‘Hazy’ or ‘jumbled’? Putting together the pieces of the bilingual puzzle

Lorna García-Pentón, Yuriem Fernández García, Brendan Costello, Jon Andoni Duñabeitia & Manuel Carreiras

Abstract:

Six commentaries (Bialystok, 2015; de Bruin & Della Sala, 2015; Green & Abutalebi, 2015; Kroll & Chiarello, 2015; Luk & Pliatsikas, 2015; Paap, 2015) were produced in relation to the review “The neuroanatomy of bilingualism: How to turn a hazy view into the full picture” (García-Pentón et al., 2015). In the review, we argued that the available evidence for structural changes in bilingualism offers ambiguous support for current neural models of bilingualism and that this shortcoming in the field is exacerbated by critical methodological differences between studies. Thus, best practices need to be established for studying and modeling bilingualism. The commentaries bring to the discussion new perspectives and highlight additional challenges. Our response addresses the issues raised under two broad topics: the need to connect structural findings with behavioral and functional data, and a series of methodological concerns that are critical if the field is to advance.

Introduction

We do not currently have a clear picture of the neuroanatomical changes associated with bilingual and multilingual experience over the lifespan. Our review (García-Pentón et al., 2015; see also others, Baum & Titone, 2014; Costa, & Sebastián-Gallés, 2014; Li, Legault, & Litcofsky, 2014) calls attention to the fact that findings that report structural changes related to bilingualism are scarce and provide an ambiguous picture. The review stressed that neuroimaging studies of bilingualism need to take greater care in the data analysis, to use more sophisticated methods of analysis and to employ consistent (and transparent) protocols. Suggestions such as (i) combining behavioral and brain measures, (ii) more complex but standardized procedures, and (iii) larger and more thoroughly characterized samples were uncontroversial and were in general endorsed by the authors of the commentaries.

Beyond providing a summary of the results obtained thus far in studies of structural brain changes in bilingualism and the proposal for greater standardization in neuroimaging approaches, our aim was also to perform a critical review that clearly laid out where inconsistencies lie and the possible causes and consequences of these inconsistences. In this sense, our review tries to put a cautionary note on the exaggerated optimism when reporting/interpreting brain findings related to bilingualism. Thus, we recast the existing debate on the bilingual advantage hypothesis from the perspective of brain structure, and this is where most of the debate and discussion was generated in the commentaries.

Put succinctly, there are two clearly opposed sides concerning the bilingual advantage hypothesis. The proponents cite evidence that bilinguals outperform monolinguals on non-linguistic executive control tasks (see Bialystok et al., 2012; Kroll & Bialystok, 2013 for a review). Conversely, the detractors point out that, studies with

large samples and/or that take into account all components of the executive control mechanism and employ multiple tasks show no differences between bilinguals and monolinguals (Anton et al., 2014; Duñabeitia et al., 2014; see also Paap et al., 2015a for a review). These critics maintain that the bilingual advantage does not exist or, if it does, its size and frequency have been inflated, and the true phenomenon arises infrequently and in very specific and undetermined circumstances (Paap et al., 2015a).

In our review we did not claim that there is no bilingual advantage but we did point out that much of the current evidence is contradictory. We suggested that the neuroanatomical perspective could provide insight that might help to clear up some of these discrepancies, but we felt it crucial to highlight the need for more compelling studies since current findings do not provide conclusive evidence one way or the other. To be clear, we consider highly plausible the existence of a differential network configuration when facing the specific cognitive demands of bi- or multilingualism, or even differential neuroanatomical substrates related to accommodating more than one language (García-Pentón et al., 2014). The brain displays such a degree of plasticity that changes related to bilingualism must consistently occur. Nevertheless, interpreting any structural changes as supporting the hypothesis of the bilingual advantage (or any other hypothesis, for that matter) without correlating the findings with behavioral or functional changes reduces such an interpretation to mere speculation.

Although the aim of the original review was to focus on structural differences between bilinguals and monolinguals, here we will discuss the relationship between behavior, function and structure, as most commentaries have mentioned this matter. Given the importance of the connections between structure, function and behavior, there is a need to look squarely at the behavioral data in order to ascertain what sorts of effects of bilingualism have been established and may be reflected as neuroanatomical

correlates. This response is structured in two main topics: first, the relationship between behavior, function and structure, in the context of the claim for a bilingual advantage. In the second section, we reexamine various methodological concerns that were picked up in the commentaries, or which the commentaries themselves suggest need reiterating.

Making sense of structure: connecting with behavior and function

Needless to say, bilingualism is a complex issue and cannot be reduced to structural brain plasticity. We agree with the various commentators who point out that the debate must be extended to the relationship between behavioral, functional and structural data. Clearly, a functional or behavioral change does not necessarily imply a structural change (although it might be enough to produce one). Conversely, a structural change does not necessarily imply a change in behavior or function (but it may also be enough to produce one). The matter of misalignment problems when considering how neural and behavioral data fit together in bilingual studies has been already discussed elsewhere (see Duñabeitia & Carreiras, 2015; Paap et al., 2015a, 2015b). Duñabeitia & Carreiras (2015) make explicit the idea that there is no direct mapping between brain structure and cognitive function, and that we are far from understanding how increases or decreases in the density or volume of a particular region or network for a specific group of participants is linked to a putative cognitive function. An example of this disconnect can be seen in the white matter studies mentioned in our review article, which show either an increase or a decrease in fractional anisotropy (FA) for the same comparisons (bilinguals vs. monolinguals) in the same regions (Cummine & Boliek 2013; Gold et al., 2013; Luk et al., 2011; Mohades et al., 2012; Pliatsikas et al., 2015).

Various commentators appear to believe that we peddle a rather simplistic idea of the link between structure and function but at no point in our original review did we

suggest that this is a straightforward issue. On the contrary, our intention was to point out that extreme caution is required when associating a function with a given structure. In this regard, some brain differences that have been linked to the bilingual advantage are sometimes accompanied by disadvantages or null differences in behavior or could be interpreted as a bilingual disadvantage (see Abutalebi & Green, 2007; Bialystok et al., 2012; García-Pentón et al., 2014). As such, it is not clear why some brain differences are labeled as a bilingual advantage. Whatever structural brain differences are found, they should not automatically be taken to support this hypothesis.

To take an example of skill-related neuroanatomical changes from another field, differences in cortical thickness of the right parahippocampus of professional diving athletes as compared to a control group (non-athletes) showed a positive correlation with years of training experience in the professional athletes (Wei et al., 2011). However, the absence of behavioral measures makes it impossible to ascertain whether the effect is related to motor expertise, general training experience, or even previous brain differences that biased some individuals to become professional athletes, and so on.

Connecting the structural data with findings from functional studies means making sure that the functional data are tapping into the right contrasts and uncovering what is specific to bilingualism. Given that monolinguals do not have to switch between languages as bilinguals (or multilinguals) do, they provide the contrast needed to answer the question of whether dealing with more than one language imposes a difference on the population. At the same time, since monolinguals use just one language, in order to get at the differences between bilinguals and monolinguals, the measures of comparison have to be non-linguistic executive control tasks or an identical linguistic executive control task (if that were possible). This is normally done in behavioral studies, but

some MRI studies (e.g. Abutalebi et al. 2013) comparing multilinguals and monolinguals have made use of discrepant switching tasks (switching between languages for the multilinguals regardless of the grammatical category and between grammatical categories but within language for the monolinguals).

Even with behavioral evidence showing that some kinds of bilinguals have better performance on certain executive control tasks than monolinguals, differences between both groups in brain activity for executive control functioning are still less fully understood (see Grady et al., 2015). The presence of both null and significant differences between bilinguals and monolinguals at the behavioral, functional and structural levels, and the difficulty in interconnecting these different types of data represent a major challenge to reaching a unified theoretical account.

Instead of only looking at structural brain changes or changes in behavior separately, we must find the correspondence between the physical brain and mental brain. Otherwise we can end up in a logical fallacy: we interpret any difference in brain structure to argue for a theoretical account (i.e., a bilingual advantage) while at the same time we take a theoretical account (bilingual advantage) to explain a difference in the brain structure. Not only is behavior needed to support the bilingual advantage, but also clear predictions for differences in specific brain regions are needed.

To this end, studies should provide a direct demonstration of a correlation between anatomical, functional and/or behavioral data. Otherwise we are just moving the same debate from the realm of cognition to the brain (Duñabeitia & Carreiras, 2015). The lack of a correlation between the differences in the brain measures and behavior is very suggestive in the context of the current disjunctive (we return to the value of negative evidence in the following section). In sum, the presence of brain

plasticity by itself (without behavioral differences) does not constitute evidence of a bilingual advantage (see also Paap et al., 2015b).

The bilingual puzzle: why some pieces don't fit together

In this section we address various methodological considerations that deserve mention in the light of the commentaries. A position shared by some of the commentators (Bialystok, 2015; Green & Abutalebi, 2015; Kroll & Chiarello, 2015; Luk & Pliatsikas, 2015) holds that instead of a hazy view what the growing body of research is showing us are isolated pieces of the very complex puzzle of bilingual neuroplasticity. Admittedly, this is possible. However, another possibility is that the current panorama is skewed by methodological issues, and factors not contemplated by the hypothesis may be driving these findings. Parallels can be drawn with the debate in the field at the behavioral level, where it is now recognized that methodological mistakes (such as small sample sizes, ignoring null results or failing to control for nuisance covariates) have been made that have unnecessarily complicated the field of bilingualism and have led to a lack of replicability (Bialystok et al., 2015; de Bruin, Trecanni, & Della Sala, 2015a, 2015b; Paap et al., 2015). Therefore, studies should try to replicate previous work using the same methodology and subsequently use more novel and sophisticated (but standardized) methods to complement or refine their analysis. It is essential that there be at least some standards if the puzzle of the effects of bilingualism is to be solved. This section focuses on three main issues concerning methodology: the sample type and size, the role of negative evidence, and questionable analysis practices.

There is no doubt that bilingualism is a multifaceted expertise and as such it will not produce a simple effect or modify a single region. As a consequence, understanding

such a complex phenomenon will require more convincing results. Everyone agrees that the structural data, especially combined with behavioral data, can provide additional evidence and contribute to understanding the causal basis of the adaptive changes in bilingualism. We are a long way from being able to make meaningful generalizations and this will require running numerous transverse and longitudinal studies with well-characterized samples using standardized and sophisticated procedures. Equally, to avoid coming to loggerheads over apparently contradictory findings, studies should be conducted in different places to test hypotheses in different bilingual populations. This would also help to prevent having the same labs produce the same results. Studies carried out in different places usually have access to populations that vary significantly in linguistic profile, making findings more difficult to reproduce across labs but also more generalizable when replication does occur.

We also concur with the need for higher sample sizes for the studies and for samples to be matched in essential variables. At the same time, we agree that samples that are small but form part of well-controlled cross-sectional studies – particularly with special populations – with carefully characterized language profiles could offer good statistical power. Large sample sizes are desirable in neuroimaging studies, although this goal may be very difficult to achieve with special populations (such as elderly and children) as participants may be difficult to find due to compatibility with the technique. In this context, the use of standard methods and statistical procedures to report significant results, as well as more sophisticated analysis, becomes even more important.

To illustrate the need for well-characterized samples and standardized procedures, we return to the case of the white matter studies (Cummine & Boliek 2013; Gold et al., 2013; Luk et al., 2011; Mohades et al., 2012; Pliatsikas et al., 2015). Contrary to what

Luk and Pliatsikas claim in their commentary, we stated in our review that chronological age and age of acquisition of the second language are clearly confounding factors between these studies. Taken together, the results are difficult to interpret as the effect may be an increase in young and older adult bilinguals or a decrease in children and older adult bilinguals. In the review we argued that it may be precisely the combination of maturation/degeneration and second language acquisition effects that bring about these disparate patterns, especially as the regions identified by these studies (i.e. the corpus callosum (CC) and inferior frontal-occipital fasciculus (IFOF)) seem to be more sensitive to age effects (Good et al., 2001; Salat et al., 2005). Additionally, there are other confounding factors related to the variability of the participants' linguistic profiles and the failure of many studies to provide the necessary information about this, issues that were widely discussed in the review. The need for more transparent and standardized reporting practices applies to the neuroimaging data themselves: it was difficult to perform a fair comparison between these studies because in many cases we could not obtain the full local maximum coordinates of the cluster of the effect. As a result, we could only perform a general visual inspection of where these effects fell along the IFOF and CC.

Related to the issue of characterizing the samples adequately, it is important to bear in mind fundamental differences that may exist between different types of bilingual populations. Although acquiring a new skill could alter the brain and behavior at any age, results at the neural and behavioral level seem to indicate that when the skill is acquired from birth, such as two simultaneous languages, it may be harder to find modifications. In this sense, native bilingualism is not exactly a case of typical expertise, such as driving a taxi, video gaming, doing sports, or playing musical instruments, and so the impact of the lifelong bilingual experience may be substantially

different to what has been found for other types of expertise acquired later in life (Duñabeitia & Carreiras, 2015).

Turning to the issue of what null effects can add to the debate, it is critical that negative results be taken into consideration and not be dismissed. By "negative result" we refer to instances where the difference between bilinguals and monolinguals is in the opposite direction to that expected (i.e. the mean measure of monolinguals is greater than the bilingual measure), or both groups perform equally and therefore no mean difference appears (see behavioral studies comparing huge samples of participants, Antón et al., 2014; Duñabeitia et al., 2014). These two kinds of results fail to provide expected results in line with the hypothesis of the bilingual advantage, and in that sense they are "negative" results for the proponents of such an advantage.

In addition to being problematic for the bilingual advantage hypothesis, to what extent do these results provide support for an alternative account? Within the framework of classical inference it is not possible to accept the null hypothesis (no difference between groups). We can only either reject the null hypothesis (by finding a difference between groups) or fail to reject the null hypothesis (by finding no difference). This means that we cannot really affirm that there is no difference between groups, because we are unable to distinguish whether the null hypothesis is true or whether we do not have enough sensitivity and/or power in our experiment to detect differences. In other words, the absence of evidence is not equivalent to evidence of absence. However, a power analysis showing a huge amount of power such that even small differences should be detected makes it possible to reliably accept the null hypothesis (that there is no difference). In addition, Bayesian inference is a complementary approach to classical inference that overcomes some of the latter's limitations (see Friston and Penny, 2003 for more details). Specifically, using a Bayesian approach we can identify the

probability of finding an effect of a determined size and show evidence for "accept" the null hypothesis, showing a low probability of an effect of specific size. Thus, power analyses and Bayesian statistics make it possible to use null results as supporting evidence (for the null hypothesis), and "negative" results need to be taken into account.

Consequently, a theoretical model of bilingualism must account for the fact that, firstly, there are bilinguals who do not show an advantage over monolinguals in several control mechanisms (Anton et al., 2014; Duñabeitia et al., 2014; Paap et al., 2015b), and, secondly, there are bilinguals who do not show differences in brain structure as compared to monolinguals (see García-Pentón et al., 2015 for review). Contrary to what some commentators claim, we do not simply focus on those studies that obtained no differences. In fact, we hold that there is a trend in the opposite direction in the field: currently, those studies not supporting the hypothesis for the bilingual advantage are overlooked and ignored. We would do well to remember that the scientific method advances through the falsifiability of theories: any result that does not conform to a given theory's predictions brings that theory into doubt and raises the need for a redefinition and generation of new questions and hypotheses (de Bruin et al., 2015a; David et al., 2013; Jennings & Van Horn, 2012). Our intention is to avoid the risk of the neuroanatomical debate falling into the same stalemate as the discussion at the behavioral level.

Finally, we wish to highlight issues of data analysis that require attention. From the 21 articles we included in the initial review, with the variability of samples, linguistic profiles and methods, finding a pattern emerging from the results would have represented an important and robust effect. Unfortunately, this was not the case. Many of these studies looked for differences using a ROI-based approach; some of them did this after reporting no significant differences at whole brain level but most went straight

to look for differences with ROIs without providing information about the whole brain level. The fact that one particular region is affected or modulated by one condition in a ROI analysis does not mean that it is more relevant with respect to other brain regions. One region could be necessary but not sufficient for a process. This is why it is essential to look first at the whole brain level and then go further in a ROI based analysis to better understand the role of the region.

The methodological and analytic issues that abound in this field can be illustrated by means of an example mentioned in one of the commentaries. Green and Abutalebi point out that converging results obtained from different neuroimaging techniques and approaches (i.e. fMRI, PET, VBM and lesion studies) would provide compelling evidence of plasticity changes linked to the acquisition of more than one language. They identify the left caudate as a case in point. However, this finding cannot be taken as representative since it is an exception amongst otherwise heterogeneous findings. Furthermore, when findings are compared across labs rather than across publications, the level of convergence is greatly reduced. Even more importantly, almost all the studies cited to support this consistent finding suffer from some of the problems that we highlight in our target article: 1) the use of small sample sizes of between 6 and 14 participants (or individual case studies)[1]; 2) the reporting of uncorrected p values[2]; 3) the use of small volume correction (SVC) or region of interest (ROI) approach[3]. As pointed out in the original review, each of these practices is problematic. The use of a small sample size reduces the power of the study and opens the door to spurious false

[1] The sample size for each study is as follows: Price et al. (1999) 6 bilinguals; Abutalebi et al. (2008) 12 bilinguals; Abutalebi et al. (2007) 12 bilinguals; Lehtonen et al. (2005) 11 bilinguals; Crinion et al. (2006) 3 different groups of 11/14/10 bilinguals; Zou et al. (2012) groups of 14/13 bilinguals and monolinguals, respectively; Abutalebi et al. (2013) groups of 14/14 bilinguals and monolinguals; Abutalebi et al. (2000; 2009) are single case studies.

[2] The following studies report uncorrected p values: Price et al. (1999); Crinion et al. (2006); Abutalebi et al. (2007); Abutalebi et al. (2008).

[3] The following studies used these techniques: Crinion et al. (2006); Zou et al. (2012); Abutalebi et al. (2013).

positives. Additionally, the use of uncorrected p values must be avoided to reach inferences about a given region in neuroimaging studies because it is crucial to correct for the multiple comparisons problem (Bennett et al., 2009). Finally, restricting analysis to a small brain volume or ROI is different from the whole brain approach and caution is needed when interpreting such findings in terms of the relevance of one specific region in relation to other regions in the brain. Furthermore, only two of these studies actually compared bilinguals vs. monolinguals (Zou et al., 2012; Abutalebi et al., 2013).

In sum, although some commentators have suggested that the diverse findings represent different pieces of a complex puzzle that need to be fitted together, we feel the need to point out that some of the pieces of this puzzle have been badly made due to the use of questionable analysis techniques, some have been taken from different boxes due to sampling issues and some pieces are simply being ignored. This puts us in a poor state of affairs for solving the puzzle, but being aware of these problems will put us on the right road. In the final section, we suggest what still needs to be done in the light of these shortcomings.

Concluding remarks

If we want to solve the puzzle of the neuroanatomical effects of bilingualism, what needs to be done? Firstly we need to make sure the pieces are not incomensurable: studies need to use samples of participants that are sufficiently large to overcome the problems of statistical power, that are carefully matched to control for demographic factors and that have well-described linguistic profiles. These are key factors to improving the field.

Secondly, we need to fit all the pieces together: studies showing negative/null results should have the same opportunities for publication as studies showing positive

results. This would make it possible to understand when, how and why the bilingual advantage appears. Thirdly, we need to make sure the pieces are well made so that they can fit together: data analysis procedures need to comply to basic standards and there must be transparency in the reporting of data so that we know what the findings represent. Additionally, studies of the whole brain network that exploit more involved methods and techniques would be a welcome development.

In addition to these methodological concerns, we have also stressed the need for more exhaustive studies that integrate functional, behavioral and structural data to get a full picture of bilingualism. Since there are so many factors that could affect the brain, and since their consequences could be functional and/or structural, the impossibility of finding differences between groups or the absence of common patterns across studies does not mean that there are no differences or commonalities in many other directions.

We are charged with failing to come up with an alternative account for the variability in the findings for neuroanatomical differences related to bilingualism. However, our point is that the methodological shortcomings and the variability in the sample profiles are acute enough to cause an unacceptable amount of noise in the data. When activation was found in the brain of a dead salmon (Bennett et al., 2009), no-one suggested adjusting the theory to account for this anomalous data point. It was taken as a call to face up to (and do something about) statistical and methodological errors. Far be it from us to "oversimplify" the matter, but we believe that these issues require attention if the field is to advance. The current alternative appears to be to "overcomplicate" the matter by constructing a model that accommodates all these questionable findings. Let us not forget that we can't make a silk purse from sows' ears. There is no denying that data only make sense in the context of a strong theoretical framework. However, there are minimum requirements on the data themselves and it

will not do to have theories driven by specious or dubious findings. We need more research that is well-designed, theoretically motivated and correctly executed and analyzed.

Acknowledgments: This article was partially supported by the Severo Ochoa program grant SEV-2015-0490, and grants PSI2012-32123 and PSI2012-31448 from the MINECO and by ERC-2011-ADG-295362 grant from the European Research Council. The authors declare no competing interests.

References

Abutalebi, J., Annoni, JM., Zimine, I., Pegna, A.J., Seghier, M.L., Lee-Jahnke, H., et al. (2008) Language Control and Lexical Competition in Bilinguals: An Event-Related fMRI Study. *Cerebral Cortex*, 18, 1496-1505

Abutalebi, J., Brambati, S.M., Annoni, JM., Moro, A., Cappa, S.F. & Perani, D. (2007). The Neural Cost of the Auditory Perception of Language Switches: An Event-Related Functional Magnetic Resonance Imaging Study in Bilinguals. *The Journal of Neuroscience*, 27(50), 13762-13769

Abutalebi, J., Della Rosa, P.A., Ding, G., Weekes, B., Costa, A., and Green, D. W (2013). Language proficiency modulates the engagement of cognitive control areas in multilinguals. *Cortex*, 49, 905-911

Abutalebi, J., Della Rosa, P.A., Tettamanti, M., Green, D.W., Cappa, S.F. (2009). Bilingual aphasia and language control: a follow-up fMRI and intrinsic connectivity study. *Brain and Language*, 109, 141-56

Abutalebi, J., & Green, D. (2007). Bilingual language production: The neurocognition of language representation and control. *Journal of Neurolinguistics, 20*(3), 242-275

Abutalebi, J., Miozzo, A., & Cappa, S.F. (2000). Do subcortical structures control 'language selection' in polyglots? Evidence from pathological language mixing. *Neurocase*, 6, 51–56.

Antón, E., Duñabeitia, J. A., Estévez, A., Hernández, J. A., Castillo, A., Fuentes, L. J., Davidson, D.J., & Carreiras, M. (2014). Is there a bilingual advantage in the ANT task? Evidence from children. *Frontiers in Psychology, 5*, 398. doi: 10.3389/fpsyg.2014.00398

Baum, S., Titone, D. (2014). Moving toward a neuroplasticity view of bilingualism, executive control and aging. *Applied Psycholinguistics*, 35, 857-894.

Bennett, C.M., Miller, M.B., & Wolford, G.L. (2009). Neural correlates of interspecies perspective taking in the post-mortem Atlantic Salmon: An argument for multiple comparisons correction. *Neuroimage*, *47*(Suppl 1), S125.

Bialystok, E. (2015). How hazy views become full pictures. *Language, Cognition and Neuroscience*. Doi: 10.1080/23273798.2015.1074255

Bialystok, E., & Barac, R. (2012). Emerging bilingualism: dissociating advantages for metalinguistic awareness and executive control. *Cognition, 122*(1), 67-73. doi: 10.1016/j.cognition.2011.08.003

Bialystok, E., Kroll, J.F., Green, D.W., MacWhinney, B., Craik, F.I. (2015). Publication Bias and the Validity of Evidence: What's the Connection? *Psychol Sci.* 26(6), 944-6. Doi: 10.1177/0956797615573759

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

Crinion, J., Turner, R., Grogan, A., Hanakawa, T., Noppeney, U., Devlin, J.T., Aso, T., Urayama, S., Fukuyama, H., Stockton, K., Usui, K., Green, D.W., Price, C.J. (2006) Language control in the bilingual brain. *Science*, 312, 1537-1540

Cummine, J., & Boliek, C. A. (2013). Understanding white matter integrity stability for bilinguals on language status and reading performance. *Brain Structure & Function, 218*(2), 595-601. doi: 10.1007/s00429-012-0466-6

David, S.P., Ware, J.J., Chu, I.M., Loftus, P.D., Fusar-Poli, P., et al. (2013) Potential Reporting Bias in fMRI Studies of the Brain. PLoS ONE 8(7): e70104. doi:10.1371/ journal.pone.0070104

de Bruin, A. & Della Sala, S. (2015) The importance of language use when studying the neuroanatomical basis of bilingualism. *Language, Cognition and Neuroscience*. Doi: 10.1080/23273798.2015.1082608

de Bruin, A., Treccani, B., & Della Sala, S. (2015a). Cognitive advantage in bilingualism an example of publication bias? *Psychol. Sci.*, *26*(1), 99-107

de Bruin, A., Treccani, B., & Della Sala, S. (2015b). The connection is in the data: we should consider them all. *Psychol. Sci.*, *26*(1), 99-107. doi:10.1177/0956797615583443

Duñabeitia, J.A. & Carreiras, M. (2015). The bilingual advantage: acta est fabula?. *Cortex*. doi: 10.1016/j.cortex.2015.06.009.

Duñabeitia, J. A., Hernández, J. A., Antón, E., Macizo, P., Estévez, A., Fuentes, L. J., & Carreiras, M. (2014). The inhibitory advantage in bilingual children revisited: Myth or reality? *Experimental Psychology, 61*(3), 234-251

Friston, K.J. & Penny, W. (2003). Posterior probability maps and SPMs. *NeuroImage*, 19, 1240-1249.

García-Pentón, L., Fernández, Y., Costello, B., Duñabeitia, J.A. & Carreiras, M. (2015). The neuroanatomy of bilingualism: how to turn a hazy view into the full picture. *Language, Cognition and Neuroscience*. Doi:10.1080/23273798.2015.1068944

García-Pentón, L., Pérez Fernández, A., Iturria-Medina, Y., Gillon-Dowens, M., & Carreiras, M. (2014). Anatomical connectivity changes in the bilingual brain. *Neuroimage, 84*, 495-504. doi: 10.1016/j.neuroimage.2013.08.064

Gold, B. T., Johnson, N. F., & Powell, D. K. (2013). Lifelong bilingualism contributes to cognitive reserve against white matter integrity declines in aging. *Neuropsychologia, 51*(13), 2841-2846. doi: 10.1016/j.neuropsychologia.2013.09.037

Good, C. D., Johnsrude, I. S., Ashburner, J., Henson, R. N., Friston, K. J., & Frackowiak, R. S. (2001). A voxel-based morphometric study of ageing in 465 normal adult human brains. *Neuroimage, 14*(1 Pt 1), 21-36. doi: 10.1006/nimg.2001.0786

Grady, C.L., Luk, G., Craik, F.I.M., Bialystok, E. (2015). Brain network activity in monolingual and bilingual older adults. *Neuropsychologia*, 66, 170-181.

Green, D.W. & Abutalebi, J. (2015). Language control and the neuroanatomy of bilingualism: in praise of variety. *Language, Cognition and Neuroscience*. Doi:0.1080/23273798.2015.1084428

Jennings, R.G. & Van Horn, J.D. (2012). Publication bias in neuroimaging research: Implications for meta-analyses. *Neuroinformatics*, 10(1), 67-80 doi:10.1007/s12021-011-9125-y.

Kroll, J. F., & Bialystok, E. (2013). Understanding the consequences of bilingualism for language processing and cognition. *Journal of Cognitive Psychology*, *25*(5), 497-514. doi: 10.1080/20445911.2013.799170

Kroll, J. and Chiarello, C. (2015). Language experience and the brain: variability, neuroplasticity, and bilingualism. *Language, Cognition and Neuroscience*. Doi:10.1080/23273798.2015.1086009

Lehtonen, M., Laine, M., Niemi, J., Thomson T., Vorobyev, V.A., and Hughdal, K. (2005). Brain correlates of sentence translation in Finnish-Norwegian bilinguals. *NeuroReport*, 16, 607-610

Li, P., Legault, J., & Litcofsky, K. A. (2014). Neuroplasticity as a function of second language learning: anatomical changes in the human brain. *Cortex, 58*, 301-324. doi: 10.1016/j.cortex.2014.05.001

Luk, G., Bialystok, E., Craik, F. I., & Grady, C. L. (2011). Lifelong bilingualism maintains white matter integrity in older adults. *The Journal of Neuroscience, 31*(46), 16808-16813. doi: 10.1523/jneurosci.4563-11.2011

Luk , G. & Pliatsikas, C. (in press). Converging diversity to unity: Commentary on The neuroanatomy of bilingualism. . *Language, Cognition and Neuroscience*.

Mohades, S. G., Struys, E., Van Schuerbeek, P., Mondt, K., Van De Craen, P., & Luypaert, R. (2012). DTI reveals structural differences in white matter tracts between bilingual and monolingual children. *Brain Research, 1435*, 72-80. doi: 10.1016/j.brainres.2011.12.005

Paap, K. (2015). The neuroanatomy of bilingualism: will winds of change lift the fog? *Language, Cognition and Neuroscience*. Doi: 10.1080/23273798.2015.1082607

Paap, K.R., Johnson, H.A. and Sawi, O. (2015a). Bilingual advantages in executive functioning either do not exist or are restricted to very specific and undetermined circumstances. *Cortex*, 69, 265-278

Paap, K.R., Johnson, H.A. and Sawi, O. (2015b). Should the search for bilingual advantages in executive functioning continue? *Cortex*. Doi:10.1016/j.cortex.2015.09.010

Pliatsikas, C., Moschopoulou, E., & Saddy, J. D. (2015). The effects of bilingualism on the white matter structure of the brain. *Proceedings of the National Academy of Sciences*, 112 (5), 1334–1337. doi:10.1073/pnas.1414183112

Price, C.J., Green, D.W. and von Studnitz, R. (1999). A functional imaging study of translation and language switching. *Brain*, 122, 2221-2235

Salat, D. H., Tuch, D. S., Greve, D. N., van der Kouwe, A. J., Hevelone, N. D., Zaleta, A. K., . . . Dale, A. M. (2005). Age-related alterations in white matter microstructure measured by diffusion tensor imaging. *Neurobiology of Aging, 26*(8), 1215-1227. doi: 10.1016/j.neurobiolaging.2004.09.017

Wei, G., Zhang, Y., Jiang, T., Luo, J. (2011). Increased Cortical Thickness in Sports Experts: A Comparison of Diving Players with the Controls. PLoS ONE 6(2): e17112. doi:10.1371/journal.pone.0017112

Zou, L., Ding, G., Abutalebi, J., Shu, H., & Peng, D. (2012). Structural plasticity of the left caudate in bimodal bilinguals. *Cortex, 48*(9), 1197-1206. doi: 10.1016/j.cortex.2011.05.022