both young and old adults (Figure 3; right panel). However, while frequency effects asymptote at higher frequencies in both models, they only level off at the lowest frequencies in the younger model, a pattern that is also observed in the empirical data: older adults are far better attuned to frequency variations in the lower range of the test-set than younger adults.

These results can be explained by considering the way the models learn in more detail. In learning, weights on the links between the cues and outcomes are adjusted in two ways: They are strengthened whenever a cue and outcome co-occur; For example, the link between the bigram WH and the lexical target WHERE is strengthened when "*where*" is encountered in reading, and the link between WH and WHELP is strengthened when the "*whelp*" is encountered. Conversely, links are weakened when cues occur but outcomes do not.

1	BLASH	11	CROME	21	TWERP	31	WHELP	41	BLEAT
2	SCHNOOK	12	GIBE	22	THWACK	32	SHUCK	42	CHIVE
3	LETCH	13	LISLE	23	DAUNT	33	MOOCH	43	WHIR
4	ZOUNDS	14	FLAYS	24	RETCH	34	JELL	44	CROON
5	JAPE	15	SPLOTCH	25	GYP	35	GROUCH	45	TAMP
6	SOUSE	16	VELDT	26	YAWL	36	AWN	46	BOSH
7	WHIG	17	SLOE	27	FLUB	37	MANSE	47	RILE
8	FILCH	18	CONK	28	STANCH	38	WRACK	48	BLANCH
9	RHEUM	19	FRAPPE	29	PAUNCH	39	HOOCH	49	LILT
10	PARCH	20	SKULK	30	JOWL	40	FLECK	50	JEER

Table 1. The 50 lowest frequency items in the set used to test the models and the older and young adults; BLASH has the lowest frequency of these items, and JEER the highest. As can be seen, many of the letter bigrams in this set of words are comparatively rare in English.

Thus when "*where*" is encountered, WH occurs without WHELP, weakening the link between WH and WHELP. The distribution of high-frequency words results in their being encountered frequently, at a fairly constant rate over time. This will consistently reinforce the link between WH and WHERE, and consistently weaken the link between WH

In contrast, low-frequency words occur sporadically, so the link between WH and WHELP is reinforced far less (and the link between WH and WHERE weakened less). These imbalances result in "selection pressures" on word types that are reflected in the distribution of orthographic (and phonetic) cues across lexical items (see Zipf, 1949): high-frequency test items are both shorter ($t(2901) = -10.58$, $p < 0.001$) and have higher mean bigram frequencies ($t(2901) = 8.98$, $p < 0.001$) than low-frequency items. This means that, on average, low-frequency words contain both more cues, and more rare cues (Table 1). Although rare cues have relatively high values in small vocabularies, they are vulnerable to competition as vocabularies grow: newer vocabulary items also have low frequencies, and are more likely to contain the same rare cues.

All the predicted empirical effects were replicated in an analysis of a second, independent dataset (Figure 4).

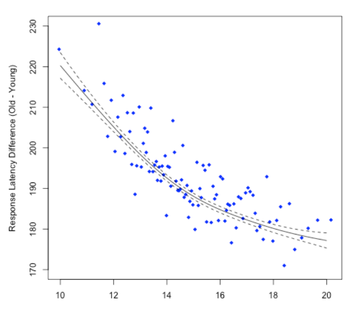

Figure 4. Average percentile RT differences (old – young) for the naming latencies of 2,820 Single Syllable words (Yap et al, 2011) by young (mean age: 22.6) and old adults (73.6), plotted against the words' log frequency in the Google 1-gram corpus, and a generalized additive model fit to the RT differences. As can be seen, the difference between older and younger readers' RTs increases as word frequency decreases.

Modeling 'decline' in a non-lexical task

To examine whether the relationship between information load and response time also holds for "non-lexical" tests, we considered the *letter classification task* (Posner & Mitchell, 1967), a standard non-lexical psychometric test in which two letters are presented in upper or lowercase (A, a, D, d, E, e, R, r, H, h) and participants judge whether they are alphabetically the same or different. Older subjects are typically slower than younger subjects in this task, a finding that is straightforwardly replicated in the NDR models once the coupling between letters and abbreviated meanings is accounted for (e.g, *H* for entropy, *R* for a statistical programming environment, *r* for correlation, etc.). The network complexity function *f(V)* in (1), which models response latencies as a function of the activation of the meanings of both letters in a letter pair, predicts longer latencies for older subjects as compared to younger subjects. In short, because the older model has a larger system of outcomes, it has more information to process, making "accessing" a letter harder, and reaction times concomitantly slower (see also Ramscar et al, 2010).

Psychometrically, letter classification is often described as an "information-processing" measure, and older adults' longer response times are taken as evidence of declining information-processing capacity. Yet information theory—which defines the workings of information-processing system—is, at heart, a set of methods for formalizing the uncertainty in distributions (be they bits of code, or vocabulary items; Shannon, 1948). Information is a property of systems, and processing demands are measured in relation to them (MacKay, 2003). In letter classification, the system comprises the task, a participant, and, crucially, what that participant knows. Because psychometric tests neglect this knowledge, they are formally incapable of measuring information-processing in this task.

Lexical knowledge and paired-associate learning

Wherever vocabulary size increases with experience, this increased knowledge predicts increasing processing costs and slower responses in psychometric tasks. As a consequence, slower latencies reflect learning, not "decline." Interestingly, this interaction between experience, vocabulary-size and response speed can also be seen in comparisons of monolingual and bilingual picture-naming (Gollan et al, 2008): the response latencies of young-bilinguals more closely resemble older-monolinguals than younger-monolinguals or older-bilinguals. Notably, slower response times and increased tip-of-the-tongue rates are not taken as evidence of cognitive decline when observed in young-bilinguals (Gollan & Acenas, 2004), but are instead seen to reflect the demands of processing the larger vocabularies that bilinguals inevitably learn.

The finding that bilinguals experience increased tip-of-the-tongue rates raises a question: could the same systemic effects of learning that account for increased lexical processing latencies explain age-related change in memory measures, such as Paired-Associate Learning (PAL; a psychometric measure of people's ability to learn and recall new information)? In PAL, e.g., the subtest of Wechsler's Memory Scale (WMS; desRosiers & Ivison, 1988) participants have to learn more or less arbitrary pairings between word cues (e.g., *baby; jury*) and responses (*cries; eagle*). Although item-level performance is highly variable (Figure 5), older adults' overall PAL performance is slower and less accurate, and it has been suggested that aging causes encoding (Gilbert, 1941; MacKay & Burke, 1990) and retrieval processing deficits (Burke & Light, 1981).

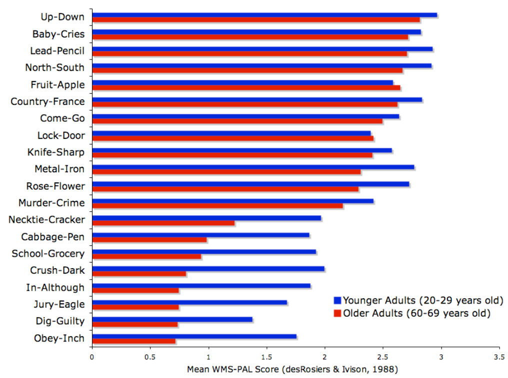

Figure 5. Mean performance by item for 100 older (age 60-69) and 100 younger (20-29) adults on forms 1 and 2 of the WMS-PAL subtest (desRosiers & Ivison, 1988). As in the lexical decision and naming data, the relationship between old and young PAL performance is nonlinear: again, older adults exhibit a more marked ability to discriminate between 'harder' (unrelated) and 'easier' (related) items than younger adults.

There is, however, no reason to think PAL performance should be age-invariant. Long-established principles of associative learning predict that well-known items should be harder to learn as Cues (w_1) than newer items (Rescorla & Wagner, 1972). Likewise, newer Response (w_2) items should support better learning than well-known, predictable items (Kamin, 1969): w_1-w_2 pairs ought to become harder to learn when w_1 and w_2 occur independently at high rates (Rescorla, 1968; compare *jury-eagle* to *baby-cries*).

To examine whether age-related PAL differences simply reflect learning, we analyzed the relationship between the age-related variance in the performance of a large sample adults on the WMS-PAL subtest (desRosiers & Ivison, 1988), and the factors that determine w_1-w_2 learnability. In a regression analysis of item score differences (mean young – mean old), w_1 predictability (log frequency; t=-4.063, p<0.001), the relationship between w_2 and w_1 predictability (log(w_2 frequency) / log(w_1 frequency); t=-2.935, p<0.01) and actual w_1-w_2 co-occurrence rates (log Google frequency; t=6.773, p<0.0001) accounted for more than 75% of the variance in item performance between 20-29 and 60-69 year-olds (F(3)=16.432, r=.87, p<0.0001).

All things being equal, the relative learnabilty of w_1-w_2 pairs can be estimated from w_1-w_2 co-occurrence and background rates. All things are *not* equal, however: Older adults have more experience, and learnability is a matter of experience. Accordingly, w_2 words will become more predictable the more they occur independently of w_1, and w_1 words will become less informative the more they occur independently of w_2; in each case, experience will make w_1-w_2 learning harder. A natural, predictable consequence of this is that PAL performance should increasingly reflect the distributional properties of the w_1-w_2 items as experience grows: if co-occurrence rates are low, a lifetime of learning that *jury* is uninformative about *eagle* should make learning *jury-eagle* harder; whereas high co-occurrence rates will reduce background rate effects, making *baby-cries* easier for older adults to learn relative to *jury-eagle*.

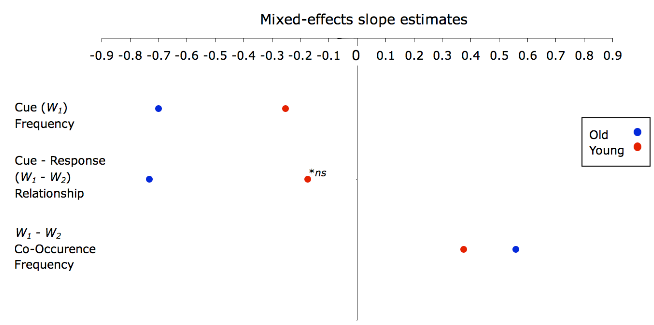

Figure 6. Mixed-effects slope estimates for the three learnability predictors and mean item performance of old (60-69) and young (20-29) adults in the WMS-PAL subtest (desRosiers & Ivison, 1988). All predictor effects and interactions in the model are significant (see supplementary materials), and all slopes (except *) are significantly different from 0 (t=>2). Older adults are more sensitive to *background rate* information (negative slopes) than young adults and, as the magnitude of the slopes shows, the overall performance of older adults reflects a far more systematic understanding of the English language.

A mixed-effects analysis of w_1-w_2 item scores by age confirmed the accuracy of this prediction (Figure 6). For each predictor, the magnitude of the slope for the older age group is greater than that for the younger age group, indicating that older subjects bring more lexical experience to the task. Consistent with our earlier findings, older adults' PAL performance reflects their greater knowledge of (and sensitivity to) the distributional properties of w_1-w_2 words, whereas younger adults' less varied performance reflects their more limited knowledge of them. As we noted above, the statistical properties of human experience makes comparing means invidious: in this case, it seems that high mean PAL performance is a measure of ignorance, not "intelligence."

Learning and Cognitive Maturation

These results suggest that older and younger adults' performance in psychometric testing largely reflects the same cognitive mechanisms, confronted with the task of processing different quantities of information. The performance of older adults on these tests is evidence of increased knowledge, not declining processing capacity.

When discussing these conclusions with colleagues, a question often arises: *"Learning seems to predict linear patterns of change, but cognitive decline seems to kick in around age 60 or 70: how do you explain this?"* To explain why, we first note that as people age, they encode less contextual information in memory (Naveh-Benjamin & Old, 2008). Although this has been taken as evidence that the processes that "bind" contextual information in memory decline with age, learning theory predicts that experience will increasingly make people insensitive to context, because ignoring less informative cues is integral to learning.

Learning is also sensitive to the environment, and its predictions change with it: If a common environmental change—e.g., retirement—was to systematically reduce the variety of contexts people typically encounter in their lives, learning theory predicts that the amount of contextual information they learn will also drop, as the background rates of cues in remaining contexts rise. If these same people were to increasingly spend their time in environments where cues already have very high background rates (e.g., family homes), this effect will be exacerbated. In other words, because learning inevitably reduces sensitivity to everyday context, retirement is likely to make memories harder to individuate and more confusable, absent any change in cognitive processing, simply because it is likely to decrease contextual variety at exactly the time when, as a result of learning and experience, the organization of older adult's memories needs it most.

Learning can explain both the apparent changes in older adults "cognitive performance" around retirement-age, and the fact that these changes are not detected in testing. Similarly, the neglect of learning in the study of cognitive aging makes it highly likely that, like *Tithonus*, many of our beliefs about cognitive decline are myths. This does not mean that the diseases that can undermine cognition in old age are also mythical: However our understanding of these diseases can only be increased by a better understanding lifelong learning, and its sensitivity to the environment.

References

Attwell D, & S.B. Laughlin. (2001) An energy budget for signaling in the grey matter of the brain. *J Cereb Blood Flow Metab* 21: 1133–1145.

Baayen, R.H. et al. (2011) An amorphous model for morphological processing in visual comprehension based on naive discriminative learning. *Psychol. Rev*. 118, 438-482.

Baayen. R.H. *Word Frequency Distributions*. (Kluwer Academic Publishers, 2001).

Balota, D.A. & D.H. Spieler. (1998) The utility of item level analyses in model evaluation. *Psychol Sci* 9, 238-240.

Burke, D. M., & Light, L. L. (1981) Memory and aging: The role of retrieval processes. *Psychol. Bulletin*, 90, 513-546.

Carroll, J.B. *Human cognitive abilities: A survey of factor-analytic studies*. (Cambridge University Press, 1993).

Danks, D. Equilibria of the Rescorla-Wagner model. (2003) *J. Math. Psychol*. 47(2), 109-121.

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

Deary, I.J. et al. (2009) Age-associated cognitive decline. *Brit. Med. Bull*. 92, 135-152.

desRosiers G, Ivison D. (1986) Paired-associate learning: normative data for differences between high and low associate word pairs. J *Clin Exp Neuropsychol*; 8: 637–42.

Dew, I.T.Z. & K.S. Giovanello. (2010) The status of rapid response learning in aging. *Psychol. Aging* 25(4), 898-910.

Gilbert, J.G. (1941) Memory loss in senescence. J. *Abnormal & Social Psychology*, 36, 73-86.

Gollan, T.H. et al. (2008) Aging, bilingualism, and the weaker links hypothesis. *J. Mem. Lang.* 58, 787-814.

Hale, S., S.D. Lima & J. Myerson. (1991) General Cognitive Slowing in the Nonlexical Domain: An Experimental Validation. *Psychol. Aging* 6(4), 512-521.

Hayes, D. P. & M. Ahrens. (1988) Vocabulary Simplification for Children: A Special Case of 'Motherese'? *J Child Lang* 15, 395-410.

Heim, A.W. (1970) *AH 4 group test of general intelligence*. NFER-Nelson.

Hentchel, H.G.E. & H.B. Barlow. (1991) Fininding minimum entropy codes with Hopfield Networks, *Network*, 2, 135-148.

Kamin L.J. (1969) Predictability, surprise, attention, and conditioning. In: Campbell B, Church R (eds). *Punishment and Aversive Behaviour*. Appleton-Century-Crofts: New York.

MacKay, D. G., & Burke, D. M. (1990) Cognition and aging: A theory of new learning and the use of old connections. In T. Hess (Ed.), *Aging and cognition: Knowledge organization and utilization* (pp. 213-263). Amsterdam: North-Holland.

Naveh-Benjamin, M. & S.R. Old. (2008) Aging and memory (humans). In J.H. Byrne, H. Eichenbaum, R. Menzel, H.L. Roediger, & D. Sweatt (Eds.). *Learning and memory: A comprehensive reference* (2008).

Posner, M.J. & R.F. Mitchell. (1967) Chronometric analsis of classification. *Psychol. Rev*. 74, 392-409.

Ramscar, M. et al. (2010) The Effects of Feature-Label-Order and their implications for symbolic learning. *Cognitive Sci*. 34(6), 909-957.

Raven, J.C. *Guide to using the Mill Hill vocabulary test with progressive matrices*. (HK Lewis, 1965).

Rescorla, R.A. & A.R. Wagner. A theory of Pavlovian conditioning. In: Black & Prokasy (Eds.), *Classical Conditioning II: Current Research and Theory*. (Appleton-Century-Crofts, 1972)

Salthouse, T.A. (1996) The processing-speed theory of adult age differences in cognition. *Psychol. Rev*. 103, 403-428.

Schultz, W. (2006). Behavioral theories and the neurophysiology of reward. *Ann. Rev. Psychol*. 57, 87-115.

Shannon, C.E. (1948) A mathematical theory of communication. *Bell Systems Technical Journal* 27(3), 379-423.

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

Verhaeghen, P. Aging and vocabulary score: A meta-analysis. *Psychol. Aging* 18, 332–339 (2003).

Watkins, K. et al. *UN Human Development Report: International Cooperation at a Crossroads-Aid, Trade and Security in an Unequal World*. (United Nations Development Programme, 2005).

Wechsler, D. *Wechsler Adult Intelligence Scale–Third Edition* (WAIS-III). (Harcourt Assessment, 1997).

Yap, M.J., D.A. Balota, D.E. Sibley & R. Ratcliff. (2011) Individual Differences in Visual Word Recognition: Insights From the English Lexicon Project. *J Exp Psychol Hum Percept Perform* 38(1), 53-79.

Zipf., G. K. *Human behavior and the principle of least-effort*. (Addison-Wesley, 1949).

For a complete list of **references** and **supplementary materials**, please consult the article copy hosted at *www.sfs.uni-tuebingen.de/*