Linguistic immersion and structural effects on the bilingual brain: a longitudinal study*

VINCENT DELUCA
School of Psychology and Clinical Language Sciences, University of Reading, UK
JASON ROTHMAN
School of Psychology and Clinical Language Sciences, University of Reading, UK
UiT the Arctic University of Norway
CHRISTOS PLIATSIKAS
School of Psychology and Clinical Language Sciences, University of Reading, UK

Learning and using additional languages can result in structural changes in the brain. However, the time course of these changes, as well as the factors the predict them, are still not well understood. In this longitudinal study we test the effects of bilingual immersion on brain structure of adult sequential bilinguals not undergoing any language training, who were scanned twice, three years apart. We observed significant increases in grey matter volume in the lower left cerebellum, mean white matter diffusivity in the frontal cortex, and reshaping of the left caudate nucleus and amygdala and bilateral hippocampus. Moreover, both prior length of immersion and L2 age of acquisition were significant predictors of volumetric change in the cerebellum. Taken together, these results indicate that bilingualism-induced neurological changes continue to take place across the lifespan and are strongly related to the quantity and quality of bilingual immersion, even in highly-immersed adult bilingual populations.

Introduction

Since the early 2000s, an increasing amount of evidence has amassed suggesting that bilingualism has an impact on brain structure (Bialystok, 2017; Pliatsikas, forthcoming, 2019), with a smaller amount of studies even suggesting an impact on the brain's default functionality (Pliatsikas & Luk, 2016). This impact has tended to be measured as a function of differences between bilingual and monolingual populations in cross-sectional designs. As might be expected given the nature of this type of research (see Bak, 2016 for discussion), results have been inconsistent across studies. Indeed, there is variability in terms of location of structural differences in the brain (e.g., hippocampus (Mårtensson, Eriksson, Bodammer, Lindgren, Johansson, Nyberg & Lövdén, 2012), versus supramarginal gyrus (SMG) (Mechelli, Crinion, Noppeney, O'Doherty, Ashburner, Frackowiak & Price, 2004), versus the left inferior frontal gyrus (Mohades, Struys, Van Schuerbeek, Mondt, Van De Craen & Luypaert, 2012)), and directionality of changes (e.g., increased versus decreased grey matter volume (Klein, Mok, Chen & Watkins, 2014) and/or white matter integrity (Gold, Johnson & Powell, 2013)). García-Pentón et al. (García-Pentón, Fernández García, Costello, Duñabeitia & Carreiras, 2016) offer a convincing view of why such inconsistencies likely occur. Even allowing for the fact that the application of neuroimaging methodologies within bilingualism studies is relatively recent, and, therefore, not abundant, García-Pentón et al. argue that non-trivial issues emerge from the lack of standardized protocols across labs. For example, they highlight that higher sample sizes, a consensus of standard scanning procedures across labs and establishing more universal, and better vetting/selection criteria within subject populations are key factors contributing to the dearth of consistency. While we do not disagree with such suggestions per se, it is not clear that dealing with them alone would have the overall effect of teasing out the signal we aim to capture from the noise that muddies it (Bialystok, 2016). Others have argued that another major contribution to the inconsistency of findings in the literature likely stems from treating bilingualism as a monolithic variable. Acknowledging that bilingualism itself is a fluid, complex and dynamic experience is a necessary first step, but acknowledging this alone is not sufficient to deal its contribution to the constituency problem. The nature of the bilingual

Address for correspondence: Christos Pliatsikas, School of Psychology and Clinical Language Sciences, University of Reading, RG6 6AL, UK
c.pliatsikas@reading.ac.uk

experience, comprising a dynamic continuum of co-varying factors, must be dealt with methodologically (Bialystok, 2017; Kroll & Chiarello, 2016; Luk & Bialystok, 2013). After all, if it is the experience of bilingualism that gives rise to adaptive change in the brain then it logically follows that various permutations of the bilingual experience should have measurably different outputs.

Although the link between structural changes in the brain and bilingualism is not fully understood, the pattern that emerges is becoming increasingly clear: bilingualism has some type of an effect on the structure of the brain. It seems reasonable that some aspects of the individual experience of being bilingual – factors that vary across individuals and/or whole subgroups – contribute to the ultimate explanation of the variance noted. Under such a view, we should assess specific factors within the bilingual experience – primarily the ones that differentiate types of bilinguals such as age of acquisition (AoA), immersion in the language(s), patterns of using the languages, level of code-switching, relative proficiency in both languages, the social milieu, etc. – with respect to potential effects on the brain. To date, exceedingly few studies have attempted to correlate experience-based factors to structural changes in the adult bilingual brain (e.g., Pliatsikas, DeLuca, Moschopoulou & Saddy, 2017) and none have done so with a longitudinal design. Such a design will help determine whether these factors, and if so which of them, are likely to explain the observed variability among various bilingual subgroups. Doing so can: (i) reveal that the present literature is less messy than a current snapshot might suggest, (ii) uncover the mechanisms by which changes occur because of bilingualism, (iii) render more precise predictions for where in the brain we might find effects, if at all, and (iv) specify for which bilingual individuals/groups this occurs. In sum, we submit that the ultimate explanation for why replication is an issue in the field may be due to the fact that bilingual samples/cohorts are not adequately measured in terms of relevant experience-based factors that would provide the link between bilingualism and its neurological effects.

Given the above discussion, we used various experience-based variables as predictors of structural changes in the brain and we focus on one in particular – immersion in the L2 context – because it turns out to be (along with L2 AoA) predictive for individual differences in changes in the bilingual brain. Immersion is an excellent factor to begin the unpacking of the catch-all label "bilingual" as often used in the neuroimaging literature for several reasons. First, immersion is a valid proxy for many things such as access to high quantity of language input, the high quality of input (because a majority of it will come from native speakers), and crucially opportunity to use both languages. The proxy of immersion thus has a clear effect on the relative juxtaposition of activation of

both languages relative to the inhibitory control needed to keep cross-linguistic influence to a minimum, which has been argued to be a likely contributor to ensuing changes in both bilingual behavior and anatomical changes to the brain (Marian & Spivey, 2003; Spivey & Marian, 1999; Thierry & Wu, 2007). Moreover, the study presented in this paper has an additional value. To our knowledge it is the first study to tackle this issue in a longitudinal manner in adults that do not undergo any kind of L2 training[1]; that is, testing the same individuals more than once with no less than 3 years in between scans. Whereas other studies have used a cross-sectional methodology, we will be able to verify changes within the same bilingual brains over time as their period of active immersion in a naturalistic L2 native community increases.

Background literature

L2 AoA has often been used as a proxy for the total amount of time one is exposed to/uses the additional language, and has also been used to investigate optimal or critical periods in brain's plasticity with respect to L2 acquisition (see Berken, Gracco & Klein, 2017 for review). L2 AoA has been found to relate to increased cortical grey matter (GM) volume, e.g., in the left SMG in bilinguals (Mechelli et al., 2004), cortical thickness in the left and right inferior frontal gyrus (IFG) and the superior parietal lobe in bilinguals (Klein et al., 2014), and GM density in the left putamen, posterior insula, bilateral occipital cortex, and right dorsolateral prefrontal cortex, and bilateral premotor cortex (Berken, Gracco, Chen & Klein, 2016). Effects of L2 AoA have also been found to manifest as increased fractional anisotropy (FA), a common index for measuring white matter (WM) integrity, in various language related tracts including the left inferior fronto-occipital fasciculus (IFOF) (Mohades et al., 2012; Rossi, Cheng, Kroll, Diaz & Newman, 2017). More recently, AoA has also been found to affect resting state connectivity (Berken, Chai, Chen, Gracco & Klein, 2016; Kousaie, Chai, Sander & Klein, 2017) (Table 1).

Informative as it is, the use of L2 AoA as a predictor of brain changes might be insufficient, for two main reasons: First, cut-off points between early and late bilingualism, very common in earlier studies, are often defined on some arbitrary age criterion, with great variability among studies. Second, simply reporting AoA does not imply active and continuous L2 usage, which in turn might be crucial for any observed brain restructuring. Indeed, it is possible that brain reorganization in bilinguals is in part, if not fully, due to the continuous language switching demands for bilinguals, rather than the mere acquisition

---

[1] To our knowledge, there is only one other study that is longitudinal and not a training study, however, this was done with children and not adults. Mohades et al. (2015) is explained in greater detail below.

Table 1. *Summary of effects of L2 AoA on brain structure and resting-state connectivity.*

Study	Participants	Measures	Findings
(Mechelli et al., 2004)	25 monolinguals 25 'early' bilinguals (L2AoA ≤ 5 yrs) 33 'late' L2 AoA 10 and 15	Voxel-based Morphometry (VBM)	-Bilinguals greater GMV in left inferior parietal lobule (IPL) -GMV increase correlates with AoA and proficiency
(Klein, Mok, Chen, & Watkins, 2014)	12 simultaneous bilinguals; 25 early sequential bilinguals (mean AoA 5 yrs), 29 late sequential bilinguals (AoA 10 yrs) 22 monolinguals	Cortical Thickness (CT)	-LIFG (pars triangularis and orbitalis) higher CT for bilingual groups over monolingual. -Late sequential bilinguals, a reduction in CT in the RIFG compared to simultaneous and early sequential bilinguals. -Early sequential bilinguals reduced CT in this area compared to monolinguals. -AoA and CT were positively correlated in the LIFG and left superior parietal lobe; negatively correlated in the RIFG.
(Berken, Gracco, Chen, & Klein, 2016)	16 simultaneous bilinguals; 25 early sequential bilinguals (mean AoA 5 yrs),	VBM (GM density)	-simultaneous bilinguals > sequential bilinguals, in the left putamen, left posterior insular cortex, left mid-occipital gyrus, right DLPFC, & right lateral occipital cortex -sequential bilinguals > simultaneous bilinguals in the bilateral premotor cortex
(Mohades et al., 2012)	15 simultaneous bilinguals children, 15 sequential bilingual children, 10 monolingual children	TBSS/DTI	-monolingual group displayed the highest FA values for the corpus callosum to occipital lobe (AC-OL) -simultaneous bilingual group highest FA values in left IFOF.
(Rossi, Cheng, Kroll, Diaz, & Newman, 2017)	24 monolingual English speakers, and 25 native English speaking, late learners of Spanish (age range: 18–27)	TBSS/DTI	-L2 learners higher FA values in IFOF, ILF, ATR, UF -FA values correlate with AoA
(Berken, Chai, Chen, Gracco, & Klein, 2016)	16 simultaneous bilinguals 18 sequential bilinguals (L2 AoA > 5 years)	Resting-state connectivity	-Simultaneous > sequential bilinguals connectivity in cerebellum, bilateral IFG, IPL, DLPFC from seeds in left and right IFG
(Kousaie, Chai, Sander, & Klein, 2017)	11 simultaneous bilinguals 10 sequential bilinguals (avg. L2 AoA 7 years)	Resting-state connectivity	-vmPFC seed in default mode network (DMN), stronger anticorrelation in connectivity in DMN for simultaneous bilingual group -degree of anticorrelation relates to Simon task performance

of an L2. To this end, L2 immersion, or the amount of time spent in a naturalistic L2 environment, has more recently been examined as a potential key factor (Stein, Winkler, Kaiser & Dierks, 2014), taken here as a proxy for continuous and intensive exposure to and use of the L2 (Pliatsikas & Chondrogianni, 2015). Comparing highly immersed to non-immersed sequential bilinguals of comparable L2 proficiency, and to monolinguals, Pliatsikas et al. (2017) reported expansions in the globus pallidus, putamen, and thalamus for the highly immersed bilinguals, but only some limited restructuring in the caudate nucleus for the bilingual group with limited

immersion. L2 immersion has also been found to affect WM integrity in language-related pathways. Comparing late-acquired, immersed, L2 learners of English to a monolingual control group, Pliatsikas, Moschopoulou & Saddy (2015) reported increased integrity in several WM tracts including bilaterally the corpus callosum, IFOF, uncinate fasciculus (UF), and superior longitudinal fasciculus (SLF). Crucially this group exhibited patterns of WM increase in a similar manner to elderly lifelong bilinguals (Luk, Bialystok, Craik & Grady, 2011; Olsen, Pangelinan, Bogulski, Chakravarty, Luk, Grady & Bialystok, 2015), also pointing to effects of immersion that are independent of the L2 AoA. Kuhl and colleagues (Kuhl, Stevenson, Corrigan, van den Bosch, Can & Richards, 2016) have also reported WM adaptations in response to L2 immersion: specifically mean diffusivity (MD) in anterior tracts of the left hemisphere were modulated by increased L2 exposure (listening), whereas production (speaking) was found to modulate FA values in the posterior section of the left hemisphere. Finally, volumetric increases in the cerebellum in bilinguals have been linked with increased exposure/proficiency in the non-native language, and were correlated with more efficient grammatical processing, which was also native-like in terms of functional activation of the cerebellum (Pliatsikas, Johnstone & Marinis, 2014a, 2014b).

Although technically not a factor that DESCRIBES the bilingual experience, but rather is a consequence thereof, L2 proficiency has also been examined as a potential predictor of neuroplasticity in bilinguals (Abutalebi, Canini, Della Rosa, Green & Weekes, 2015; Nichols & Joanisse, 2016). Increased L2 proficiency has been found to relate to GM volume increases in regions such as the anterior cingulate cortex (ACC) in ageing bilingual populations (Abutalebi et al., 2015), and WM and functional adaptations in a variety of regions in bilingual adults (Nichols & Joanisse, 2016). Moreover, GM density in the cerebellum has been found to relate to levels of semantic and phonemic fluency in both languages for bilingual adults (Grogan, Green, Ali, Crinion & Price, 2009). It is worth noting here that all results corresponding to different measures of proficiency are inextricably related to the measure itself, in the absence of an objective, or at least commonly agreed, way to measure proficiency. It is therefore possible that these effects only reflect the acquisition of the particular skill that the chosen proficiency measure taps on or is more reflective of the fact that increased proficiency is likely to correlate with multiple sub-factors of L2 language use and exposure.

Two proposals have attempted to model the processes of neural adaptation: the Adaptive Control Hypothesis (ACH) (Abutalebi & Green, 2016; Green & Abutalebi, 2013), and the Bilingual Anterior to Posterior and Subcortical Shift (BAPSS) model (Grundy, Anderson & Bialystok, 2017). The ACH discusses brain adaptation

as a result of linguistic experience by proposing that the brain adapts to the conversational/linguistic contexts in which one is engaged routinely, including both single and dual language contexts (Green, 2011). These contexts call on varying subsets of cognitive control, including planning, inhibitory, and engagement processes (Green & Abutalebi, 2013). As modulated by changes to experience, a reshaping of cortical and subcortical structures occurs, including the left inferior frontal gyrus (LIFG) left caudate, putamen, and thalamus. The cerebellum is also implicated, forming part of the language control network with the LIFG, caudate, and putamen via the thalamus (Abutalebi & Green, 2016). The BAPSS model examines the overarching effect of bilingual/non-native language use through time. It states that initial stages of contact with an additional language incur reliance on frontal areas, due to increased demands on working memory and several language/executive control processes. As L2 exposure and proficiency increase, reliance on, and use of, the frontal regions shifts to subcortical and posterior regions. Naturalistic linguistic immersion is an ideal environment in which to test the models, given the opportunity of continuous and sustained exposure to the additional language (Grundy et al., 2017). However, both theories have mostly based their predictions on the synthesis of a huge variety of functional and structural brain data which have come either from cross-sectional studies comparing bilinguals to monolinguals, or from studies that report significant correlations between indices of brain structure and function and self-reported demographic variables. Longitudinal studies would be an ideal method to examine how specific experience-based factors of bilingualism manifest in the brain, and crucially how they change through time. Given the main comparison is within-subjects, changes to experience can be isolated and thus examined directly without necessarily collapsing other aspects of bilingualism across one another (Luk & Pliatsikas, 2016; Wong, Yin & Brien, 2016). A handful of studies in the past decade have examined neuroplastic effects of L2 acquisition and use from a longitudinal perspective. Notably, all the available longitudinal studies have focused on the brain outcomes of intensive training programs to participants that were newly acquiring an L2.

Several studies have found cortical GM volume to be affected during the acquisition of the L2 (Bellander, Berggren, Mårtensson, Brehmer, Wenger, Li, Bodammer, Shing, Werkle-Bergner & Lövdén, 2016; Mårtensson et al., 2012; Osterhout, Poliakov, Inoue, McLaughlin, Valentine, Pitkanen, Frenck-Mestre & Hirschensohn, 2008; Stein, Federspiel, Koenig, Wirth, Strik, Wiest, Brandeis & Dierks, 2012). For example Mårtensson et al. (2012) studied effects of initial stages L2 acquisition in military interpreters undergoing a 10-month intensive language training course. They report significant increases in GM volumes in the left superior temporal gyrus

(STG), left medial frontal gyrus (MFG), and right hippocampus (RHC) after three months, the extent of which correlated with proficiency (subjects' course performance). Differences in WM integrity have also been found as a result of L2 acquisition and/or use over a range of time periods, from hours (Hofstetter, Friedmann & Assaf, 2016) to 12–18 months (Hosoda, Tanaka, Nariai, Honda & Hanakawa, 2013). A recent study by Mamiya and colleagues examined the relationship of white matter tracts, immersion in an L2 environment, and genetic factors related to the growth of WM integrity (Mamiya, Richards, Coe, Eichler & Kuhl, 2016). Regarding effects of immersion, increases in FA values in the right SLF were positively correlated with both time spent in the language course and proficiency measures. These values were also found to decline after the course was completed.

While the results are encouraging, available studies only present a partial picture in terms of relating structural change to L2 acquisition and use, for a variety of reasons: first, they primarily examine early stages of L2 acquisition or use (Mamiya et al., 2016; Stein et al., 2012); second, they examine the effects of intensive linguistic training (Hofstetter et al., 2016; Hosoda et al., 2013; Mårtensson et al., 2012), thus report brain adaptations to experiences that stem from a highly controlled linguistic environment; third, and perhaps most important, the majority of these studies focused on one aspect of L2 acquisition, very often vocabulary acquisition, so they do not account for the full repertoire of both L2 acquisition and L2 control (lexicon, grammar, phonology, pragmatics) which immersion in an L2-speaking environment brings about. To date, only one study has longitudinally examined the effects of naturalistic language use on the brain. Mohades et al. (Mohades, Van Schuerbeek, Rosseel, Van De Craen, Luypaert & Baeken, 2015) compared simultaneous bilingual with early sequential bilingual (mean AoA: 3 years) and monolingual children at two points, (two years apart) across several language-related WM pathways. They report that sequential bilinguals had the greatest increases in mean FA values in the left IFOF after the two-year period, while the simultaneous bilinguals displayed the highest overall FA values in this region at the second scan. Other than that, no study has longitudinally examined the effect of naturalistic immersion on neural plasticity, especially in already-proficient bilingual adults over a long-term period.

The present study

The aim of this study is to examine the effects of naturalistic linguistic immersion on brain structure in adult bilinguals over an extended period. Highly proficient non-native speakers of English were scanned three years apart while residing in the UK. Importantly, these participants did not undergo any linguistic training in their L2 or any other language during their immersion period, but had been continuous residents in the L2-dominant environment. We specifically examined potential changes in language demographics, neural structure, and functional connectivity over the three-year period. We also assessed whether any changes in the brain were potentially predicted by differences in linguistic experience – specifically overall length of immersion, L2 proficiency, and L2 AoA.

Given results of previous studies showing effects of L2 immersion (e.g., Pliatsikas et al., 2017), we predict neuroplastic changes to occur, largely in either the cerebellum or subcortical regions, specifically the putamen, thalamus, and globus pallidus (Berken et al., 2016; Grogan et al., 2009; Pliatsikas et al., 2014b; Pliatsikas et al., 2017), as processing/production efficiency increases with prolonged, sustained, exposure to the non-native language (Grundy et al., 2017). We would also predict higher FA values in WM tracts connecting the subcortical structures with posterior regions including the cerebellum (Abutalebi & Green, 2016; Grundy et al., 2017). With respect to cortical GM, we do not predict any significant changes, given the majority of reported effects were seen in non-immersed bilingual subjects or those acquiring their L2 (e.g., Mårtensson et al., 2012; Mechelli et al., 2004; Osterhout et al., 2008). Finally, we wanted to test whether any observed changes in structural connectivity are related to changes in functional connectivity, as evidence suggests (Luk, Bialystok, et al., 2011), so we included a resting state fMRI scan.

Materials and methods

Participants

Twenty-six healthy bilingual participants that had already been scanned for previous projects were invited back three years later[2]. The inclusion criterion was continuous residence in the UK between the two testing points. Of these, nine participated in this study (8 females; Mage 35.33 years, SD 8.12) and the two scanning sessions were approximately three years apart (average time between Visits 1 and 2: 3.22 years; $SD = 0.12$). At Visit 1, the participants had a variety of L1s, were already residing

---

[2] A group of monolingual native English-speaking participants (n = 25, 14 female, Mage 28.16, SD 5.3) were also scanned initially as a control group and were also invited to this study. Of them, only six participants (4 female, Mage 33 yrs, SD 2.38) returned for Visit 2. The same analytical procedure was applied to this group as for the bilingual participants, and no significant neural changes across the longitudinal period were found for this group. Although this is not a surprise finding, and it would have indeed been our prediction, we do not discuss this cohort in detail primarily due to the small sample size, but also for several additional reasons, which are covered in the Discussion section.

Table 2. *Demographic information for subjects at Visit 1.*

Subject	Age	Sex	Education	Native Language	UK Residence (months)	QPT Score (%)	L2 AoA
1	34	M	Graduate School	Greek	99	98	8
2	50	F	College (BA/BS)	Chinese	374	92	7
3	37	F	Graduate School	Polish	114	70	4
4	34	F	Graduate School	French	144	93	16
5	40	F	Graduate School	Polish	92	73	19
6	31	F	Graduate School	Italian	4	62	8
7	26	F	Graduate School	Greek	5	79	9
8	49	F	Graduate School	German	175	100	11
9	44	F	College (BA/BS)	Spanish	196	100	4

in the UK (*M* length residence: 134 months, *SD* 111.7 months, range = 4 months to 196 months), and they reported an average AoA of L2 English of 10.5 years (*SD* 4.6) (see Table 2 for details).

Procedure, data acquisition, and analysis

In both Visits, prior to their scan participants were assessed for their English proficiency via the Oxford Quick Placement Test (QPT) (Geranpayeh, 2003). The same scanning protocol was implemented for both Visits 1 and 2. Brain images were collected with a 3.0-Tesla Siemens MAGNETOM Trio MRI scanner with Syngo software and 32-channel Head Matrix coil. T1-weighed MPRAGE (Magnetization Prepared Rapid Gradient Echo) images were collected from each participant (192 sagittal slices, 1 mm slice thickness, in-plane resolution 250 × 250, acquisition matrix of 246 × 256 mm, echo time (TE) = 3.02 ms, repetition time (TR) = 2020 ms, inversion time = 900 ms, flip angle = 9°). The scan lasted approximately 10 minutes. T2-weighted Diffusion-Tensor Imaging (DTI) images were also collected (60 transversal slices, 2 mm slice thickness, acquisition matrix 256 × 256, in-plane resolution 128 × 128, 2 averages, TE = 93 ms, TR = 8200 ms). The scan lasted approximately 11 minutes. A resting-state Echo Planar Imaging (EPI) sequence was also run for each participant (200 volumes, 56 transversal slices, 3 mm slice thickness, in-plane resolution 64x64, acquisition matrix of 192 × 192 mm, TE = 30 ms, TR = 3000 ms, flip angle = 90°). This scan lasted approximately 10 minutes. The MRI scanning session did not involve any tasks. However, subjects were advised to keep their eyes open during the resting-state scan.

T1-weighted images were pre-processed with the FSL_anat software pipeline in FSL (Jenkinson, Beckmann, Behrens, Woolrich & Smith, 2012). Images were reoriented to the Montreal Neurological Institute (MNI)-152 orientation, automatically cropped, bias-field corrected, and non-linearly-registered to MNI space. Grey matter volume was calculated via the voxel-based morphometry (VBM) pipeline in FSL (Ashburner & Friston, 2000). Pre-processed images were brain extracted and grey matter segmented. A study specific template was then created using the average of the GM images. Native GM images were registered to this template, and modulated to correct for local expansions and contractions due to the non-linear component of registration. They were spatially smoothed with an isotropic Gaussian Kernel of 3mm. A voxel-wise general linear model (GLM) was applied to test for differences between Visits 1 and 2, using permutation-based non-parametric testing. Two contrasts were examined: Visit 2 > Visit 1 and Visit 1 > Visit 2 to assess directionality of any significant differences; that is, whether any differences were increased or decreased GM volume between the two sessions. Corrections for multiple comparisons across the brain were done using threshold-free cluster enhancement (TFCE) (Smith & Nichols, 2009).

For the subcortical structures, a vertex-based analysis was applied via the FIRST software package of FSL (Patenaude, Smith, Kennedy & Jenkinson, 2011). Several structures were extracted including the bilateral hippocampus, amygdala, thalamus, globus pallidus, putamen, and caudate nucleus. Although changes in the amygdala have not been reported in the bilingual literature, we decided to add them to this analysis due to the close proximity and relationship with the hippocampus (Schumann, 1990, 2001). Quality of the extractions was then assessed. No images were discarded. The resulting images were then submitted to vertex analyses. Per standard procedure, each structure underwent a 6 degrees of freedom rigid body transformation to study-specific template in standard space. The vertex coordinates of individuals were then projected onto the average coordinates of the two scanning scans. For each participant, this created maps signifying positive (outside the surface) or negative (inside the average surface) values

for each structure. Finally, the values were analyzed using a voxel wise GLM (Smith & Nichols, 2009). The contrasts examined were Visit 2 > Visit 1 and Visit 1 > Visit 2, to examine what changes occurred (both expansions and contractions) between the two scanning points. Age and sex were included in the model as covariates of no interest. Corrections for multiple comparisons were done via the Randomise pipeline within FIRST (Winkler, Ridgway, Webster, Smith & Nichols, 2014). This created maps of within-groups differences, thresholded at $p \leq 0.05$.

For the analysis of structural connectivity, we looked at FA and MD values across the brain as the most commonly used indices of white matter integrity. FA and MD values were calculated using the FDT (Behrens, Woolrich, Jenkinson, Johansen-Berg, Nunes, Clare, Matthews, Brady & Smith, 2003) and TBSS (Smith et al., 2006) pipelines within FSL. Images were corrected for head motion and eddy-current distortions, then a diffusion-tensor model was fit for each voxel of the corrected data, using DTIFIT. This resulted in individual FA and MD images for each participant from Visits 1 and 2 respectively. Using TBSS, the FA images were (in a combined process) non-linearly registered to a standard space FA target image and affine-transformed to MNI standard space. This resulted in a 4D image which consisted of each FA image from the participants. An FA skeletonisation program was used to create an FA skeleton that included the voxels identified as white matter (WM) in each FA image, thresholded at 0.2. MD images were then nonlinearly registered to standard space and were then warped and registered into a single 4D file which was projected onto the mean FA skeleton. Finally, a within-groups voxel wise analysis was applied to test for differences between Visit 1 and 2 for the FA and MD data respectively. The contrasts examined were Visit 2 > Visit 1 and Visit 1 > Visit 2, to assess the directionality of the changes. Age and sex were included in the model as nuisance covariates. This resulted in a whole-brain t-statistical image of significant differences in FA and MD between the two scan points.

For the resting-state data, we conducted an independent components analysis (ICA), as this offers a data-driven approach to analysis, and mitigates the risks of potential biases in the data imposed by the more traditional seed-based resting state analysis (for discussion see Beckmann & Smith, 2004). The data were analyzed using the Multivariate Exploratory Linear Optimized Decomposition into Independent Components (MELODIC) pipeline within FSL (Beckmann & Smith, 2004, 2005). Resting-state data was preprocessed including motion corrections, corrections for field distortions, and registered to standard space. The processed datasets were then decomposed into spatial and temporal components using a multi-session temporal concatenation. This process involves the concatenation

of all subjects' preprocessed 4D datasets, which subsequently are submitted to an independent components analysis. This generates a series of spatial maps which correspond to common components across each subject. Once components were established, these were submitted to group-level analysis via the dual_regression pipeline within FSL (Beckmann, Mackay, Filippini & Smith, 2009). This involves a two-stage process in which the group-level spatial maps (common components) are regressed into each subject's 4D dataset to render a series of time courses. The time courses are in turn regressed into the same dataset to create a subject-specific set of spatial maps. These spatial maps are then compared across subjects via a within-subjects GLM, assessing Visit 2 > Visit 1 and Visit 1 > Visit 2, with age and sex included in the model as nuisance covariates. Correction for family-wise error was done with Randomise permutation testing (Winkler et al., 2014).

Results

Proficiency

We first looked at whether the participants' proficiency changed as a function of linguistic immersion. Results of a Wilcoxon's signed rank test showed that no significant change in the QPT scores occurred over the three-year period (Mdifference = 3.481%, SD = 12.03, p = .515). However, this is not surprising, as our participants were already at a proficiency ceiling at Visit 1, as measured by their performance on the QPT (Table 2). Spearman's Rho correlations between immersion (time spent in the UK at Visit 1), English (L2) proficiency (QPT score at Visit 1), and L2 AoA showed a marginally significant correlation for immersion and proficiency, $r_s = .661$, p = .053. None of the other correlations approached significance (proficiency and AoA: $r_s = -.371$, p = .325; AoA and immersion: $r_s = -.017$, p = .996). Finally, we also observed that the proficiency x immersion correlation disappeared at Visit 2 ($r_s = .085$, p = .753).

Grey matter

Results of voxel-based morphometry (VBM) analysis showed a significant, corrected, increase in GM volume for Visit 2 over Visit 1 in the lower VIIb region of the left cerebellum (peak coordinates-32, -68, -58; 115 voxels; p = .033) (Fig. 1). No significant decreases (Visit 1 > Visit 2) were found. We assessed the normality of the distribution of the extracted volume changes within the cerebellum. This was done to ensure that the effect was not being driven artificially by a small portion of the participant sample. A Shapiro-Wilk test showed the changes in GM volume to be normally distributed (p = .887), with no outliers.
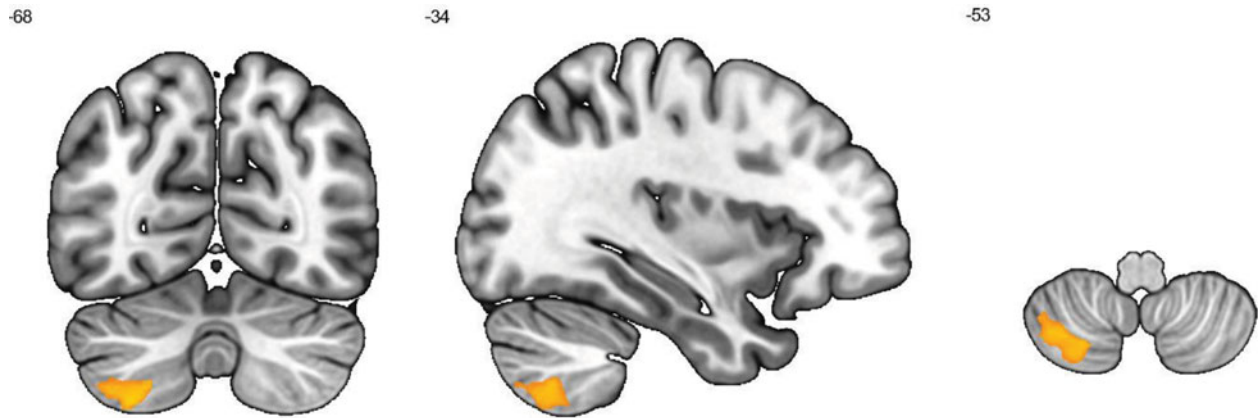
Figure 1. Results of VBM analysis. Significant grey matter volume change (orange) within the right cerebellum. Coordinates listed in MNI-space.

Table 3. *Results of vertex analysis corrected for multiple comparisons (thresholded p≤.0021)*

Structure	Hemisphere	Direction	Voxels	p	x	y	z
Caudate	L	Contraction	597	0.002	−18	−8	21
Hippocampus	R	Expansion	15	<0.001	20	−29	−10
	L	Contraction	85	0.001	−27	−19	−13
	R	Contraction	138	<0.001	23	−22	−19
Amygdala	L	Contraction	83	<0.001	−28	−5	−18

Our subcortical analysis produced no significant effects in the corrected results from the permutation analysis, which is not surprising given the sample size. Significant (uncorrected) contractions and expansions from the vertex analysis were thresholded at $p \leq 0.002$ to account for multiple comparisons (dividing the target significance threshold by the number of tests run (24- two tests per 12 structures)). Expansions and contractions were found in several structures, including a contraction in the left caudate nucleus, a contraction in the left amygdala, a small expansion and larger contraction in the ventral anterior portion of the right hippocampus, and a contraction the dorsal anterior portion of the left hippocampus. Table 3 illustrates these effects. Changes in the putamen, globus pallidus or the thalamus did not survive thresholding.

White matter

No significant differences of FA values (increase or decrease) were found between Visits 1 and 2 for any WM tracts. Significant, corrected, increases in MD values were found in two clusters in the left forceps minor (-20, 44, 3, cluster size: 174 1 mm^3 voxels; and -19, 34, 16, cluster size: 27 1 mm^3 voxels). These effects are illustrated in Figure 2.

Resting state networks

No significant differences (increase or decrease) were found in functional connectivity for any resting state networks between Visits 1 and 2.

Regression analyses

To determine the role of the subjects' language experience in shaping brain structure, we ran multiple regression analyses on the affected brain regions using immersion (months of residence in the UK at Visit 1) and L2 AoA as predictor variables, and age and sex as nuisance covariates. Specifically, the volumetric change in the cerebellum and the MD change in forceps minor were analyzed respectively with the above predictors. Given that proficiency at Visit 1 was found to correlate with length of immersion, it was not included in the final model to avoid issues of multicollinearity.

The model was found to significantly predict the longitudinal cerebellar GM increase, $F(4,4) = 10.73$,
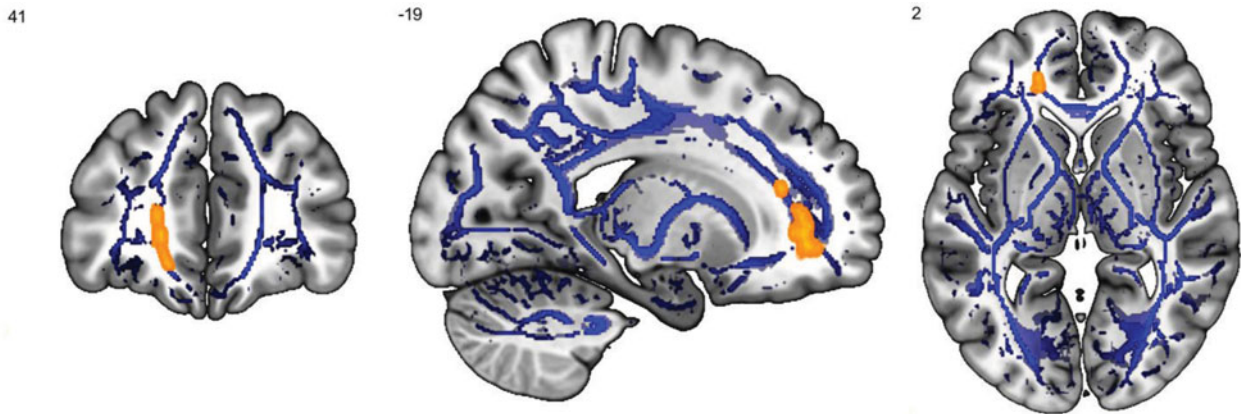
Figure 2. Results of MD analysis. Increases in MD values (orange) in the left forceps minor (blue). Coordinates listed in MNI-space.
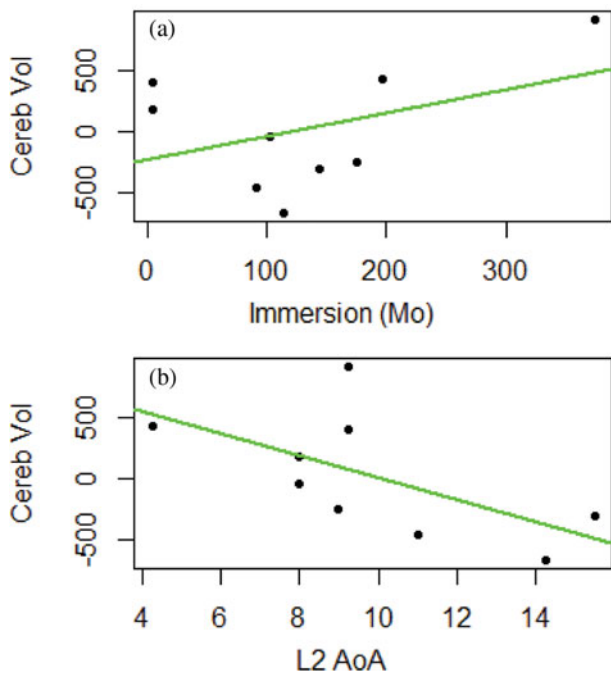

Figure 3a,b. (Colour online) Relationship between cerebellar volume change and Immersion at Visit (a), and L2 AoA (b).

$p<0.021$, $r^2 = .829$. Specifically, L2 AoA was found to negatively correlate with the cerebellar increase ($\beta = -.815$, $p = .0066$) and length of immersion was found to positively correlate with it ($\beta = 1.391$, $p = .00815$) suggesting that the greater the immersion, and the younger the AoA, the more plastic the cerebellum became. Figure 3 illustrates this. The model did not significantly predict the changes in MD values in forceps minor (all ps > 0.1).

Discussion

The working hypothesis of this paper has been that it is not bilingualism *per se*, but potentially particular variables (and not others) related to the dynamics of the bilingual experience that induce anatomical brain changes. If the hypothesis is on the right track, the interplay of certain experience-based variables and their relative weight should correlate to individual subject and cross-aggregate differences outcomes in individual studies and, by extension, explain at least some of the disparities across the literature. Herein, bilingual subjects were scanned twice with a three-year interval between scans while residing in the UK. Recall that there was a range of exposure in the same L2 immersion environment at Visit 1 (range = 4 months to 12 years); however, all subjects share a crucially common experience – that is, the same amount of time in between Visits 1 and 2. Also recall that our analysis was focused on change that took place from the point of each individual's own baseline (change between Visit 1 and 2), which crucially did not coincide to any major change in the lifestyle and/or experiences of our participants (e.g., arrival in the UK), or at least any change that applied to the entire sample.

Several changes were seen between the two scan points: a significant increase in grey matter volume was found in the lower left cerebellum, and reshaping in several subcortical structures including the (bilateral) hippocampus, left amygdala, and left caudate. Moreover, we observed a significant increase in MD values within the left forceps minor. These results add to a growing body of literature supporting the role of linguistic immersion in neuroplasticity related to bilingualism (Kuhl et al., 2016; Pliatsikas et al., 2017, 2015). Taken together, the grey matter adaptations support aspects of both the BAPSS model (Grundy et al., 2017) and ACH (Abutalebi & Green, 2016). The changes in the cerebellum, forceps minor, caudate, and hippocampus

show an overall adaptation within the language control network towards a more automated system of processing and production in the L2, as demonstrated below.

The increase in cerebellar GM volume partly replicates findings from previous work comparing bilinguals with limited immersion to age-matched monolinguals (Pliatsikas et al., 2014b) – specifically the same region of the cerebellum was found to expand for both studies. The cerebellar expansion over the three year period is in line with aspects of the BAPSS and ACH models – specifically that the increased reliance on subcortical/posterior structures in the language control network is commensurate with increased time spent intensively using the L2 (Abutalebi & Green, 2016; Grundy et al., 2017), which in turn leads to increases in GM volume in these areas. It should be noted that the BAPSS model does not explicitly include the cerebellum in the posterior structures that are implicated in the shift with increased L2 exposure. However, the cerebellum has been implicated in several functions related to language including automated processing of grammatical rules in one's non-native language (Pliatsikas et al., 2014b; Ullman, 2004), language control between the L1 and L2 (Filippi, Richardson, Dick, Leech, Green, Thomas & Price, 2011), and articulatory processes for bilinguals (Frenck-Mestre, Anton, Roth, Vaid & Viallet, 2005; Grogan et al., 2009). The connection with the time spent using the L2 is corroborated by the correlations with both L2 AoA and immersion seen in the multiple regression. The negative correlation with L2 AoA suggests that the earlier one acquires one's L2 the greater plasticity in the cerebellum may be predicted. Similarly, the positive correlation between immersion and the cerebellar increase indicates that the longer one is immersed in the environment of the L2, the more likely one is to experience change in this region. The cerebellar expansion related to L2 AoA and immersion appears to support an account of increased reliance on the cerebellum, as processing and production in the L2 becomes more automated. In any case, the present pattern of effects suggests that the more experienced in an L2 one is, the more plastic the cerebellum becomes.

The reduction in the left caudate likely indicates a decreased switching cost between the participants' languages (Elmer, Hänggi & Jäncke, 2014). The caudate has been implicated in a number of control processes in bilingual language processing and production (Abutalebi & Green, 2016; Luk, Green, Abutalebi & Grady, 2011). The reduction found in the caudate may thus reflect an optimization of the system to the language control demands within the immersion environment. An alternative explanation for this comes from the CONDITIONAL ROUTING model (Stocco, Yamasaki, Natalenko & Prat, 2014), which suggests that increased automation in language switching and selection occurs in

cortical regions, thus decreasing demand on the caudate, resulting in contractions. Similarly, recall that both expansions and contractions within the right hippocampus were found. It should be noted that the ventral anterior contraction in the right hippocampus overlaps with the expansion found by Bellander and colleagues (2016), who report the increased GM volume in the right hippocampus to be predicted by number of hours spent acquiring new words. In the case of the immersed bilinguals, the contraction likely indicates a return to baseline from an expansion at the first stages of L2 acquisition and/or immersion (e.g., Bellander et al., 2016; Mårtensson et al., 2012). The contraction in the right hippocampus may also be evidence for proceduralization within the L2 (Pliatsikas et al., 2014b; Ullman, 2004). Increased reliance on the procedural system in the processing and production in the non-native language (thus increased use on the cerebellum) would entail decreased use of the declarative memory systems (thus a decreased use of the hippocampus), leading to contractions in this structure.

The contraction in the left amygdala was not predicted, as it has not been typically reported in studies about bilingualism-induced structural changes, nor is it typically implicated in language processing and control in bilinguals (see e.g., Abutalebi & Green, 2016; Grundy et al., 2017; Stocco & Prat, 2014). The only exception is that Li and colleagues (Li, Abutalebi, Emmorey, Gong, Yan, Feng, Zou & Ding, 2017) found increased GM volume in the left amygdala/hippocampus for older bimodal bilinguals who were active daily users of both languages. It has been proposed that the amygdala forms part of a corticofugal pathway for memory formation, and plays a role in regulation, stimulus appraisal, and motivation (based on emotional valence) in the formation of new memories in L2 acquisition (Schumann, 1990, 2001). The contraction in the amygdala, then, indicates a lower reliance on short-term/declarative memory formation procedures, in line with the contractions found in the hippocampus. However, a lack of neurolinguistic data currently exists to support this interpretation. It is also worth reiterating that the effects found in the hippocampus, caudate, and amygdala were found not to survive correction for multiple comparisons, thus should be interpreted with caution.

No differences were seen in the globus pallidus, putamen, or thalamus during the three year period, which is not in line with our predictions (Berken, Gracco, et al., 2016; Burgaleta, Sanjuán, Ventura-Campos, Sebastián-Gallés & Ávila, 2016; Pliatsikas et al., 2017). This lack of difference between Visits 1 and 2 for these structures can be interpreted as no change in the reliance on them during this period. Demographically, our group is highly similar to the group in Pliatsikas et al. (2017), which reported structural changes in both structures for bilinguals vs. monolinguals. If the thalamus is crucial

in selecting among lexical and semantic alternatives (Abutalebi & Green, 2016; Llano, 2013), the absence of any longitudinal structural changes probably reflects that the control needs did not change between Visit 1 and Visit 2, which is to be expected as in both cases our participants were immersed in the same environment. The putamen has been implicated in phonological monitoring in bilingual language production (Abutalebi, Della Rosa, Castro Gonzaga, Keim, Costa & Perani, 2013). Similarly to the thalamus, the lack of change in the putamen indicates that the demands on the underlying system will not have changed. Finally, a similar explanation may also be given for the absence of change in the globus pallidus: that the demands on semantic control and selection would not have changed, even after periods of extensive immersion. An alternative explanation for the lack of changes in these structures might be that the selection and monitoring mechanisms were already optimized at Visit 1 and, as the relevant cognitive demands did not change between visits, the structure of the implicated regions remained stable. This would align with tenets of the ACH (Abutalebi & Green, 2016), specifically that continued plasticity of given brain regions would be dependent on changes in language use and/or exposure.

No significant changes in FA values were found in any region of the brain. This finding is not in line with the proposals of the BAPSS model (Grundy et al., 2017), which would predict increased use of subcortical to posterior tracts, and thus increased myelination in those tracts linking the subcortical structures to those regions. It is important again to remember that our group in this study was demographically very similar to the group from Pliatsikas et al. (2015), meaning that higher levels of FA can already be assumed at Visit 1, as an effect of continuous prior exposure. What we didn't find here is further FA increase OR decrease WITHIN participants. An explanation for this lack of changes could be the same as the one for the absence of thalamic effects: the same needs for efficient communication between brain areas applied to both test points, so an already optimized system retained its structure. Not only would this explain the lack of FA DECREASES in Visit 2, but also reconciles with the fact that increased FA persists even in older bilinguals compared to monolinguals, while other effects typically disappear. In the same vein, the lack of INCREASES in FA might reflect either that the system has reached and maintained the appropriate level of "reinforcement" needed to deal with the demands of the bilingual experience, or more simply, that there are physical constraints in white matter restructuring, that have now been reached. The small increases in MD in the forceps minor are less readily explainable. Forceps minor is the most anterior part of the corpus callosum, which projects to the frontal lobes, and has been associated with executive control (Kuhl et al., 2016). While several anterior and medial

parts of the corpus callosum are shown to increase their FA and/or volume as a result of bilingualism (Coggins, Kennedy & Armstrong, 2004; Felton, Vazquez, Ramos-Nunez, Greene, Macbeth, Hernandez & Chiarello, 2017; Pliatsikas et al., 2015), changes in the MD of the forceps minor have only been reported in two studies: Cummine and Boliek (2013) reported decreased MD values for bilinguals vs. monolinguals in the left forceps minor, while Kuhl et al. (2016) reported negative correlations between MD in this region and the bilinguals' residence in the L2 speaking country, in that the smaller the immersion the larger the MD values. Higher MD values are typically explained as higher white matter diffusivity, which might signify reduced myelination. With this in mind, an interesting pattern emerges: the participants in both Cummine and Boliek and Kuhl et al.'s studies were at initial stages of L2 immersion, and showed increases in myelination, expressed as decreases in MD (but not increases in FA). On the other hand, our highly immersed participants showed small increases in MD but not decreases in FA. Although the exact correspondence between FA and MD is not fully understood, it appears that initial immersion causes changes in the MD of frontal areas, which are crucial for cognitive control. With increased immersion, it appears that overall diffusivity is also increased, possibly reflecting less reliance in the region because efficient control of languages has been achieved, while the directionality of the diffusivity, which is better expressed by the FA, remains unchanged. This suggestion is congruent with the BAPPS prediction of reduced reliance in frontal regions as an effect of L2 immersion. Nevertheless, more research is needed to understand the complex relationships between the FA and the MD, and how these are affected by the bilingual experience.

No significant differences in resting-state connectivity were seen between Visits 1 and 2. As this analysis was exploratory, we did not have any specific predictions about how the resting-state networks would change between Visits 1 and 2.

It is also worth noting that L2 proficiency was not found to increase between visits. Regarding change in proficiency, it is prudent to note a few things that render the lack of effects less surprising. Firstly, the participants were already highly proficient in English at the point of Visit 1 (see 'participants' section for details). Given that L2 learners tend to show ceiling effects in ultimate attainment that differ from typical monolingual acquisition (see Long, 2007; Rothman, 2008 for opposing views on how to interpret such finds), their initial proficiency was likely high enough that continued immersion effects would not result in changes as measured by the QPT. In this context then, we need to keep in mind what the QPT and measures like it are, and what its purpose is in our study. The QPT is designed to gauge so-called global proficiency

and such measures tend to focus on properties of the formal grammar, especially lexical-semantic knowledge, as opposed to colloquial language, overall verbal fluency and other language aspects that are likely to be positively affected by continued immersion, even in the case that someone has a highly proficient grammar at the outset. By GRAMMAR we are referring to the set of underlying mental representations and rules that form the computational system enabling an individual to comprehend (decode linguistic information received) and produce (encoding information onto language specific forms to be uttered) intended messages for any given language. The purpose of a proficiency measure is most crucial at Visit 1 to get an initial baseline, ensure a minimum level of proficiency for inclusion in a study, and for purposes of comparison across studies (present and future ones) who have similar subject profiles, including L2 proficiency. Were there to be a measure that could tease apart gains in the above-mentioned domains of language use for L2 learners, we might expect three additional years of immersion would show considerable gains. The fact that there is no observed increase in proficiency, despite clear changes to the brain in the same time span, already suggests that measures like the QPT are unable to capture the full picture of what is going on. Although L2 grammatical knowledge as measured by the QPT can reach a ceiling effect for such measurement, clearly continued exposure is still having significant effects that would otherwise not be captured, save for the use of different methodologies that are more fine-grained (such as MRI). For these reasons, and because linguistic immersion and proficiency are not independent to each other, as our results showed, we chose not to include proficiency as a predictor in our models.

This said, from a linguistic perspective it is not clear why proficiency per se would predict differences, provided that we are looking at individuals with at least a minimum level of exposure to and proficiency in the L2 (both serving as proxies for opportunity to develop a mental L2 grammar). This is evidenced by the correlation between length of L2 immersion and proficiency levels at Visit 1 and the fact that this correlation disappears after the three year immersion period, as by then all would have surpassed the minimum threshold. Even more important, however, is the fact that when one has high degrees of verbal fluency in an L2, even if extremely different from what natives of the target grammar display, it is regarded as a complete grammar itself, referred to in the literature as INTERLANGUAGE (Selinker, 1972). The basic idea is that at any given point in L2 development, the L2 grammar is a full grammar, much as it is for children developing their L1. If on the right track, then L2 proficiency is merely a measure of how native-like one is on a continuum, not a measure of how complete (in the sense of being considered a mental computational grammar itself) the developing system is. In this sense, the activation vs.

inhibition tension that likely underlies changes to the bilingual brain (behavioral and anatomical) starts and is functionally much earlier than might otherwise be expected. The brain does not know when a grammar is target-like, it simply knows when there is a need to inhibit a competing system regardless of its stage of development. From such a perspective then, it is not clear that relative proficiency should matter. This is, of course, an empirical question.

Taken together, our results show that immersive bilingualism continuously affects the structure of a dynamic system including the cerebellum, the hippocampus and the basal ganglia, as well as the integrity of the white matter, even after years of bilingual immersion. The reported patterns generally support the predictions of the BAPSS model and ACH, with some potential modifications. The expansions in the cerebellum and the reshaping/contractions in the caudate nucleus and hippocampus, and potentially the MD increase in the forceps minor, serve as neurological markers of increased efficiency and automation of processing and production in the L2 (Abutalebi & Green, 2016; Booth, Wood, Lu, Houk & Bitan, 2007; Filippi et al., 2011; Pliatsikas et al., 2017), while the overall stability of the white matter diffusivity across the entire brain, paired with the structural stability of the thalamus and the putamen, suggest that the language control demands persist irrespective of the amount of L2 immersion. Future work should consider the cerebellum as a key structure in brain adaptation to L2 exposure and use, particularly at later stages of development.

An obvious limitation of our study was the high attrition rate of participants between Visits 1 and 2, thus our sample size ended up being fairly low (n = 9). Nevertheless, the cerebellar and MD effects survived statistical corrections even with such a small sample, such that they are nonetheless reliable. A larger group at Visit 2 would have probably produced more robust findings in both cortical and subcortical regions. The GMV increase in the cerebellum would likely remain, along with decreases in GMV in frontal regions (Grundy et al., 2017). The contractions in the caudate and hippocampus would likely also remain. Finally, we might have expected to see modulation in white matter integrity in pathways connecting the cerebellum to the basal ganglia and frontal regions. However, these predictions require further investigation with an appropriate number of participants. Future longitudinal studies should strive to implement appropriate strategies in order to retain a greater number of participants in their final cohort. This of course likely requires a larger number of participants at the outset since attrition is common in studies like this, especially given a duration of three years between testing sessions. Future studies should also incorporate an extensive behavioral battery tapping at both executive functions and language abilities in L1 and L2, in order to study how these change

as a factor of immersion, and whether they are linked with the observed structural changes.

Finally, a control group was not reported in this study, for several reasons. Despite our attempts to test a monolingual native-English speaking cohort as the control group (see footnote 1), we would still treat any findings with caution. Inclusion of such a cohort represents a comparative fallacy in examining specific neural effects of linguistic immersion, as these cannot be teased apart from any other effects related to the bilingual experience more generally. In other words, if we included a control group of this kind our null hypothesis would have been that the observed effects should appear in both language groups, or that the cerebellum continues to expand for everyone, no matter the language status. However, we certainly don't have such a prediction, but we do have the valid prediction that the cerebellum reshapes as a result of the bilingual experience. Since our participants act as their own controls, we fail to see what the addition of a monolingual control group could add to the narrative. Conversely, the appropriate control group to use here, and the one we would suggest for future studies, would have been a highly-proficient bilingual cohort living in a country where their L2 (in this case English) is not the dominant language for communication, which would allow for the examination of effects of linguistic immersion independent of other potentially conflating experiences in bilingualism. Second – recall that the bilingual cohort in this was not homogeneous – they exhibited both a range of AoAs of English and length of immersion at Visit 1, and furthermore stemmed from a variety of professions. It is increasingly likely, then, that the one major commonality in their experiences (and thus neural outcomes) could be related to linguistic immersion within the three-year (longitudinal) period. Potential alternative explanations of our findings would include them being a result of major lifestyle changes in our participants' lives that might induce neural adaptations, such as taking up a sport or learning a musical instrument. To the best of our knowledge no such activity was taken up by all of our participants after Visit 1. Other potential causes of structural changes include ageing, which is hard to establish given the wide age range of our sample (and the fact that we added age as a covariate in our models), major changes of environment, such as recent migration, and the emotional imbalance it may bring about, which clearly does not apply to our already immersed participants, and major changes in general quality of life. The latter is a very broad concept encompassing a variety of factors (socioeconomic status, family experiences, education, general health), which is difficult (if not impossible) to test systematically; still, any changes to quality of life would have to apply to all of our participants (and to an entire control group for that matter) in order to produce these group effects.

Conclusions

In this study – the first to look at the longitudinal effects of L2 immersion in a group of highly proficient adult bilinguals – we showed that L2 exposure and use continue to impact brain structure beyond acquisition and initial stages of use. Crucially, bilingualism (and immersion where it applies) is a dynamic process in which brain adaptations are modulated through time by exposure and changes in efficiency of production and processing. Our approach and results also support current arguments that future research should move away from traditional cross-sectional comparative (bilingual vs. monolingual) designs, and turn its attention to the experience of bilingualism itself, with a focus on experience-based factors to be used as predictors in assessing the specific impact of bilingualism on brain structure and function. Although exceedingly difficult to shift towards true longitudinal studies where change can be tracked within individuals as the dynamics of bilingualism unfold over time, the sacrifices (e.g., numbers of participants in light of attrition, the temporal length of the studies themselves, etc.) one will need to make will be overshadowed by the increased precision and comparability that intragroup comparisons provide.

References

Abutalebi, J., Canini, M., Della Rosa, P. A., Green, D. W., & Weekes, B. S. (2015). The neuroprotective effects of bilingualism upon the inferior parietal lobule: A Structural Neuroimaging Study in Aging Chinese Bilinguals. *Journal of Neurolinguistics*, *33*, 3–13. https://doi.org/10.1016/j.jneuroling.2014.09.008

Abutalebi, J., Della Rosa, P. A., Castro Gonzaga, A. K., Keim, R., Costa, A., & Perani, D. (2013). The role of the left putamen in multilingual language production. *Brain and Language*, *125*(*3*), 307–315. https://doi.org/10.1016/j.bandl.2012.03.009

Abutalebi, J., & Green, D. W. (2016). Neuroimaging of language control in bilinguals: neural adaptation and reserve. *Bilingualism: Language and Cognition*, (April), 1–10. https://doi.org/10.1017/S1366728916000225

Ashburner, J., & Friston, K. J. (2000). Voxel-Based Morphometry – The Methods. *NeuroImage*, *11*(*6*), 805–821. https://doi.org/10.1006/nimg.2000.0582

Bak, T. H. (2016). Cooking pasta in La Paz. *Linguistic Approaches to Bilingualism*, *5*(*2016*), 1–19. https://doi.org/10.1075/lab.16002.bak

Beckmann, C. F., Mackay, C. E., Filippini, N., & Smith, S. M. (2009). Group comparison of resting-state FMRI data using multi-subject ICA and dual regression. *NeuroImage*, *47*(*Suppl 1*), S148. https://doi.org/10.1073/pnas.0811879106

Beckmann, C. F., & Smith, S. M. (2004). Probabilistic Independent Component Analysis for Functional Magnetic Resonance Imaging. *IEEE Transactions on Medical Imaging*, *23*(*2*), 137–152. https://doi.org/10.1109/TMI.2003.822821

Beckmann, C. F., & Smith, S. M. (2005). Tensorial extensions of independent component analysis for multisubject FMRI analysis. *NeuroImage*, *25*(*1*), 294–311. https://doi.org/10.1016/j.neuroimage.2004.10.043

Behrens, T. E. J., Woolrich, M. W., Jenkinson, M., Johansen-Berg, H., Nunes, R. G., Clare, S., Matthews, P. M., Brady, J. M., & Smith, S. M. (2003). Characterization and Propagation of Uncertainty in Diffusion-Weighted MR Imaging. *Magnetic Resonance in Medicine*, *50*(*5*), 1077–1088. https://doi.org/10.1002/mrm.10609

Bellander, M., Berggren, R., Mårtensson, J., Brehmer, Y., Wenger, E., Li, T. Q., Bodammer, N. C., Shing, Y. L., Werkle-Bergner, M., & Lövdén, M. (2016). Behavioral correlates of changes in hippocampal gray matter structure during acquisition of foreign vocabulary. *NeuroImage*, *131*, 205–213. https://doi.org/10.1016/j.neuroimage.2015.10.020

Berken, J. A., Chai, X. J., Chen, J.-K., Gracco, V. L., & Klein, D. (2016). Effects of Early and Late Bilingualism on Resting-State Functional Connectivity. *Journal of Neuroscience*, *36*(*4*), 1165–1172. https://doi.org/10.1523/JNEUROSCI.1960-15.2016

Berken, J. A., Gracco, V. L., Chen, J. K., & Klein, D. (2016). The timing of language learning shapes brain structure associated with articulation. *Brain Structure and Function*, *221*(*7*), 3591–3600. https://doi.org/10.1007/s00429-015-1121-9

Berken, J. A., Gracco, V. L., & Klein, D. (2017). Early bilingualism, language attainment, and brain development. *Neuropsychologia*, *98*, 220–227. https://doi.org/10.1016/j.neuropsychologia.2016.08.031

Bialystok, E. (2016). The signal and the noise: Finding the pattern in human behavior. *Linguistic Approaches to Bilingualism*, *6*(*5*), 517–534. https://doi.org/10.1075/lab.15040.bia

Bialystok, E. (2017). The Bilingual Adaptation: How Minds Accommodate Experience. *Psychological Bulletin*, *143*(*3*), 233–262. https://doi.org/10.1037/bul0000099

Booth, J. R., Wood, L., Lu, D., Houk, J. C., & Bitan, T. (2007). The role of the basal ganglia and cerebellum in language processing. *Brain Research*, *1133*, 136–144. https://doi.org/10.1016/j.brainres.2006.11.074

Burgaleta, M., Sanjuán, A., Ventura-Campos, N., Sebastián-Gallés, N., & Ávila, C. (2016). Bilingualism at the core of the brain. Structural differences between bilinguals and monolinguals revealed by subcortical shape analysis. *NeuroImage*, *125*, 437–445. https://doi.org/10.1016/j.neuroimage.2015.09.073

Coggins, P. E., Kennedy, T. J., & Armstrong, T. A. (2004). Bilingual corpus callosum variability. *Brain and Language*, *89*(*1*), 69–75. https://doi.org/10.1016/S0093-934X(03)00299-2

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

Elmer, S., Hänggi, J., & Jäncke, L. (2014). Processing demands upon cognitive, linguistic, and articulatory functions pro-mote grey matter plasticity in the adult multilingual brain: Insights from simultaneous interpreters. *Cortex*, *54*(*1*), 179–189. https://doi.org/10.1016/j.cortex.2014.02.014

Felton, A., Vazquez, D., Ramos-Nunez, A. I., Greene, M. R., Macbeth, A., Hernandez, A. E., & Chiarello, C. (2017). Bilingualism influences structural indices of interhemispheric organization. *Journal of Neurolinguistics*, *42*(May), 1–11. https://doi.org/10.1016/j.jneuroling.2016.10.004

Filippi, R., Richardson, F. M., Dick, F., Leech, R., Green, D. W., Thomas, M. S. C., & Price, C. J. (2011). The Right Posterior Paravermis and the Control of Language Interference. *Journal of Neuroscience*, *31*(*29*), 10732–10740. https://doi.org/10.1523/JNEUROSCI.1783-11.2011

Frenck-Mestre, C., Anton, J. L., Roth, M., Vaid, J., & Viallet, F. (2005). Articulation in early and late bilinguals' two languages: evidence from functional magnetic resonance imaging. *Neuroreport*, *16*(*7*), 761–765. https://doi.org/10.1097/00001756-200505120-00021

García-Pentón, L., Fernández García, Y., Costello, B., Duñabeitia, J. A., & Carreiras, M. (2016). The neuroanatomy of bilingualism: how to turn a hazy view into the full picture. *Language, Cognition and Neuroscience*, *3798*(September), 1–25. https://doi.org/10.1080/23273798.2015.1068944

Geranpayeh, A. (2003). A quick review of the English Quick Placement Test. *UCLES Research Notes*, (*12*), 8–10.

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. https://doi.org/10.1016/j.neuropsychologia.2013.09.037

Green, D. W. (2011). Language control in different contexts: The behavioral ecology of bilingual speakers. *Frontiers in Psychology*, *2*(MAY), 2009–2012. https://doi.org/10.3389/fpsyg.2011.00103

Green, D. W., & Abutalebi, J. (2013). Language control in bilinguals: The adaptive control hypothesis. *Journal of Cognitive Psychology*, *25*, 1–16. https://doi.org/10.1080/20445911.2013.796377

Grogan, A., Green, D. W., Ali, N., Crinion, J. T., & Price, C. J. (2009). Structural correlates of semantic and phonemic fluency ability in first and second languages. *Cerebral Cortex*, *19*(*11*), 2690–2698. https://doi.org/10.1093/cercor/bhp023

Grundy, J. G., Anderson, J. A. E., & Bialystok, E. (2017). Neural Correlates of Cognitive Processing in Monolinguals and Bilinguals. *Annals of the New York Academy of Sciences*. https://doi.org/10.1111/nyas.13333

Hofstetter, S., Friedmann, N., & Assaf, Y. (2016). Rapid language-related plasticity: microstructural changes in the cortex after a short session of new word learning. *Brain Structure and Function*. https://doi.org/10.1007/s00429-016-1273-2

Hosoda, C., Tanaka, K., Nariai, T., Honda, M., & Hanakawa, T. (2013). Dynamic Neural Network Reorganization Associated with Second Language Vocabulary Acquisition: A Multimodal Imaging Study. *The Journal of Neuroscience*, *33*(*34*), 13663–13672. https://doi.org/10.1523/JNEUROSCI.0410-13.2013

Jenkinson, M., Beckmann, C. F., Behrens, T. E. J., Woolrich, M. W., & Smith, S. M. (2012). Fsl. *NeuroImage*, *62*(2), 782–790. https://doi.org/10.1016/j.neuroimage.2011.09.015

Klein, D., Mok, K., Chen, J. K., & Watkins, K. E. (2014). Age of language learning shapes brain structure: A cortical thickness study of bilingual and monolingual individuals. *Brain and Language*, *131*, 20–24. https://doi.org/10.1016/j.bandl.2013.05.014

Kousaie, S., Chai, X. J., Sander, K. M., & Klein, D. (2017). Simultaneous learning of two languages from birth positively impacts intrinsic functional connectivity and cognitive control. *Brain and Cognition*, *117*(June), 49–56. https://doi.org/10.1016/j.bandc.2017.06.003

Kroll, J. F., & Chiarello, C. (2016). Language experience and the brain: variability, neuroplasticity, and bilingualism. *Language, Cognition and Neuroscience*, *31*(3), 345–348. https://doi.org/10.1080/23273798.2015.1086009

Kuhl, P. K., Stevenson, J., Corrigan, N. M., van den Bosch, J. J. F., Can, D. D., & Richards, T. L. (2016). Neuroimaging of the bilingual brain: Structural brain correlates of listening and speaking in a second language. *Brain and Language*, *162*, 1–9. https://doi.org/10.1016/j.bandl.2016.07.004

Li, L., Abutalebi, J., Emmorey, K., Gong, G., Yan, X., Feng, X., Zou, L., & Ding, G. (2017). How bilingualism protects the brain from aging: Insights from bimodal bilinguals. *Human Brain Mapping*, *4124*, 4109–4124. https://doi.org/10.1002/hbm.23652

Llano, D. A. (2013). Functional imaging of the thalamus in language. *Brain and Language*, *126*(1), 62–72. https://doi.org/10.1016/j.bandl.2012.06.004

Long, M. H. (2007). *Problems in SLA*. Mahwah, NJ.: Lawrence Erlbaum Associates.

Luk, G., & Bialystok, E. (2013). Bilingualism is not a categorical variable: Interaction between language proficiency and usage. *Journal of Cognitive Psychology*, *25*(5), 605–621. https://doi.org/10.1080/20445911.2013.795574

Luk, G., Bialystok, E., Craik, F. I. M., & Grady, C. L. (2011). Lifelong Bilingualism Maintains White Matter Integrity in Older Adults. *Journal of Neuroscience*, *31*(46), 16808–16813. https://doi.org/10.1523/JNEUROSCI.4563-11.2011

Luk, G., Green, D. W., Abutalebi, J., & Grady, C. L. (2011). Cognitive control for language switching in bilinguals: A quantitative meta-analysis of functional neuroimaging studies. *Language and Cognitive Processes*, *27*(10), 1479–1488. https://doi.org/10.1080/01690965.2011.613209

Luk, G., & Pliatsikas, C. (2016). Converging diversity to unity: commentary on The neuroanatomy of bilingualism. *Language, Cognition and Neuroscience*, *31*(3), 349–352. https://doi.org/10.1080/23273798.2015.1119289

Mamiya, P. C., Richards, T. L., Coe, B. P., Eichler, E. E., & Kuhl, P. K. (2016). Brain white matter structure and COMT gene are linked to second-language learning in adults. *Proceedings of the National Academy of Sciences*, 1–6. https://doi.org/10.1073/pnas.1606602113

Marian, V., & Spivey, M. J. (2003). Competing activation in bilingual language processing: Within- and between-language competition. *Bilingualism: Language and Cognition*, *6*(2), 97–115. https://doi.org/10.1017/S1366728903001068

Mårtensson, J., Eriksson, J., Bodammer, N. C., Lindgren, M., Johansson, M., Nyberg, L., & Lövdén, M. (2012). Growth of language-related brain areas after foreign language learning. *NeuroImage*, *63*(1), 240–244. https://doi.org/10.1016/j.neuroimage.2012.06.043

Mechelli, A., Crinion, J. T., Noppeney, U., O'Doherty, J., Ashburner, J., Frackowiak, R. S. J., & Price, C. J. (2004). Structural plasticity in the bilingual brain. *Nature*, *431*(7010), 1 p following 757; discussion following 757. https://doi.org/10.1038/nature03016

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. https://doi.org/10.1016/j.brainres.2011.12.005

Mohades, S. G., Van Schuerbeek, P., Rosseel, Y., Van De Craen, P., Luypaert, R., & Baeken, C. (2015). White-matter development is different in bilingual and monolingual children: A longitudinal DTI study. *PLoS ONE*, *10*(2), 1–16. https://doi.org/10.1371/journal.pone.0117968

Nichols, E. S., & Joanisse, M. F. (2016). Functional activity and white matter microstructure reveal the independent effects of age of acquisition and proficiency on second-language learning. *NeuroImage*, *143*, 15–25. https://doi.org/10.1016/j.neuroimage.2016.08.053

Olsen, R. K., Pangelinan, M. M., Bogulski, C. A., Chakravarty, M. M., Luk, G., Grady, C. L., & Bialystok, E. (2015). The effect of lifelong bilingualism on regional grey and white matter volume. *Brain Research*, *1612*, 128–139. https://doi.org/10.1016/j.brainres.2015.02.034

Osterhout, L., Poliakov, A., Inoue, K., McLaughlin, J., Valentine, G., Pitkanen, I., Frenck-Mestre, C., & Hirschensohn, J. (2008). Second-language learning and changes in the brain. *Journal of Neurolinguistics*, *21*(6), 509–521. https://doi.org/10.1016/j.jneuroling.2008.01.001

Patenaude, B., Smith, S. M., Kennedy, D. N., & Jenkinson, M. (2011). A Bayesian model of shape and appearance for subcortical brain segmentation. *NeuroImage*, *56*(3), 907–922. https://doi.org/10.1016/j.neuroimage.2011.02.046

Pliatsikas, C. (forthcoming, 2019). Multilingualism and brain plasticity. In J. W. Schweiter (Ed.), *The Handbook of the Neuroscience of Multilingualism* (pp. 0–9). Wiley Blackwell.

Pliatsikas, C., & Chondrogianni, V. (2015). Editorial: Learning a non-native language in a naturalistic environment: insights from behavioral and neuroimaging research. *Frontiers in Psychology*, *6*(July), 1–3. https://doi.org/10.3389/fpsyg.2015.01009

Pliatsikas, C., DeLuca, V. F., Moschopoulou, E., & Saddy, J. D. (2017). Immersive bilingualism reshapes the core of the brain. *Brain Structure and Function*, *222*(4), 1785–1795. https://doi.org/10.1007/s00429-016-1307-9

Pliatsikas, C., Johnstone, T., & Marinis, T. (2014a). fMRI evidence for the involvement of the procedural memory system in morphological processing of a second language. *PLoS ONE*, *9*(5). https://doi.org/10.1371/journal.pone.0097298

Pliatsikas, C., Johnstone, T., & Marinis, T. (2014b). Grey matter volume in the cerebellum is related to the processing of grammatical rules in a second language: A structural voxel-based morphometry study. *Cerebellum*, *13*(*1*), 55–63. https://doi.org/10.1007/s12311-013-0515-6

Pliatsikas, C., & Luk, G. (2016). Executive control in bilinguals: a concise review of fMRI studies. *Bilingualism: Language and Cognition*, *53*(*9*), 1689–1699. https://doi.org/10.1017/CBO9781107415324.004

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. https://doi.org/10.1073/pnas.1414183112

Rossi, E., Cheng, H., Kroll, J. F., Diaz, M. T., & Newman, S. D. (2017). Changes in White-Matter Connectivity in Late Second Language Learners: Evidence from Diffusion Tensor Imaging. *Frontiers in Psychology*, *8*(November), 1–15. https://doi.org/10.3389/fpsyg.2017.02040

Rothman, J. (2008). Why all counter-evidence to the critical period hypothesis in second language acquisition is not equal or problematic. *Language and Linguistics Compass*, *2*(*6*), 1063–1088. https://doi.org/10.1111/j.1749-818X.2008.00098.x

Schumann, J. H. (1990). The role of the amygdala as a mediator of affect and cognition in second language acquisition. In J. E. Alatis (Ed.), *Georgetown University Round Table on Language and Linguistics* (pp. 169–176). Washington, DC: Georgetown University Press. Retrieved from https://repository.library.georgetown.edu/bitstream/handle/10822/555483/GURT_1990.pdf?sequence=1#page=179

Schumann, J. H. (2001). Appraisal psychology, neurobiology, and langauge. *Annual Review of Applied Linguistics*, *21*, 23–44. https://doi.org/https://doi.org/10.1017/S0267190501000022

Selinker, L. (1972). Interlanguage. *International Review of Applied Linguistics*, *10*(*3*), 209–232.

Smith, S. M., Jenkinson, M., Johansen-Berg, H., Rueckert, D., Nichols, T. E., Mackay, C. E., Watkins, K. E., Ciccarelli, O., Cader, M. Z., Matthews, P. M., & Behrens, T. E. J. (2006). Tract-based spatial statistics: Voxelwise analysis of multi-subject diffusion data. *NeuroImage*, *31*(*4*), 1487–1505. https://doi.org/10.1016/j.neuroimage.2006.02.024

Smith, S. M., & Nichols, T. E. (2009). Threshold-free cluster enhancement: Addressing problems of smoothing, threshold dependence and localisation in cluster inference. *NeuroImage*, *44*(*1*), 83–98. https://doi.org/10.1016/j.neuroimage.2008.03.061

Spivey, M. J., & Marian, V. (1999). Cross Talk Between Native and Second Languages: Partial Activation of an Irrelevant Lexicon. *Psychological Science*, *10*(*3*), 281–284. https://doi.org/10.1111/1467-9280.00151

Stein, M., Federspiel, A., Koenig, T., Wirth, M., Strik, W., Wiest, R., Brandeis, D., & Dierks, T. (2012). Structural plasticity in the language system related to increased second language proficiency. *Cortex*, *48*(*4*), 458–465. https://doi.org/10.1016/j.cortex.2010.10.007

Stein, M., Winkler, C., Kaiser, A. C., & Dierks, T. (2014). Structural brain changes related to bilingualism: Does immersion make a difference? *Frontiers in Psychology*, *5*(SEP), 1–7. https://doi.org/10.3389/fpsyg.2014.01116

Stocco, A., & Prat, C. S. (2014). Bilingualism trains specific brain circuits involved in flexible rule selection and application. *Brain and Language*, *137*, 50–61. https://doi.org/10.1016/j.bandl.2014.07.005

Stocco, A., Yamasaki, B. L., Natalenko, R., & Prat, C. S. (2014). Bilingual brain training:. A neurobiological framework of how bilingual experience improves executive function. *International Journal of Bilingualism*, *18*(*1*), 67–92. https://doi.org/10.1177/1367006912456617

Thierry, G., & Wu, Y. J. (2007). Brain potentials reveal unconscious translation during foreign-language comprehension. *Proceedings of the National Academy of Sciences*, *104*(*30*), 12530–12535. https://doi.org/10.1073/pnas.0609927104

Ullman, M. T. (2004). Contributions of memory circuits to language: The declarative/procedural model. *Cognition*, *92*(*1–2*), 231–270. https://doi.org/10.1016/j.cognition.2003.10.008

Winkler, A. M., Ridgway, G. R., Webster, M. A., Smith, S. M., & Nichols, T. E. (2014). Permutation inference for the general linear model. *NeuroImage*, *92*, 381–397. https://doi.org/10.1016/j.neuroimage.2014.01.060

Wong, B., Yin, B., & Brien, B. O. (2016). Neurolinguistics:. Structure, Function, and Connectivity in the Bilingual Brain. *BioMed Research International*, *2016*.