Semantic representations in monolingual and bilingual connectionist networks

Nicholas Rendell (n.rendell@bbk.ac.uk)
Eddy J. Davelaar (e.davelaar@bbk.ac.uk)

Department of Psychology, Birkbeck College. University of London
Malet Street, London WC1E 7HX

Abstract

A neural network model is presented which investigates the suggestion that being multilingual contributes to an individual's level of cognitive reserve. Two versions of this model were produced, one which learnt the names of input representations in a single language and another model which learnt input representations in two languages. The languages in both models were split further into two semantic categories. The relationship between the representations of both semantic categories in the first language of each of the two versions was investigated. Further manipulations came in the form of changing the sensitivity of the artificial neurons in the network and varying the hidden layer size with respect to variable levels of brain reserve. Findings were not immediately interpretable in terms of age-related decline in the absence of a behavioral measure. However, the variance in trajectories of category separation provide a cautionary tale against the interpretation of any measures gained at discreet intervals.

Keywords: ageing; connectionist model; language; bilingualism

Introduction

Recent studies of bilingual and multilingual individuals have demonstrated some offsetting of normal cognitive ageing (Kavé, Eyal, Shorek, & Cohen-Mansfield, 2008) and protective effects again the onset of the cognitive symptoms of dementia (Bialystok, Craik, & Freedman, 2007). These effects are presented in opposition to linguistic deficits also reported in bilinguals (Bialystok, 2008). This study examines both of these consequences by comparing representations of picture categories in monolingual and bilingual networks.

The contribution of language in offsetting age-related cognitive deficits is one of a number of factors, known collectively as the latent variable cognitive reserve (Stern, 2003, 2009). The existence of the variability in levels of protective factors is evidenced in studies which have demonstrated a poor relationship between an individual's cognitive intactness *in vivo* and levels of brain pathology post mortem (Mortimer, 1997; Valenzuela & Sachdev, 2006). Given the number of different factors which contribute to cognitive reserve, it may be difficult to untangle them. Years of education has been linked to the ability to stave off age-related decline (Albert et al., 1995; Barnes, Tager, Satariano, & Yaffe, 2004; Scarmeas, Albert, Manly, & Stern, 2006). Therefore, it is reasonable to assume years of education, childhood intelligence or, more simply, a nurturing environment may be a moderating factor for the relationship between multilingualism and cognitive reserve.

More recent studies of multilingualism and its association with cognitive reserve have attempted to control for education and intelligence. Bak, Nissan, Allerhand, & Deary (2014) utilised the Lothian Birth Cohort, a group of English native speakers of European origin who were initially tested for a level of intelligence at age 11 in 1947. This allowed the authors to control for childhood intelligence, gender and socioeconomic status. The participants, now 73, were tested on fluid intelligence, memory, speed of information processing, reading and verbal fluency. The results demonstrated a protective effect of bilingualism with no negative effects of having more than one language. Reading verbal fluency and general intelligence were the most affected and general intelligence in particular was related to improvement in executive processes. Of note was the similarity in performance between active (using second language) and passive (not required to use second language) bilinguals. This contradicts a view of cognitive reserve resulting from the continual practice of cognitive mechanisms. However, the increase in general intelligence in both active and passive bilinguals suggests that the effect of initial use of a second language is sufficient to upgrade cognitive processes. This is also demonstrated in the advantages for acquiring a second language in later life.

To understand why bilingualism confers an advantage to cognitive ageing, the cognitive mechanisms involved in speaking more than one language must be unpacked. In terms of nonverbal effects, these are wholly positive. Initial findings in a study comparing English only speaking Canadian children with their French- English speaking counterparts on verbal and nonverbal tests found that the bilingual children outperformed the monolinguals in almost all aspects, especially the nonverbal intelligence tests (Peal & Lambert, 1962). Equivalence was found in visual perception but advantages were found in symbol manipulation. Such early findings may subscribe to the criticism of a lack of control for potential confounds (Bialystok, 2001). However, the study demonstrated at least the potential for cognitive improvement in bilingual individuals. Further, the apparent improvement in nonverbal abilities for increased effort in the cognitive domain of language refutes the modular notion of cognitive processing (Bialystok, Craik, Green, & Gollan, 2009). Studies in metalinguistic capabilities have uncovered the mechanisms behind the cognitive advantage in bilingual individuals. For example, Bialystok (1988) found that bilingual children demonstrated an advantage in tasks requiring cognitive control. Further, in error checking and explanation of ungrammatical sentences, Galambos & Goldin-Meadow (1990) found that bilingual children performed better in the trials which required a change in the focus of attention. However, both monolinguals and bilinguals performed equally on the actual explanation of the errors. Nonlinguistic studies have also demonstrated benefits for bilinguals in executive control tasks using perceptual stimuli (Costa, Hernández, & Sebastián-Gallés, 2008). This suggests that bilingualism provides a holistic strengthening of executive control processes. This assertion has been supported with neuroimaging studies which demonstrate stronger resting-state connectivity in the frontal lobe for bilingual rather than monolingual individuals (Grady, Luk, Craik, & Bialystok, 2015; Luk, Bialystok, Craik, & Grady, 2011). Gold (2014) asserts that increased activity with frontal regions as a result of bilingualism serves to protect against age-related decline within those circuits related to executive processing.

Whilst the cognitive advantages of bilingualism appear well documented, the linguistic deficits associated with having more than one language are equally well-researched. For example, it is generally accepted that one of the predominant negative effects of bilingualism is the vocabulary size. This is generally smaller compared to monolinguals for both languages spoken (Mahon & Crutchley, 2006; Portocarrero, Burright, & Donovick, 2007). However, equivalence in vocabulary size for L1 between monolinguals and bilinguals has been found in very young children (age 24 months; Poulin-Dubois, Bialystok, Blaye, Polonia, & Yott, 2013). In addition to size of lexicon, bilinguals also appear to have more trouble accessing particular words. Picture naming tasks have shown that bilinguals are slower than their monolingual counterparts (e. g. Gollan, Montoya, Fennema-Notestine, & Morris, 2005). Further, verbal fluency tasks in which participants are asked to name as many words as possible for a given category or categories, have demonstrated a disadvantage for bilinguals (e.g. Rosselli et al., 2000). Tip of the tongue (Gollan & Acenas, 2004) errors are also more frequent in multilingual speakers and it is also reported that bilinguals have trouble identifying specific words through noise (Rogers, Lister, Febo, Besing, & Abrams, 2006). The aim of this study was to investigate any differences in the development of storage of representations between monolingual and bilingual groups of simulants. Two neural network models were trained to remember the names of a number of 'pictures' in one (monolingual) or two (bilingual) languages. Ageing of the networks was simulated by adjusting the gain of the transfer function (Li, Lindenberger, & Sikström, 2001; Servan-Schreiber, Printz, & Cohen, 1990).

Method

Architecture

The model in this study was a simple three layer, feedforward back propagating neural network. Two versions were produced, a monolingual version and a bilingual version. The hidden layer was varied in size for this investigation since it represented a view of passive reserve (Stern, 2009) which could easily be manipulated. For both models, the hidden layer size was manipulated to contain 5, 10, 15 or 20 nodes.

Stimulus Patterns

Given the focus of study for the models was representation storage in the hidden layer rather than performance, a compromise between an artificial language and a realistic corpus was used for input. The inputs used in both models were patterns of 26 binary digits. The first 20 digits were randomised with the addition of a further six inputs. The first three of these represented a language tag. This was added to the experimental paradigm to guarantee separation of the two sets of pictures in the bilingual model since the rest of the input consisted of a random pattern. The final three binary digits of each input presentation related to the membership of semantic category A or B. 34 input patterns were used in the monolingual model and the monolingual input set was augmented with a further 34 patterns for the bilingual model, making 64 in total for the bilingual model.

The input 'words' used for the monolingual network were taken from a dataset of English phonemes which had been converted to a binary input set using a set of 19 features (Thomas & Karmiloff-Smith, 2003). The input set for the monolingual network comprised of 34 English words with a further 34 Greek words produced for the bilingual model. The English words (L1) were used both in the monolingual and bilingual model and the Greek words represented the second language in the bilingual model (L2). Both monolingual and bilingual models had 40 output nodes. In the monolingual model, the first 19 nodes in each output vector were taken up by English words whilst the rest of the output nodes were left at zero. This set of output vectors was the same for the first 34 output vectors in the bilingual model. However, for the second 34 output vectors the last 21 units were used in the rest of the outputs since they related to the Greek names for the input patterns.

Training

Both networks were initially trained for 800 epochs. For comparison, test data was introduced to both monolingual and bilingual networks in the form of both categories of L1 only. 50 simulants were trained in this manner, the starting weights for each was seeded randomly from a uniform distribution of between 0 and 1. All of the following analyses represent mean scores. Training for both networks took around 200 epochs for the error to reach an asymptotic state. Overall, error settled at a slightly higher level in the bilingual network. This can be attributed to the increase in constraints in the bilingual network as it needed to accommodate the same amount of 'pictures' as the monolingual network but in both languages. Given that an asymptotic state was achieved around 200 epochs, it was decided that at 220 epochs the network was considered mature. It may be that there is a considerable 'grace period' during which the brain does not immediately decline after maturity. However, for the purposes of this study, aging can be said to begin upon reaching an asymptotic state, both for the purposes of analysis and interventions.

Analysis

Using a methodology similar to Thomas (1998) the difference between monolingual and bilingual modes in terms of the storage of representations for both categories of L1 was investigated. To this end, a scatterplot was produced by carrying out multidimensional scaling on the Euclidean distances between the activation vectors in response to each picture input (see Figures 1. & 2.). After multidimensional scaling was applied, the inputs were divided into the categories A & B and the three dimensions were plotted to illustrate any differences of semantic storage in representational space.

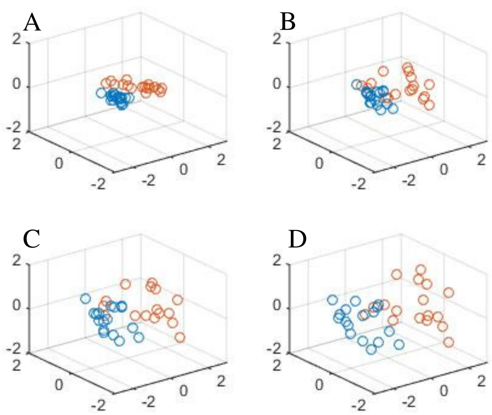

Figure 1: Scatterplots representing the distributions of representations of categories A and B within L1 of the monolingual network. Each graph refers to hidden layer sizes of five (A), ten (B), fifteen (C) and twenty (D) nodes. The blue dots relate to category A and the red dots relate to picture representations in category B.

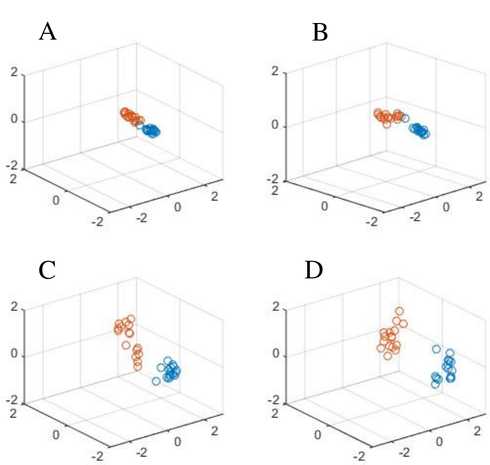

Figure 2: Scatterplots representing the distributions of representations of categories A and B within L1 of the bilingual network. Each graph refers to hidden layer sizes of five (A), ten (B), fifteen (C) and twenty (D) nodes. The blue dots relate to category A and the red dots relate to picture representations in category B.

The main analysis within this study relied on the online calculation of the separation of categories A and B within L1. This was achieved through the calculation of a single centroid L1. This provided the distances to an overall mean. Further, distances to the representations within each category from the overall centroid were calculated. This provided a measure of within and overall variance upon which to calculate an F value. This was calculated at each epoch for each simulant over the four different hidden layer sizes.

Results

The overall trend in categorical separation is driven by the feature differences between the two categories. Spreading of the categories can be seen progressing over the spectrum of hidden layer sizes for both models. However, the effect appears greater with the monolingual model. Conversely, clustering of representations within the hidden layer of the bilingual model is tighter. This was confirmed by cluster analysis of category A from L1 in both monolingual and bilingual networks carried out over the period of training (Figure 3.). This demonstrates that overall, higher hidden layer size networks showed the greatest spacing between representations with the two most dispersed categories belonging to the monolingual network. A 2*2 ANOVA carried out on the distances from the individual scores of the simulants at maturity with monolingual or bilingual as one factor and hidden layer as the other. The analysis demonstrated main effects for both hidden layer ($F(3,392) = 1539$, $p<.001$) and type of network ($F(1,392) = 516$, $p<.001$) together with an interaction between the two ($F(3,392) = 1953$, $p<.001$). Therefore, it appears that such an effect may be due to the differences in space cause by the additional constraints of second language storage offset only by a higher storage capacity.

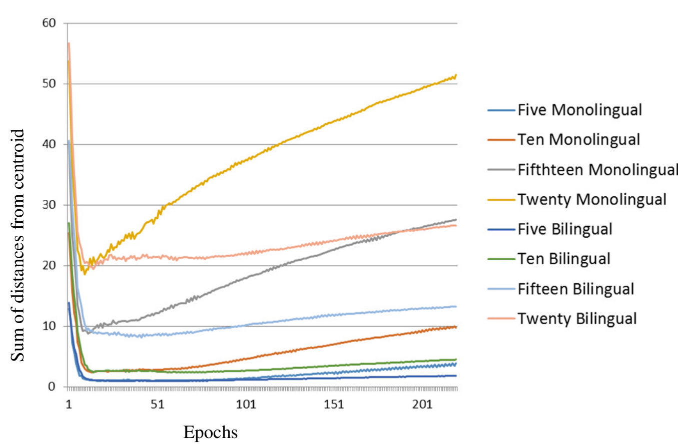

Figure 3: Line graph demonstrating the projection of the sum of the distances from calculated centroid in category for monolingual and bilingual models over all hidden layer sizes. Lines represent mean score of 50 simulants.

In order to provide a more valid interpretation of biological change over lifespan, the model integrated a gradual decline in the slope of the log sigmoidal transfer function. This manipulation reflects an age related reduction in dopamine which in turn relates to cognitive decline and a generally greater susceptibility to neural noise as the network becomes less discerning (Bäckman et al., 2010; Li et al., 2001; Servan-Schreiber et al., 1990).

Firstly, gain was set to decline gradually in steps of .0015 from the beginning of training. However, differentiation of either category could not be achieved. Therefore, a necessary and more valid representation of dopamine attenuation over lifespan was achieved by initiating the decline of gain after maturity, in this case, 220 epochs (Figure 4.)

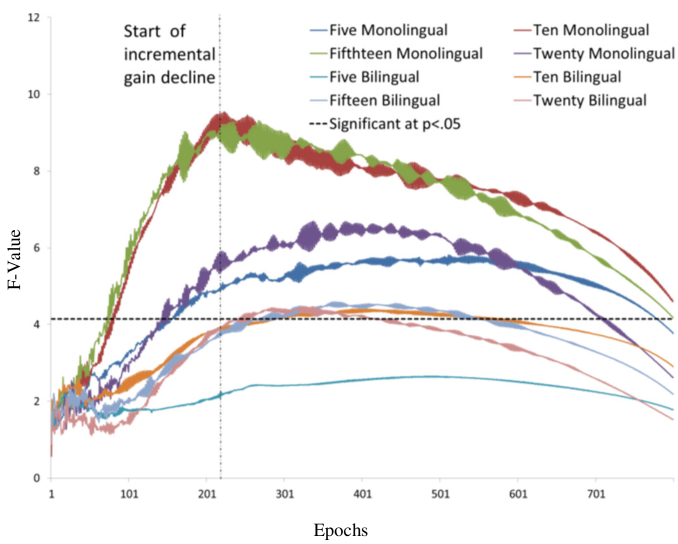

Figure 4: Projections of F-values as a measure of the dedifferentiation between categories A & B in L1 for both models. The age-related decrease in sensitivity of the artificial neurons in the network is represented by the gain decline. This started at 220 epochs and decreased gradually until the end of training at 800 epochs.

Inline analysis of the separation items from both categories demonstrated an increasing separation of semantic information up until maturity for all groups. However, characteristics of the different projections differ in response to gain decline. Dedifferentiation appears almost immediately for ten and fifteen node monolingual models. However, for the other sizes of hidden layer for monolingual and all bilingual hidden layer sizes, dedifferentiation in the representational separation of categories for all hidden layer sizes for both monolingual and bilingual models. Differences in the projections over time were investigated between the ten hidden layer networks for monolingual and bilingual networks. Firstly a bounded line graph was produced (Figure 5.) This demonstrated a greater variability in scores for the monolingual simulants.

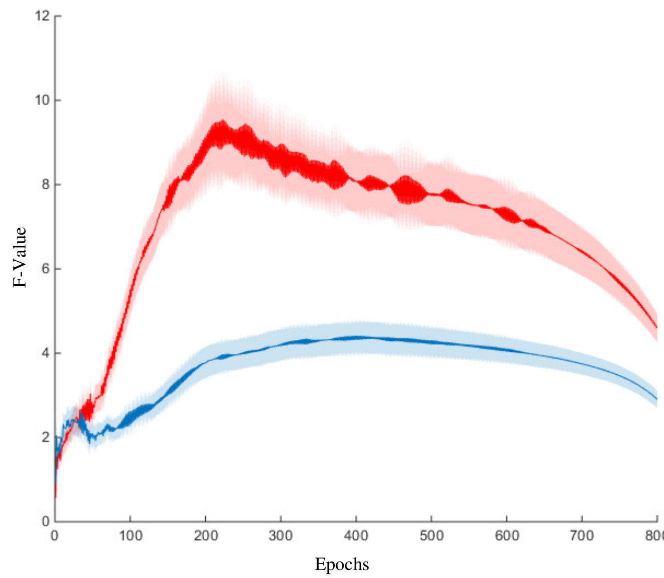

Figure 5: Bounded line graph of monolingual (red) and bilingual (blue) mean F-Value scores for ten unit hidden layer versions only. The shaded area around each line is one standard error of the mean.

Multilevel analysis was used in which individual scores were used as the dependent variable with the epoch and as the first level predictor. The second level was grouped by whether the model was monolingual or bilingual. This was estimated as a random effect due to the differing projections between monolingual and bilingual.

$$Y_{ij} = \gamma_{00} + u_{0j} + r_{ij}$$

Where Y represents the F-value at Epoch i for group j. A Likelihood ratio test demonstrated a significantly better fit with the inclusion of the random component ‘group’ (p<.001).

Discussion

This study represents an initial attempt at exploring the way in which bilingualism influences how categories within languages are represented. The models presented in this paper represent the first attempt at modelling and analyzing multilingualism from a representational perspective. This study explored differences between monolingual and bilingual models, each with differing levels of hidden layer size, taken as a proxy of brain reserve capacity. Both models were trained over a number of epochs, representing a lifespan. Further, a change in the sigmoidal transfer function slope was included after asymptote. The results of this study provide commentary on the way in which picture representations separate according to input characteristics. Further, the results also provide an explanation of the behavioral observations of multilingual individuals.

The analysis carried out in this study suggests two main effects of bilingualism on semantic memory. Firstly, Categories within a single representational space in a bilingual speaker are sensitive to space. Equivalence on hidden layer size or brain reserve capacity leads to a greater contraction and overlap of representations that the monolingual equivalent. Secondly, gain change produces differing effects according to the constraining factors. Ten and fifteen node hidden layer monolingual networks declined in representational separation almost immediately. All other networks continued to separate to a degree before declining.

A trend toward poorer separation and greater clustering together of the bilingual versions could lead to the observed deficits in lexical access. The slower reaction times observed in verbal fluency tasks (Bialystok, 2008) may be related to the inability to separate the individual representations to the same degree of success as their monolingual counterparts. This provides a different perspective to the weaker link hypothesis (Gollan, Montoya, Cera, & Sandoval, 2008). The weaker link hypothesis suggests that links between representations are weaker due to the relative lack of use compared to monolingual individuals who would use one language all of the time rather than share communication between two languages. Here, we suggest that links may be too close between representations. Further, recall errors in lexical decision making tasks may be due to the smaller distances and greater overlap observed in bilingual representational space. What might be expected from smaller clustering of representations is an increased search speed. However, the speed accuracy tradeoff may account for the lower than expected speeds observed in behavioral studies of lexical recall.

Given the executive nature of the positive contribution of multilingualism to cognitive reserve, it is difficult to relate the changes occurring in representational space to executive processes. However, what this study has indicated is that caution must be made when interpreting behavioral results gained from individuals over discreet periods of time. This is due to both the within group variance and the between category differences in projections over differing amounts of passive reserve and cognitive reserve, the former represented by differing amounts of hidden layer units and the latter represented by multilingualism. Specifically, a result of note is the steeper trajectory of decline near end of lifespan. This differed between mono and bilingual groups with a steeper decline observed in monolingual models. Further research is required to relate the representational spacing to behavioral outcomes. Further, it is important to note that the variability, especially within monolingual simulants, demonstrates the importance that must be placed on monitoring the progression of clinical outcomes for a single individual. The difference in projections demonstrate that behavioral outcomes may vary according to point in an individual's lifespan that testing has occurred as well as the individual differences within a group. However, in order to fully relate performance to the changes in representational clustering demonstrated in this study, further research is required.

A novel analysis of the representational space was carried out on simple three layer networks portraying monolingual and bilingual speakers. The results demonstrate that some of the negative effects of multilingualism may be due to space constraints rather than relative underuse of multiple languages. Further, this research raises questions as to the efficacy of testing at discreet time points over multiple individuals where continuous assessment may provide a clearer picture of the contribution of multilingualism to offsetting cognitive decline.

References

Albert, M. S., Jones, K., Savage, C. R., Berkman, L., Seeman, T., Blazer, D., & Rowe, J. W. (1995). Predictors of cognitive change in older persons: MacArthur studies of successful aging. *Psychology and Aging*, *10*(4), 578–579.

Alladi, S., Bak, T. H., Duggirala, V., Surampudi, B., Shailaja, M., Shukla, A. K., Kaul, S. (2013). Bilingualism delays age at onset of dementia, independent of education and immigration status. *Neurology*, *81*(22), 1938–44.

Bäckman, L., Lindenberger, U., Li, S.-C., & Nyberg, L. (2010). Linking cognitive aging to alterations in dopamine neurotransmitter functioning: recent data and future avenues. *Neuroscience and Biobehavioral Reviews*, *34*(5), 670–7.

Bak, T. H., Nissan, J. J., Allerhand, M. M., & Deary, I. J. (2014). Does bilingualism influence cognitive aging? *Annals of Neurology*, *75*(6), 959–63.

Barnes, D. E., Tager, I. B., Satariano, W. A., & Yaffe, K. (2004). The Relationship Between Literacy and Cognition in Well-Educated Elders. *The Journals of Gerontology Series A: Biological Sciences and Medical Sciences*, *59*(4), 390–395.

Bialystok, E. (1988). Levels of bilingualism and levels of linguistic awareness. *Developmental Psychology*, *24*, 560–567.

Bialystok, E. (2001). *Bilingualism in Development: Language, Literacy, and Cognition* (Vol. 8, p. 288). Cambridge University Press.

Bialystok, E. (2008). Bilingualism: The good, the bad, and the indifferent. *Bilingualism: Language and Cognition*, *12*(01), 3-11.

Bialystok, E., Craik, F. I. M., & Freedman, M. (2007). Bilingualism as a protection against the onset of symptoms of dementia. *Neuropsychologia*, *45*(2), 459–64.

Bialystok, E., Craik, F. I. M., Green, D. W., & Gollan, T. H. (2009). Bilingual Minds. *Psychological Science in the Public Interest* , *10* (3), 89–129.

Chertkow, H., Whitehead, V., Phillips, N., Wolfson, C., Atherton, J., & Bergman, H. (2010). Multilingualism (but not always bilingualism) delays the onset of Alzheimer disease: evidence from a bilingual community. *Alzheimer Disease and Associated Disorders*, *24*(2), 118–25.

Costa, A., Hernández, M., & Sebastián-Gallés, N. (2008). Bilingualism aids conflict resolution: evidence from the ANT task. *Cognition*, *106*(1), 59–86.

Galambos, S. J., & Goldin-Meadow, S. (1990). The effects of learning two languages on levels of metalinguistic awareness. *Cognition*, *34*(1), 1–56.

Gold, B. T. (2014). Lifelong bilingualism and neural reserve against Alzheimer's disease: A review of findings and potential mechanisms. *Behavioural Brain Research*, *281C*, 9–15.

Gollan, T. H., & Acenas, L. A. R. (2004). What is a TOT? Cognate and translation effects on tip-of-the-tongue states in Spanish-English and tagalog-English bilinguals. *Journal of Experimental Psychology. Learning, Memory, and Cognition*, *30*(1), 246–69.

Gollan, T. H., Montoya, R. I., Cera, C., & Sandoval, T. C. (2008). More use almost always a means a smaller frequency effect: Aging, bilingualism, and the weaker links hypothesis. *Journal of Memory and Language*, *58*(3), 787–814.

Gollan, T. H., Montoya, R. I., Fennema-Notestine, C., & Morris, S. K. (2005). Bilingualism affects picture naming but not picture classification. *Memory & Cognition*, *33*(7), 1220–1234.

Gollan, T. H., Montoya, R. I., & Werner, G. A. (2002). Semantic and letter fluency in Spanish-English bilinguals. *Neuropsychology*, *16*(4), 562–76.

Grady, C. L., Luk, G., Craik, F. I. M., & Bialystok, E. (2015). Brain network activity in monolingual and bilingual older adults. *Neuropsychologia*, *66*, 170–81.

Kavé, G., Eyal, N., Shorek, A., & Cohen-Mansfield, J. (2008). Multilingualism and cognitive state in the oldest old. *Psychology and Aging*, *23*(1), 70–78.

Li, S. C., Lindenberger, U., & Sikström, S. (2001). Aging cognition: from neuromodulation to representation. *Trends in Cognitive Sciences*, *5*(11), 479–486.

Luk, G., Bialystok, E., Craik, F. I. M., & Grady, C. L. (2011). Lifelong bilingualism maintains white matter integrity in older adults. *The Journal of Neuroscience : The Official Journal of the Society for Neuroscience*, *31*(46), 16808–16813.

Mahon, M., & Crutchley, A. (2006). Performance of typically-developing school-age children with English as an additional language on the British Picture Vocabulary Scales II. *Child Language Teaching and Therapy*, *22*(3), 333–351.

Mortimer, J. A. (1997). Brain reserve and the clinical expression of Alzheimer's disease. *Geriatrics*, *52 Suppl 2*, S50–3.

Peal, E., & Lambert, W. E. (1962). The Relation Of Bilingualism To Intelligence. *Psychological Monographs: General and Applied*, *76*(27), 1–23.

Portocarrero, J. S., Burright, R. G., & Donovick, P. J. (2007). Vocabulary and verbal fluency of bilingual and monolingual college students. *Archives of Clinical Neuropsychology : The Official Journal of the National Academy of Neuropsychologists*, *22*(3), 415–22.

Poulin-Dubois, D., Bialystok, E., Blaye, A., Polonia, A., & Yott, J. (2013). Lexical access and vocabulary development in very young bilinguals. *The International Journal of Bilingualism : Cross-Disciplinary, Cross-Linguistic Studies of Language Behavior*, *17*(1), 57–70.

Roberts, P. M., Garcia, L. J., Desrochers, A., & Hernandez, D. (2002). English performance of proficient bilingual adults on the Boston Naming Test. *Aphasiology*, *16*(4-6), 635–645.

Rogers, C. L., Lister, J. J., Febo, D. M., Besing, J. M., & Abrams, H. B. (2006). Effects of bilingualism, noise, and reverberation on speech perception by listeners with normal hearing. *Applied Psycholinguistics*, *27*(03), 465–485.

Rosselli, M., Ardila, A., Araujo, K., Weekes, V. A., Caracciolo, V., Padilla, M., & Ostrosky-Solís, F. (2000). Verbal fluency and repetition skills in healthy older Spanish-English bilinguals. *Applied Neuropsychology*, *7*(1), 17–24.

Scarmeas, N., Albert, S. M., Manly, J. J., & Stern, Y. (2006). Education and rates of cognitive decline in incident Alzheimer's disease. *Journal of Neurology, Neurosurgery & Psychiatry*, *77*(3), 308–316.

Servan-Schreiber, D., Printz, H., & Cohen, J. (1990). A network model of catecholamine effects: gain, signal-to-noise ratio, and behavior. *Science*, *249*(4971), 892–895.

Stern, Y. (2003). The concept of cognitive reserve: a catalyst for research. *J Clin.Exp.Neuropsychol.*, *25*(5), 589–593.

Stern, Y. (2009). Cognitive reserve. *Neuropsychologia*, *47*(10), 2015–2028.

Thomas, M. S. (1998, January). Distributed representations and the bilingual lexicon: One store or two?. In *4th Neural Computation and Psychology Workshop, London, 9–11 April 1997*.

Thomas, M. S. C., & Karmiloff-Smith, A. (2003). Modeling language acquisition in atypical phenotypes. *Psychological Review*, *110*(4), 647–82.

Valenzuela, M. J., & Sachdev, P. (2006). Brain reserve and dementia: a systematic review. *Psychological Medicine*, *36*(04), 441–454.