Age and Proficiency in the Bilingual Brain Revisited: Activation Patterns Across Different L2-Learner Types

Tomasina M. Oh [1,2*†], Steven Graham [3], Philina Ng [1], Ing Berne Yeh [4], Bernard P. L. Chan [5] and Andrew M. Edwards [6]

[1] Department of English Language and Literature, National University of Singapore, Singapore, Singapore, [2] School of Sport, Health and Wellbeing, Plymouth Marjon University, Plymouth, United Kingdom, [3] Functional Brain Imaging Lab, Department of Psychology, A*STAR-NUS Clinical Imaging Research Centre, Yong Loo Lin School of Medicine, National University of Singapore, Singapore, Singapore, [4] Ng Teng Fong General Hospital, Jurong Health Services, Singapore, Singapore, [5] Division of Neurology, National University Hospital, Singapore, Singapore, [6] School of Human and Life Sciences, Canterbury Christ Church University, Canterbury, United Kingdom

Second language (L2) proficiency and age of acquisition (AoA) are important when examining L2 neural representation. However, they are potential confounds of each other and their relative importance remains unclear. How L2 was learnt (informally vs. formally at school) is another complication that potentially contributes further to this lack of clarity. Careful control of these variables is crucial for teasing apart their effects, yet almost all previous neuroimaging studies have studied one or the other in isolation. Thirty five participants of varying proficiency and AoA were scanned using fMRI while performing an English (L2) past tense task; all were formal L2 learners. Early high proficiency bilinguals (EAHP) were contrasted with late high proficiency (LAHP) in three conditions: (i) regular inflection; (ii) irregular inflection; and (iii) regularity × AoA. In line with previous findings, LAHP (vs. EAHP) bilinguals showed more extensive activation across multiple regions for both regular and irregular inflection. The left inferior frontal gyrus (IFG; BA47) was one region that showed significant activation in condition (iii). EAHPs engaged this region selectively for regular but not irregular inflection while LAHPs activated it during both types of inflection. Late high and late low proficiency (LALP) bilinguals were also contrasted in three conditions: (i) regular inflection; (ii) irregular inflection; and (iii) regularity × proficiency. In all regions showing significant differences, LAHPs showed greater activation relative to LALPs (regular and irregular conditions). In the regularity × proficiency condition the left IFG was also a significantly activated region. Previous studies suggest this region is positively associated with high proficiency but this has not always been replicated. LAHPs showed increased activation in BA45 but not BA44, suggesting L2 is a controlled rather than automatic process in this group despite being highly proficient. Our study suggests AoA and proficiency *both* influence bilingual brain activation independently, an important replication given only two other neuroimaging studies have experimentally manipulated both variables within the same study. We also provide evidence for how different AoA influences left IFG engagement during L2 processing, and for the hypothesis that BA45 is associated with high proficiency when degree of automaticity is lower.

Keywords: bilingual brain, second language, age of acquisition, age of L2 learning, proficiency, mode of learning, fMRI, automaticity

INTRODUCTION

In the study of the bilingual brain it has been shown that activated regions underlying a bilingual's first (L1) and second (L2) languages do not always overlap (e.g., Kim et al., 1997) and attempts have been made to identify the factors that contribute to these differences. L2 age of learning/acquisition (AoA) and L2 proficiency are two factors that have received close attention over the last 20 years. However, despite this close scrutiny we are actually not that much closer to knowing exactly how each factor influences the bilingual brain (Watkins et al., 2017). A specific question that remains unanswered is which of the two is more crucial in determining the pattern of brain activation during L2 processing, and related to that is how each factor influences functional activation. Addressing such questions has a potential impact on larger theoretical questions such as those about neural plasticity and the effects of experience as well as degree of automaticity on the bilingual brain (Li et al., 2014; Nichols and Joanisse, 2016; Vissiennon et al., 2017).

One difficulty when examining either L2 AoA or proficiency is that the two are closely intertwined and therefore possible confounds of each other. Further complicating the picture is that AoA may be confounded with the way L2 was learnt, referred to as either "modality of learning" (Fabbro, 2001) or "mode of learning" (Marrero et al., 2002). Many early L2 bilinguals learn their L2 informally or via immersion while late bilinguals tend to learn in more formal, non-immersion environments. But this is obviously not always the case. Thus, many previous studies may have thought they were studying a homogenous group of early or late bilinguals but it is possible that there were hidden differences related to mode of learning within each group. Previously reported differences could therefore have been a reflection of mode of learning rather than AoA *per se*.

In order to avoid the confound between AoA and proficiency, previous neuroimaging studies have usually held one of these constant while the other was examined, i.e., they examined only one of the factors in isolation. Perani et al. (1998) were one of the first groups to do this, using positron emission tomography (PET). An earlier study of theirs (Perani et al., 1996) had shown that auditory processing of stories in the L1 (vs. L2) of late acquisition (age >7 years)/low (L2) proficiency Italian-English bilinguals caused more extensive activation in the temporal lobes and temporoparietal cortex. But these findings were difficult to interpret because both AoA and proficiency were different in each language—L1 had an early AoA plus high proficiency while L2 had a late AoA plus low proficiency. When Perani et al. (1998) addressed this issue by holding proficiency constant in their subsequent study, they found no differences in activation between the two groups. Because previously found activation differences disappeared once proficiency was equated, they concluded that L2 proficiency rather than AoA determines differences in L2 brain activation.

The debate has continued for years since, not helped—as Waldron and Hernandez (2013) point out—by the fact that few (neuroimaging) studies examining L2 AoA have adequately controlled for L2 proficiency properly. Some studies have found support for Perani et al.'s (1998) conclusion on proficiency

being the determining factor (e.g., Frenck-Mestre et al., 2005; Hesling et al., 2012). Others have not found such support; AoA effects have been demonstrated even when proficiency was held constant (e.g., Saur et al., 2009; Consonni et al., 2013; Archila-Suerte et al., 2015; Hernandez et al., 2015). A smaller number have argued that both play a role but affect the brain differently, for example proposing that AoA influences regions involved in grammatical processing while proficiency affects those used in semantic processing (Wartenburger et al., 2003). A recent meta-analysis of neuroimaging studies spanning the period 1998–2014 (Liu and Cao, 2016) examined how L2 AoA affects L1 vs. L2 networks and reported that late (vs. early) bilinguals rely on additional regions when processing L2 compared to their L1. However, while this meta-analysis confirmed AoA does affect the bilingual brain it focused exclusively on studies that recruited high proficiency bilinguals. Papers with low or moderate proficiency bilinguals were excluded, making it difficult to draw any firm conclusions about L2 proficiency effects.

Interestingly, very few imaging studies have manipulated *both* L2 AoA and proficiency in the same study—as described above they have tended to manipulate one factor while keeping the other constant. Two notable exceptions are Wartenburger et al. (2003) and Nichols and Joanisse (2016). Wartenburger et al. (2003) experimentally manipulated the two factors by comparing three groups of German-Italian bilinguals differing in their AoA (early/late) and proficiency (high/low): an early acquisition, high proficiency group (EAHP; $n = 11$), a late acquisition, high proficiency group (LAHP; $n = 12$) and a late acquisition, low proficiency group (LAHP; $n = 9$). The early bilinguals had been exposed to their L2 since birth, while mean AoA for late bilinguals was 18.9 years (LAHP) and 20.4 (LALP) years, respectively. Participants were scanned using functional magnetic resonance imaging (fMRI) while performing grammatical and semantic judgements on presented sentences. Wartenburger et al. (2003) found that L2 AoA was more important in determining brain regions involved in L2 grammatical processing, while L2 proficiency was more important in determining those involved in L2 semantic processing. This finding that grammatical processing might be influenced differently to semantic processing is perhaps not so surprising; as Wartenburger et al. (2003) point out, findings in other domains (monolingual electrophysiological, lesion and functional neuroimaging studies) have led to the idea that grammar is acquired incidentally and implicitly while lexical-semantic processing is carried out by the explicit memory and knowledge system (Paradis, 1994; Ullman, 2001, 2004, 2005, 2016; Lebrun, 2002; Hernandez and Li, 2007; Abutalebi, 2008). But what is significant about Wartenburger et al.'s (2003) study is that by directly manipulating each factor it was able to show what the respective contributions of AoA and proficiency might be.

Nichols and Joanisse's (2016) study is the other to have tested the independent effects of both L2 AoA and proficiency on L1 and L2 neural activation. Twenty two Mandarin-English bilinguals of varying L2 AoA and proficiency were scanned while performing a lexical-semantic task (picture-word matching); both functional and structural data were collected. The bilinguals in this study were shown to have a weak relationship between

AoA and proficiency. This allowed the researchers to take a different approach to Wartenburger et al. (2003), in that L2 AoA and proficiency were treated as continuous variables. Like Wartenburger et al. (2003), Nichols and Joanisse (2016) also found both L2 AoA and proficiency effects although in this study AoA effects were found in the realm of semantic processing. AoA was found to uniquely predict activity differences in seven areas that included the right parahippocampal gyrus, bilateral superior temporal gyrus (STG) and bilateral inferior frontal gyrus while proficiency was uniquely associated with activation in the left parahippocampal gyrus and right cingulate (Nichols and Joanisse, 2016).

That there are only two studies that have specifically looked at both factors is surprising given the potential confound that exists between the two and the current lack of consistent findings among neuroimaging studies on bilinguals (see also Wong et al., 2016 for a review). The importance of replication in neuroimaging has recently been highlighted (Evans, 2017), particularly given that the high cost of data collection often leads to smaller sample sizes. Both direct and conceptual replication studies have a crucial role to play in increasing the confidence in neuroimaging findings (Open Science Collaboration, 2015), with the latter referring to a study in which a previous result or hypothesis is tested with different methods (Schmidt, 2009; Evans, 2017).

The current study was conducted to examine (i) the independent contributions of L2 AoA and proficiency to L2 neural representation, and (ii) the nature of those contributions and what light it might shed on questions about the L2 learning such as plasticity and automaticity of processing. Given the first aim, our study could also be viewed as an attempt to conceptually replicate the findings of Wartenburger et al. (2003) and Nichols and Joanisse (2016), and contribute to neuroimaging data on L2 AoA and proficiency. Following Wartenburger et al. (2003) *both* L2 AoA and proficiency levels in our study were experimentally manipulated. Mode of learning was kept constant across all participants in recognition of its status as a potential confound. We used a past-tense task to ensure task uniformity across different language processes (computational for regular verbs vs. lexical access for irregular verbs; see Ullman, 2001) and to keep processing to a single-word level (see section Stimuli for further explanation).

MATERIALS AND METHODS

Participants

Thirty five bilingual Mandarin-English participants, all of whom learnt English at school (i.e., formal setting) were assigned to one of three groups: Early Acquisition High Proficiency (EAHP; $n = 12$), Late Acquisition High Proficiency (LAHP; $n = 12$) or Late Acquisition Low Proficiency (LALP; $n = 11$). While there are advantages to Nichols and Joanisse's (2016) approach of treating AoA and proficiency as continuous variables we chose not to do this given we could not guarantee the decoupling of L2 AoA and proficiency in our sample in the way they were able to. The study received ethical approval from the local Institutional Review Board (Domain Specific Review Board

(A), National Healthcare Group, National University Hospital, Singapore) and all individuals provided written informed consent prior to participation.

Following previous neuroimaging studies (Perani et al., 2003; Meisel, 2004; Waldron and Hernandez, 2013; Klein et al., 2014) the cut-off for early bilinguals was set at age 7, although we do acknowledge that there is disagreement over what this cut-off age ought to be and in some cases debate over the traditional notion of the cut-off age itself. L2 (English) proficiency was determined via the Self Report Classification Tool (Lim et al., 2008), which has been validated on 198 undergraduates studying at a Singapore university and was chosen because it best reflected our bilingual context; our participants were also all students at the National University of Singapore. Speaking and listening proficiency ratings in the questionnaire were used for the purpose of this study; in a small number of cases where the two were different from each other the average of the two ratings were taken. Li et al. (2019) and Tomoschuk et al. (2018) have discussed the downside of relying purely on self-reported proficiency. To minimize subjectivity in ratings and to help our participants identify which rating best suited them, all participants were provided with a detailed guide explaining with examples what they should be able to do in their L2 for each rating. Research assistants were also on hand to address any doubts participants had. Participants' self-ratings were consistent with their General Certificate of Education (GCE) Ordinary Level English examination grades, which we also recorded as part of each participant's profile. The GCE paper examines written and oral production and comprehension.

EAHP and LAHP bilinguals were matched on chronological age and proficiency (**Table 1**).

The two groups differed significantly in their mean L2 AoA; mean AoA for EAHP was 4.75 years vs. 9.38 years for LAHP ($p = 0.002$).

LAHP and LALP bilinguals were matched on chronological age and L2 AoA (**Table 1**) but were significantly different in proficiency (LAHP mean proficiency rating = 6.17; maximum score = 7); LALP mean proficiency rating = 4.00, $p < 0.001$).

All participants were right-handed and had no history of neurological damage.

TABLE 1 | Participant characteristics.

	EAHP	LAHP	LALP
n (female, male)	12 (4, 8)	12 (3, 9)	11 (5, 6)
Mean age (s.d)	23.5 (2.19)	25.91 (9.38)*	25.64 (9.23)**
Mean AoA (s.d.)	4.75 (1.14)	9.38 (4.16)	9.59 (1.74)
Mean proficiency rating (s.d.) [1 = Low; 7 = High]	6.25 (0.72)	6.21 (0.72)	4.00 (0.37)
Place of learning	School	School	School

*EAHP, Early high proficiency bilinguals; LAHP, Late high proficiency bilinguals; LALP, Late low proficiency bilinguals. *All but two of the participants in this group were in their early 20's; one participant was 53 years old, the other 35. **All participants were in their early to mid 20's except for one, who was also 53.*

Experimental Paradigm

While lying in the MRI scanner (Siemens (Erlangen, Germany) 1.5T Symphony MRI scanner), participants were presented with the present tense of a verb (regular or irregular). Participants were told to think of the past tense form but not answer aloud until they saw a row of fixation crosses ("++++++"), which appeared during the following 2 s silent period. Participants produced their answers aloud during this silent period (i.e., overt rather than covert generation) in order to minimize artifacts associated with head movement.

Stimuli were presented using Presentation software (version 0.60, Neurobehavioural Systems, USA). Lowercase words appeared for 3 s in white Arial font (size 60) against a black background. Participants received the same oral instructions at the beginning of the experiment not to speak until the fixation crosses were presented, in accordance with our previous study on monolingual speakers (Oh et al., 2011). Participants were also reminded at the start of each run to provide the past tense form of the presented words aloud during the silent period.

Responses were audio recorded by one of the experimenters using an Olympus digital audio recorder (model DM20), and then transcribed. Any errors were noted and excluded from the imaging analysis.

Participants were given a familiarization session outside the scanner to familiarize them with the experiment. The 20 verbs used in the familiarization session were not used in the actual scanning session.

Stimuli

The stimuli in this experiment were mainly based on those used by Oh et al. (2011). Stimuli were made up of 40 regular and 40 irregular verbs (total $n = 80$). Past tense inflection is one way to examine both lexical access (via irregular inflection, thought to tap into semantic processes) and "grammatical" or morphological processing (via regular inflection, thought to reflect computational processes; see Ullman, 2001, 2004, 2005, 2016). In other words, past tense inflection allowed us to examine two different processes in the way Wartenburger et al. (2003) did, but at the single-word level and within a single task. It has been pointed out that the interpretation of neuroimaging data becomes more difficult as complexity of the task increases

(Fabbro, 2001). Thus, using this past tense task helped avoid the interpretation problem that comes with more complex sentence-level processing tasks or comparing activation across different tasks (see Fabbro, 2001).

Verbs were matched (pair-wise) for log frequency (of the past tense form), number of phonemes and phonological complexity. The latter was defined in terms of Consonant-Vowel (CV) structure (see Bird, 2003), where a CCVCC structure (as in the word "trust") would be more complex than a CVC structure present in a word such as "sit." Frequency of words was determined according to the Singapore International Corpus of English (ICE) database (http://ice-corpora.net/ice/index.html), which is part of the International Corpus of English (Greenbaum, 1996). There were four runs, each consisting of 10 regular and 10 irregular verbs randomly presented. The two groups of 40 verbs were randomly assigned to each of the four runs. Please see **Appendix A** for the full list of verbs.

Image Acquisition and Data Analysis

Images were acquired using blipped gradient-echo planar imaging (flip angle $= 90°$; 64×64 pixel matrix; FOV $= 192 \times 192$ mm). Acquisition time was 3,000 ms, followed by 2,000 ms of silence. During acquisition, participants saw a verb and were instructed to think of the past tense of the verb in silence. Once the word was replaced by a series of crosses "+++++" (i.e., during the silent period), participants provided their answers aloud. Time to repetition (TR) was therefore 5,000 ms.

Seventy two images (32 oblique axial 3-mm slices with 0.3 mm gap, descending interleaved) were collected per run, depending on the randomly inserted (for jittering purposes) baseline period where no words appeared; only a "++++++" was seen on the screen during these baseline periods. The periods could last 5,000, 10,000, or 15,000 ms. In each run there was a two-in-six chance of a 5,000 ms period appearing after a trial, a three-in-six chance of a 10,000 ms period appearing and a one-in-six chance of a 15,000 ms period appearing. The average length of a run was 355 s (minimum 350 s, maximum 360 s). In addition, 50 s and at least 30 s of fixation were presented at the beginning and end of each run, respectively, to ensure signal homogeneity (the first four acquisitions were ignored) and sufficient baseline periods.

Brain Voyager QX (version 2.3, Brain Innovation, Holland) was used to analyze images. Slice scan time correction, motion correction, spatial smoothing (8 mm FWHM) and linear trend removal were applied to these functional images. These were then registered to the MPRAGE (magnetization prepared rapid acquisition gradient echo) images, and the realigned data then transformed into Talairach space. A random effects multi subject general linear model (GLM) was then computed. This hierarchical analysis entailed a first level analysis in which all experimental conditions for each subject were modeled as separate regressors. Each regressor was convolved with a canonical haemodynamic response function (HRF) peaking 5 s after onset of word presentation (Henson and Friston, 2007). The resulting GLM thus contained 3 regressors per subject: GenReg (generation of regular past tense verbs), GenIrreg (generation of irregular past tense verbs) and Others (for errors). Each regressor was then analyzed at a second level using separate

TABLE 2 | Behavioral results, participants producing the past tense of presented verbs.

Group	Mean % accuracy (s.d.)		
	Overall (combined regular and irregular verbs)	Regular verbs (computational/ grammar task)	Irregular verbs (lexical access/ semantic task)
	$n = 80$	$n = 40$	$n = 40$
EAHP	90.94 (4.86)	97.29 (2.91)	84.58 (9.40)
LAHP	90.10 (7.64)	97.71 (2.71)	82.50 (14.89)
LALP	78.98 (12.17)*	92.95 (8.28)	65.00 (19.24)*

*EAHP, Early high proficiency bilinguals; LAHP, Late high proficiency bilinguals; LALP, Late low proficiency bilinguals. *Significant difference (p < 0.05).*

group-level random-effects t-tests: EAHP vs. LAHP, as well as LAHP vs. LALP. The resultant group-level statistical parametric t-maps were corrected for multiple comparisons using cluster-size thresholding, described below (Forman et al., 1995; Goebel et al., 2006).

Each map was initially thresholded at a voxel-wise p-value ($p < 0.01$, uncorrected) that yielded distinct segregated regions of interest (ROIs). These maps were then subjected to a whole brain (no mask) correction criterion based on the estimate of the map's spatial smoothness (the FWHM was estimated by BVQX to be 1.417 in native voxel resolution for all contrasts) and 1,000 iterations of Monte Carlo simulation to determine the minimum cluster size threshold. The thresholds determined for each contrast are listed, respectively, in the appropriate tables (**Tables 3**, **4**). These cluster-size thresholds were then applied to the group-level statistical t-maps to yield a corrected 5% false positive rate. We selected any voxels that were activated above the indicated threshold ($p < 0.05$, corrected) and reported the peak for each significantly activated cluster (**Tables 3**, **4**). In addition, z-normalized regressor values (averaged across all ROI voxels) for each condition for each group were extracted and further interrogated by plotting each regressor relative to the fixation baseline (zero). The inclusion of a fixation baseline also allowed the estimation of HRF predictors for each of these conditions of interest for each group of participants.

Between-group comparisons for behavioral data were made using independent sample t-tests as appropriate. Within-group comparisons were evaluated using paired t-tests. Statistical significance was accepted at $p < 0.05$.

RESULTS

Behavioral Results

Table 2 shows the mean accuracy rate and standard deviation for each of the three groups in the past tense task. There were no differences in performance between the early and late high proficiency bilinguals. However, low proficiency bilinguals made significantly more errors in the tasks. Given that participants had to wait till the 2,000 ms silent period (after the word was presented for 3,000 ms) to provide their answer in the scanner, no reaction time data are provided here.

No significant difference in overall accuracy rate was found between the two high proficiency groups (EAHP = 90.94% and LAHP = 90.10%), despite one group learning their L2 significantly later than the other. The LAHP group performed significantly better overall than the LALP group [$t_{(22)} = 2.54$; $p = 0.02$]. This difference was attributable to the performance difference in irregular rather than regular verbs. Analysis showed that LAHP performed significantly better on irregular verbs relative to LALP [$t_{(22)} = 2.40$; $p = 0.03$] but not on regular verbs ($p > 0.05$).

All participants performed significantly better when inflecting regular verbs [EAHP [$t_{(11)} = 4.42$; $p = 0.001$]; LAHP [$t_{(11)} = 3.52$; $p = 0.005$]; LALP [$t_{(10)} = 5.49$; $p = 0.0003$] as compared to irregular verbs. Regular and irregular verbs were controlled for word frequency and phonological complexity and so this finding

TABLE 3 | Regions activated by the EAHP vs. LAHP groups in the (1) Generate Regular Past Tense, (2) Generate Irregular Past Tense, and (3) Contrast between Irregular and Regular Tense generation tasks, respectively.

Regions	Brodmann Area	Talairach (x, y, z)			Cluster size (mm³)
		x	y	z	
1. Generation of Regular Verbs					
Threshold 717 mm³ (p < 0.05, corrected)					
1a. LAHP > EAHP					
L post central gyrus	BA3	−51	−16	28	795
L post central gyrus	BA3	−38	−20	46	794
R middle frontal gyrus	BA9	24	34	27	1,752
R medial frontal gyrus	BA6	6	−7	57	805
2. Generation of Irregular Verbs					
Threshold 789 mm³ (p < 0.05, corrected)					
2a. LAHP > EAHP					
L inferior frontal gyrus	BA47	−39	17	−3	1,645
L & R medial frontal gyrus	BA6	±1	−22	57	2,233
L inferior parietal lobe	BA40	−54	−36	39	1,921
L supramarginal gyrus	BA40	−45	−43	33	1,919
L & R precuneus	BA7	±5	−54	39	995
R inferior frontal gyrus	BA47	42	41	1	2,130
R middle frontal gyrus	BA8/9	29	37	32	1,478
R middle frontal gyrus	BA9	45	25	32	1,101
R insula	BA13	34	16	1	830
R medial frontal gyrus	BA6	8	−4	65	812
R post central gyrus	BA3	47	−17	39	1,213
R superior temporal gyrus	BA42	61	−28	12	1,008
3. Interaction (Regularity × AoA)					
Threshold 598 mm³ (p < 0.05, corrected)					
L inferior frontal gyrus	BA47	−40	16	−4	906
L calcarine sulcus	BA17	−21	−75	16	888
L lingual gyrus	BA18	−18	−76	−8	645

EAHP, Early high proficiency bilinguals; LAHP, Late high proficiency bilinguals; L, Left; R, Right.

is not attributable to regular verbs being more familiar or to them being less phonologically complex.

Errors in irregular past tense forms tended to be over-regularizations (e.g., providing "slided" as the past tense for "slide") or, in a very small number of cases the wrong tense (e.g., providing the perfect past instead of simple past: "shrunk" for "shrink"). In a small number of other cases (for both regular and irregular verbs), no response was made. Errors and trials on which no response was made were modeled as a dummy variable for the imaging analysis.

Imaging Results

Our imaging findings show that both L2 AoA and proficiency influence brain activation, independently of each other. How

TABLE 4 | Regions activated by the LAHP vs. LALP groups in the (1) Generate Regular Past Tense, (2) Generate Irregular Past Tense, and (3) Contrast between Irregular and Regular Tense generation tasks, respectively.

Regions	Brodmann area	Talairach (x, y, z)			Cluster size (mm^3)
		x	**y**	**z**	
1. Generation of Regular Verbs					
Threshold 781 mm^3 (p < 0.05, corrected)					
1a. LAHP > LALP					
R superior frontal gyrus	BA9	17	43	29	1,177
R superior frontal gyrus	BA8	13	36	40	1,184
R inferior frontal gyrus	BA44/45	32	28	8	1,149
R medial frontal gyrus	BA6	4	−6	56	901
R middle frontal gyrus	BA6	26	−9	51	901
2. Generation of Irregular Verbs					
Threshold 991 mm^3 (p < 0.05, corrected)					
2a. LAHP > LALP					
L superior frontal gyrus	BA9	−8	45	30	1,311
L medial frontal gyrus	BA9	−13	43	27	1,311
L inferior frontal gyrus	BA45	−49	21	19	1,171
L cingulate	BA32	−13	21	35	3,290
L cingulate	BA24	−11	13	27	3,335
L post central Gyrus	BA3	−55	−19	32	3,500
L cingulate	BA31	−5	−35	32	2,495
L precuneus	BA7	−2	−53	41	5,863
L precuneus	BA7	−21	−60	32	1,611
L cerebellum		−12	−70	−17	1,178
R superior frontal gyrus	BA9	14	45	28	2,116
R medial frontal gyrus	BA8	6	35	42	2,582
R anterior cingulate		10	29	6	1,738
R post central gyrus	BA3	46	−19	39	5,510
R transverse temporal gyrus	BA41	33	−27	11	1,238
R inferior parietal lobule	BA40	43	−31	24	1,559
R superior parietal lobule	BA7	23	−67	33	1,178
3. Interaction (Regularity × Proficiency)					
Threshold 745 mm^3 (p < 0.05, corrected)					
L middle frontal gyrus	BA9/46	−35	31	25	1,448
L cingulate	BA32	−9	17	30	1,603
L inferior frontal gyrus	BA44	−50	15	25	1,303
L middle frontal gyrus	BA6	−57	4	18	1,341
L cingulate	BA24	−2	3	44	2,461
L superior temporal gyrus	BA22	−60	−18	6	1,316
L precuneus	BA7	−1	−57	47	1,797

LAHP, Late high proficiency bilinguals; LALP, Late low proficiency bilinguals; L, Left; R, Right.

each of the factors affects neural activation is reported below; the first section (EAHP vs. LAHP: How Does L2 AoA Influence Neural Activation in the Bilingual Brain?) describes how early

vs. late L2 AoA in our participants altered neural activation; the effects of high vs. low L2 proficiency are reported in the following section.

EAHP vs. LAHP: How Does L2 AoA Influence Neural Activation in the Bilingual Brain?

To examine AoA effects, we compared brain regions activated by EAHP vs. LAHP bilinguals in three conditions: (i) during regular inflection (see Part 1 in **Table 3**); (ii) during irregular inflection (Part 2 in **Table 3**) and (iii) regularity × AoA (Part 3 in **Table 3**). A whole brain analysis was applied and significantly activated regions in each of these conditions are reported in **Table 3**. Only those regions significantly activated above threshold for cluster size (at a corrected $p < 0.05$) are reported. Regions are arranged anterior to posterior and left regions are reported first.

Significantly greater activation was found for LAHP bilinguals relative to EAHP bilinguals in both the regular and irregular conditions, respectively (1a and 2a in **Table 3**). EAHP bilinguals did not differ significantly from baseline in either condition and so no significantly activated regions are reported for this group. LAHP bilinguals showed activation across the left and right hemispheres for both regular and irregular inflection. During regular past tense generation, the LAHP group (vs. EAHP) showed greater activation in the left post central (2 regions), right middle and medial frontal gyri. During irregular past tense generation LAHPs showed greater activation (vs. EAHP) in the left and right inferior frontal gyrus (BA47), medial frontal gyrus (BA6), left inferior parietal lobe (BA40) and right middle frontal gyrus (BA9).

In the third condition (regularity × AoA), three regions including the left IFG (BA47) were significantly activated (see Part 3 **Table 3**). The frontal cortex has been acknowledged for its important role in language processing (see Li et al., 2014). We interrogated voxels at the peak of this frontal region to examine the nature of the interaction (**Figure 1**). The pattern of activation shown in **Figure 1** suggests that LAHPs engage left BA47 as much for both regular and irregular generation (computational process and lexical access) while EAHPs activate this region for regular inflection but not for irregular inflection. When the peak voxels of other significantly activated regions were interrogated, a similar pattern was observed in other frontal regions such as the right medial frontal gyrus (BA6; 8, −4, 65; **Figure 2**), right IFG (BA47; 42, 41, 1), the right MFG (BA 9; 45, 25, 32), and left insula (−37, 16, 0).

LAHP vs. LALP: How Does L2 Proficiency Affect Neural Activation in the Bilingual Brain?

Table 4 identifies regions found during (i) regular inflection only (Part 1); (ii) irregular inflection only (Part 2); and (iii) regularity × proficiency (Part 3). As in section "EAHP vs. LAHP: How Does L2 AoA Influence Neural Activation in the Bilingual Brain?" a whole brain analysis was applied and only regions significantly activated above cluster threshold size ($p < 0.05$, corrected) in each condition were reported.

Significant differences were found between LAHP and LALP bilinguals in several brain regions for both regular and irregular conditions. In each of these regions, LAHPs showed greater
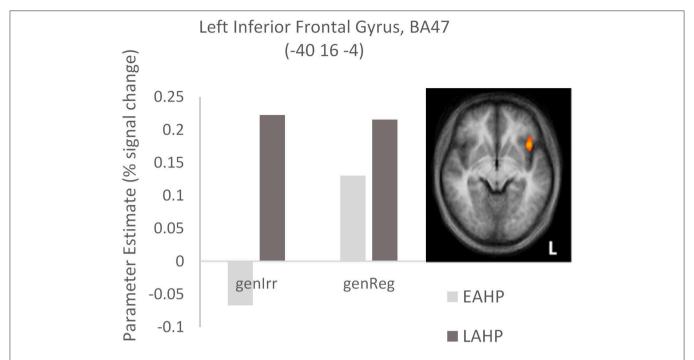
FIGURE 1 | Activation differences in BA47 between the early (EAHP) vs. late (LAHP) bilinguals for the Regularity × AoA condition. Significant activation is attributed to: (i) the difference between EAHP and LAHP for irregular inflection (left of the graph) and (ii) the difference between regular and irregular inflection for the EAHP bilinguals (see the lighter bars). Activated regions are those activated above the indicated threshold ($p < 0.05$, corrected).
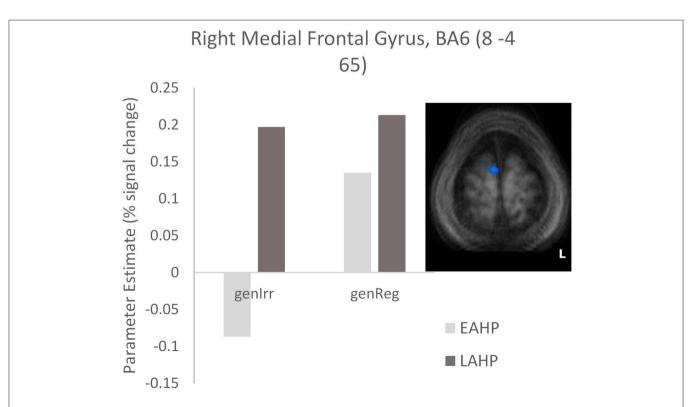

FIGURE 2 | Activation differences in BA6 between EAHP and LAHP bilinguals, during the generation of irregular past tense verbs (EAHP vs. LAHP). Activated regions are those activated above the indicated threshold ($p < 0.05$, corrected).

activation compared to LALPs (1a and 2a, **Table 4**). When compared to the LALP group, LAHP showed extensive right frontal activation (superior BA8; superior BA9; inferior BA 44/45; medial BA6; middle BA6) for regular inflection. Activation was more bilateral for irregular inflection; in the left hemisphere activation was found in frontal regions (superior BA9; medial BA9, inferior BA45; see **Figure 3**), several left cingulate regions (BA24; BA 31; BA32), two regions in the left precuneus and the left cerebellum. Right frontal regions were also seen (superior BA9; medial BA8), together with right parietal (inferior BA41; superior BA40), post central (BA3) and transverse temporal gyrus (BA41).

Finally, interactions were examined to ascertain whether or not there was a significant difference across high and low proficiency bilinguals when activation differences between regular and irregular generation in each group were compared (Part 3, **Table 4**). Significantly activated regions were mainly in the left hemisphere, including the left inferior frontal gyrus (BA44), the left middle frontal gyrus (BA9/46) and the left superior temporal gyrus (BA22). As before we interrogated the voxels at the peak in these areas to examine the nature of the interaction. We found that high proficiency bilinguals barely engaged the left IFG (BA44; see **Figure 4**) during regular or irregular inflection; in fact, this group demonstrated a small amount of deactivation in both conditions. Low proficiency bilinguals on the other hand deactivated this region for irregular generation but activated it for regular generation; this difference was significantly different ($p < 0.05$, corrected) from the non-difference in the LAHP group.

DISCUSSION

This study set out to tease apart the effects of L2 AoA and proficiency on neural activation in bilinguals. We scanned three groups of bilinguals—each with a different combination of L2 AoA and proficiency but all matched for mode of L2 learning—while they generated English (L2) past tense verbs in the scanner. We experimentally manipulated these two variables so that each could be examined in isolation within the same study. We found that both AoA and proficiency influence L2 neural activation independently, replicating the finding in similarly designed studies that both factors play equally crucial roles in the bilingual brain (Wartenburger et al., 2003; Nichols and Joanisse, 2016). We discuss our main findings related to L2 AoA and L2 proficiency effects below, beginning with the former.

Firstly, our study showed that late bilinguals produce greater overall activation compared to early bilinguals when generating both regular and irregular past tense in their L2; this effect was independent of L2 proficiency. This result is consistent with Liu and Cao's (2016) meta-analysis of previous neuroimaging studies, which concluded that late high proficiency bilinguals show greater overall activation than early high proficiency bilinguals when processing their L2.

Our finding of significantly greater activation for late vs. early bilinguals during regular verb inflection (previously argued to reflect computational/grammatical processing; Oh et al., 2011; Ullman, 2016) also agrees with Wartenburger et al.'s (2003) study which showed the same pattern when participants performed a grammatical judgement task. Interestingly, our findings differ in the realm of "semantic processing" (i.e., irregular verb inflection in our study). Wartenburger et al.'s (2003) reported that EAHP and LAHP groups did not differ in activation during the semantic judgement task and concluded there are no AoA effects for semantic processing. Our study however suggests that AoA effects are present across extensive brain regions during the production of irregular past tense verbs, which some models interpret as being a semantic or lexical access task (e.g., Ullman's DP model, 2001, 2004). In support of our findings, we note that Nichols and Joanisse (2016) who used a semantic-lexical task similarly reported increased L2 activity relative to L1 in bilateral IFG as a function of later AoA. It is possible however that the difference between our studies and Wartenburger's may be due to task differences: while we used single-word stimuli (Nichols and Joanisse's task was a picture-word matching task), Wartenburger et al. (2003) used a semantic judgement task involving sentences and therefore more processes in addition to lexical retrieval. Perhaps it is the case when sentence-level processing is measured, AoA and proficiency begin to play more specialized roles such as AoA affecting grammatical processing and proficiency affecting semantic processing. This is a line of investigation that requires further exploration.

There are at least two extant explanations as to why activation differences may exist between early and late language learning. One is to do with automaticity, which has been associated with decreased activity in the inferior frontal cortices and right middle frontal gyrus (Poldrack et al., 2005). Another proposed reason for this observed difference in neural activation is that it reflects maturational changes in neural plasticity when learning of language happens later, as compared to first language acquisition (Perani et al., 1996; Weber-Fox and Neville, 1996; Wartenburger et al., 2003; Mechelli et al., 2004; Abutalebi, 2008; Pakulak and Neville, 2011; Klein et al., 2014; Li et al., 2014). We speculate that our early bilinguals had a greater degree of L2 automaticity by virtue of having learnt it early in life and possibly because the learning coincided with a time of greater plasticity for language learning. We return to this question of L2 automaticity below.

Our second finding was that late bilinguals who are highly proficient are also distinguishable from their early counterparts by how they engage with the left prefrontal cortex (PFC) during irregular past tense inflection. As reported above we observed a significant interaction between AoA and regularity (i.e., regular vs. irregular inflection) in the left IFG (BA47). Interrogation of peak voxels revealed that while both groups of bilinguals engaged the left IFG during regular inflection, only the late bilinguals continued to engage this region during irregular inflection (**Figure 1**). This was a recurring pattern in other significantly activated frontal regions as well (see EAHP vs. LAHP: How Does L2 AoA Influence Neural Activation in the Bilingual Brain?; **Figure 2**). We suggest that early high proficiency bilinguals use frontal regions in a more task-specific way, i.e., there is selective activation depending on whether they are inflecting regular (computational process) or irregular (lexical retrieval) verbs. This involvement of the left IFG in regular but not irregular verb inflection for early L2 learners is consistent with the Declarative Procedural model's prediction that early L2 grammar is similar to L1 grammar acquisition
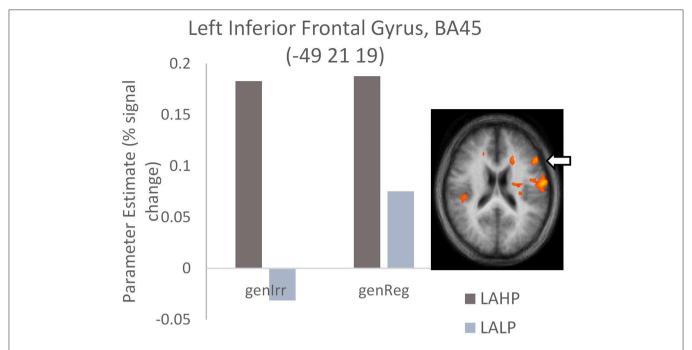
FIGURE 3 | High proficiency bilinguals (LAHP) showing increased activation relative to low proficiency bilinguals (LALP) in BA45 during the generation of irregular past tense verbs (LAHP vs. LALP). Activated regions are those activated above the indicated threshold ($p < 0.05$, corrected).
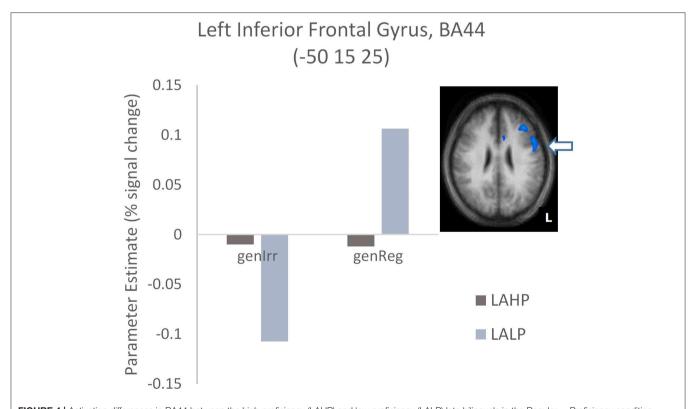

FIGURE 4 | Activation differences in BA44 between the high proficiency (LAHP) and low proficiency (LALP) late bilinguals in the Regular × Proficiency condition. Activated regions are those activated above the indicated threshold ($p < 0.05$, corrected).

(Ullman, 2001, 2004, 2005). We have previously shown that the left IFG (BA47; −31, 24, 3) was significantly more activated when monolingual native English speakers were generating regular (compared to irregular) past tense verbs (Oh et al., 2011). The pattern shown by our early bilinguals in the current study is therefore similar to that shown by the L1 English speakers in that study.

Our third finding is with regard L2 proficiency effects: we found that L2 proficiency affects neural activation, independently of the age our participants began learning their L2. Late L2 high proficiency bilinguals (LAHP) produced greater overall activation when compared to late L2 low proficiency bilinguals (LALP), in both the regular and irregular generation conditions. LAHP bilinguals activated right frontal areas (superior, middle, medial and inferior frontal) more than their LALP counterparts while generating regular past tense verbs. When generating irregular past tense verbs, LAHP (>LALP) bilinguals showed more bilateral activation (especially for superior frontal and post central gyri); regions such as left IFG (BA 45) were found too (**Table 4**; **Figure 3**). This finding fits with other neuroimaging studies, both structural and functional, that have shown increased volume or activation in the left inferior frontal gyrus with better proficiency. For example, various structural studies (Mårtensson et al., 2012; Stein et al., 2012, 2014) have reported that changes in L2 proficiency are positively related with volume increase in the left IFG. Functionally, activation in the left IFG has been seen to increase with increased proficiency, for example in the learning of language-like rules in an artificial grammar (Opitz and Friederici, 2003). Other studies have likewise argued for either decreased activation for less fluent bilinguals or increased activation in some regions to be positively correlated with increased proficiency, possibly in earlier stages of L2 learning (Hasegawa et al., 2002; Sakai et al., 2004; Xue et al., 2004; although see Saur et al., 2009 for a different view; Li et al., 2014; Nichols and Joanisse, 2016). One of these studies (Sakai et al., 2004) also used the past tense task to examine the functional role of the IFG and its relationship to proficiency in L2 learners. Like us, they found that high (vs. low) bilinguals showed increased activation in this area (their coordinates were close to ours) during irregular inflection. The inferior parietal lobe (Chou et al., 2006; Booth et al., 2008) is another region in which increased activations have been reported as a function of increased skill and development (Hernandez et al., 2015). Besides increased activation in the IFG, LAHP bilinguals in our study showed increased activation in the right inferior parietal lobe relative to LALP bilinguals.

We might ask why *both* early L2 high proficiency (EAHP) and late low proficiency (LALP) bilinguals show less extensive activation, respectively, relative to LAHP bilinguals. We propose that in the case of EAHP, relatively less activation is due to the language task being more automatic for them and therefore requiring less neural effort. For the LALPs relatively lower activation is caused by a different reason, e.g., it is possible that they found the task much harder and perhaps too hard, leading them to disengage or not engage adequately, especially in a task that is timed. This is a possibility that requires further study.

Returning to the question of automaticity, Jeon and Friederici (2015) propose that the degree of automaticity could be the critical factor in the functional organization of the prefrontal cortex. According to their account, anterior prefrontal regions (e.g., BA47, BA45) are involved during processes that require more control and have a low degree of automaticity (such as an L2). Conversely, more posterior prefrontal regions (e.g., BA44) are associated more with processes that have a high degree of automaticity (e.g., L1). Put another way: there is a reliance on BA44 in adult native language processing, which is considered to be highly automatic. In children however (and by extension L2 learners) language learning is still in "development" and there is greater reliance during this period on BA45, a more anterior part of the IFG (Hahne et al., 2004; Vissiennon et al., 2017).

Our findings provide support for such a view. The greater overall activation for LAHP (vs. EAHP) bilinguals in left anterior prefrontal regions (e.g., BA47) suggests that while our LAHP bilinguals were highly proficient and indistinguishable from EAHP bilinguals on past tense accuracy, their late acquired language processing was less automatic than their early counterparts. Similarly, we found significant activation in left BA45 (an anterior PFC region) for LAHP bilinguals (this time in relation to LALP bilinguals). We did not however find this pattern of activity in the more posterior region of left PFC (BA44). BA44—which showed differences in the LAHP vs. LALP contrast—showed very little activation (relative to baseline activation) for the LAHP group during regular or irregular verb inflection. Our previous findings with this same task in L1 native speakers (Oh et al., 2011) showed that BA44 was significantly activated for both regular and irregular verb inflection conditions. If one were to argue that a more automatic L1 was more activated than a high proficiency but less automatic L2 in this region, then this supports the anterior-posterior automaticity gradient in PFC proposed by Jeon and Friederici (2015).

CONCLUSION

In summary our study found that *both* L2 AoA and proficiency independently influence L2 functional activation. We saw L2 AoA affect neural activation in two ways. First, late L2 learning was associated with greater overall activation; this was interpreted as reflecting a lesser degree of automaticity in L2 processing even though high proficiency levels had been attained. We also saw early L2 bilinguals using the left IFG (BA47) selectively depending on whether they were generating regular or irregular past tense verbs. The activation patterns of these early L2 learners were similar to L1 native speakers generating the past tense (Oh et al., 2011). In terms of L2 proficiency effects, our study confirmed that high L2 proficiency was associated with increased activation in a number of regions, including bilateral frontal regions and the left IPL. This was not always the case in the left IFG however, in particular BA44. Increased activation in left BA45 but not BA44 is a pattern that is seen in

the "developmental" stage of language learning in children or when processing is controlled rather than automatic (Jeon and Friederici, 2015; Vissiennon et al., 2017). We propose that the pattern of frontal activation we saw in our LAHP bilinguals was a function of degree of automaticity in L2.

DATA AVAILABILITY

The datasets generated for this study are available on request to the corresponding author.

ETHICS STATEMENT

Full name of the ethics committee that approved the study: Domain Specific Review Board, National Healthcare Group, Singapore.

Consent Procedure

Once ethical approval was received, participants were recruited via advertisements. Participants who responded and fulfilled selection criteria were allowed time to read the participant information sheet (a copy was sent via email or posted), the research assistant then went through the information sheet with them. Signed consent was then obtained but only when the participant was satisfied all their questions and concerns had been addressed. A copy of the signed consent was given to the participant and a copy retained by us. Participants could withdraw at any point without explanation if they so chose.

Additional Considerations

No vulnerable populations were involved in the study reported in this paper.

AUTHOR CONTRIBUTIONS

All authors have each contributed significantly to the manuscript and consent to be named authors. TO was principal investigator (PI) of the grant that funded the study, designed all the studies in the funded project, managed the entire project, and analyzed and interpreted data. SG was co-PI, co-designed the fMRI study, co-wrote the paper, and analyzed and interpreted data. PN oversaw participant recruitment and data collection/scanning and ensured ethical guidelines were strictly adhered to by all involved in the study and contributed to data analysis. IY contributed to data analysis and design of the fMRI study and manuscript. BC contributed to data interpretation, manuscript, and advised on relevant aspects of the study. AE co-wrote the paper. All authors contributed to manuscript revision, read, and approved the submitted version.

FUNDING

ACKNOWLEDGMENTS

REFERENCES

Abutalebi, J. (2008). Neural aspects of second language representation and language control. *Acta Psychol.* 128, 466–478. doi: 10.1016/j.actpsy.2008.03.014

Archila-Suerte, P., Zevin, J., and Hernandez, A. E. (2015). The effect of age of acquisition, socioeducational status, and proficiency on the neural processing of second language speech sounds. *Brain Language* 141, 35–49. doi: 10.1016/j.bandl.2014.11.005

Bird, H. (2003). Deficits in phonology and past-tense morphology: what's the connection? *J. Memory Language* 48, 502–526. doi: 10.1016/S0749-596X(02)00538-7

Booth, J. R., Mehdiratta, N., Burman, D. D., and Bitan, T. (2008). Developmental increases in effective connectivity to brain regions involved in phonological processing during tasks with orthographic demands. *Brain Res.* 1189, 78–89. doi: 10.1016/j.brainres.2007.10.080

Chou, T. L., Booth, J. R., Bitan, T., Burman, D. D., Bigio, J. D., Cone, N. E., et al. (2006). Developmental and skill effects on the neural correlates of semantic processing to visually presented words. *Human Brain Mapping* 27, 915–924. doi: 10.1002/hbm.20231

Consonni, M., Cafiero, R., Marin, D., Tettamanti, M., Iadanza, A., Fabbro, F., et al. (2013). Neural convergence for language comprehension and grammatical class production in highly proficient bilinguals is independent of age of acquisition. *Cortex* 49, 1252–1258. doi: 10.1016/j.cortex.2012.04.009

Evans, S. (2017). What has replication ever done for us? Insights from neuroimaging of speech perception. *Front. Human Neurosci.* 11:41. doi: 10.3389/fnhum.2017.00041

Fabbro, F. (2001). The bilingual brain: cerebral representation of languages. *Brain Language* 79, 211–222. doi: 10.1006/brln.2001.2481

Forman, S. D., Cohen, J. D., Fitzgerald, M., Eddy, W., Mintun, M., and Noll, D. C. (1995). Improved assessment of significant activation in functional magnetic resonance imaging (fMRI): use of a cluster-size threshold. *Magn. Resonance Med.* 33, 636–647. doi: 10.1002/mrm.1910330508

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

Goebel, R., Esposito, F., and Formisano, E. (2006). Analysis of functional image analysis contest (FIAC) data with Brain Voyager QX: from single-subject to cortically aligned group general linear model analysis and self-organizing group independent component analysis. *Human Brain Mapping* 27, 392–401. doi: 10.1002/hbm.20249

Greenbaum (1996). *Comparing English Worldwide: The International Corpus of English*. Oxford: Clarendon Press.

Hahne, A., Eckstein, K., and Friederici, A. D. (2004). Brain signatures of syntactic and semantic processes during children's language development. *J. Cogn. Neurosci.* 16, 1302–1318. doi: 10.1162/0898929041920504

Hasegawa, M., Carpenter, P. A., and Just, M. A. (2002). An fMRI study of bilingual sentence comprehension and workload. *Neuroimage* 15, 647–660. doi: 10.1006/nimg.2001.1001

Henson, R., and Friston, K. (2007). "Convolution models for fMRI," in *Statistical Parametric Mapping*, eds K. Friston, J. Ashburner, S. Kiebel, T. Nichols, and W. Penny (London: Academic Press), 178–192. doi: 10.1016/B978-012372560-8/50014-0

Hernandez, A. E., and Li, P. (2007). Age of acquisition: its neural and computational mechanisms. *Psychol. Bull.* 133, 638–650. doi: 10.1037/0033-2909.133.4.638

Hernandez, A. E., Woods, E. A., and Bradley, K. A. L. (2015). Neural correlates of single word reading in bilingual children and adults. *Brain Language* 143, 11–19. doi: 10.1016/j.bandl.2015.01.010

Hesling, I., Dilharreguy, B., Bordessoules, M., and Allard, M. (2012). The neural processing of second language comprehension modulated by the degree of proficiency: a listening connected speech fMRI study. *Open Neuroimag. J.* 6, 44–54. doi: 10.2174/1874440001206010044

Jeon, H. A., and Friederici, A. D. (2015). Degree of automaticity and the prefrontal cortex. *Trends Cogn. Sci.* 19, 244–250. doi: 10.1016/j.tics.2015.03.003

Kim, K. H., Relkin, N. R., Lee, K. M., and Hirsch, J. (1997). Distinct cortical areas associated with native and second languages. *Nature* 388, 171–174. doi: 10.1038/40623

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

Lebrun, Y. (2002). "Implicit competence and explicit knowledge," in *Advances in the Neurolinguistics of Bilingualism*, ed F. Fabbro (Udine: Forum-Udine University Press), 299–313.

Li, P., Legault, J., and 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

Li, P., Zhang, F., Yu, A., and Zhao, X. (2019). Language History Questionnaire (LHQ3): an enhanced tool for assessing multilingual experience. *Bilingualism Language Cogn.* 8, 1–7. doi: 10.1017/S1366728918001153

Lim, V. P. C., Rickard Liow, S., Lincoln, M., Chan, Y. H., and Onslow, M. (2008). Determining language dominance in English–Mandarin bilinguals: development of a self-report classification tool for clinical use. *Appl. Psycholinguist.* 29, 389–412. doi: 10.1017/S0142716408080181

Liu, H., and Cao, F. (2016). L1 and L2 processing in the bilingual brain: a meta-analysis of neuroimaging studies. *Brain Language* 159, 60–73. doi: 10.1016/j.bandl.2016.05.013

Mårtensson, J., Eriksson, J., Bodammer, N. C., Lindgren, M., Johansson, M., Nyberg, L., et al. (2012). Growth of language related brain areas after foreign language learning. *Neuroimage* 63, 240–244. doi: 10.1016/j.neuroimage.2012.06.043

Marrero, Z. M., Golden, C. J., and Espe-Pfeifer, P. (2002). Bilingualism, brain injury, and recovery: implications for understanding the bilingual and for therapy. *Clin. Psychol. Rev.* 22, 463–478. doi: 10.1016/S0272-7358(01)00109-X

Mechelli, A., Crinion, J. T., Noppeney, U., O'Doherty, J., Ashburner, J., Frackowiak, R. S., et al. (2004). Structural plasticity in the bilingual brain. *Nature* 431, 757. doi: 10.1038/431757a

Meisel, J. (2004). "The bilingual child," in *The Handbook of Bilingualism*, eds T. Bhatia and W. Ritchie (Oxford: Blackwell Publishing Ltd), 91–113.

Nichols, E. S., and 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. doi: 10.1016/j.neuroimage.2016.08.053

Oh, T. M., Tan, K. L., Ng, P., Berne, Y. I., and Graham, S. (2011). The past tense debate: is phonological complexity the key to the puzzle? *Neuroimage* 57, 271–280. doi: 10.1016/j.neuroimage.2011.04.008

Open Science Collaboration (2015). Estimating the reproducibility of psychological science. *Science* 349:aac4716. doi: 10.1126/science.aac4716

Opitz, B., and Friederici, A. D. (2003). Interactions of the hippocampal system and the prefrontal cortex in learning language-like rules. *Neuroimage* 19, 1730–1737. doi: 10.1016/S1053-8119(03)00170-8

Pakulak, E., and Neville, H. J. (2011). Maturational constraints on the recruitment of early processes for syntactic processing. *J. Cogn. Neurosci.* 23, 2752–2765. doi: 10.1162/jocn.2010.21586

Paradis, M. (1994). "Neurolinguistic aspects of implicit and explicit memory: implications for bilingualism," in *Implicit and explicit learning of Second Languages*, eds N. Ellis (London: Academic Press), 393–419.

Perani, D., Abutalebi, J., Paulesu, E., Brambati, S., Scifo, P., Cappa, S. F., et al. (2003). The role of age of acquisition and language usage in early, high-proficient bilinguals: an fMRI study during verbal fluency. *Human Brain Mapping* 19, 170–182. doi: 10.1002/hbm.10110

Perani, D., Dehaene, S., Grassi, F., Cohen, L., Cappa, S. F., Dupoux, E., et al. (1996). Brain processing of native and foreign languages. *Neuroreport* 7, 2439–2444. doi: 10.1097/00001756-199611040-00007

Perani, D., Paulesu, E., Galles, N. S., Dupoux, E., Dehaene, S., Bettinardi, V., et al. (1998). The bilingual brain. Proficiency and age of acquisition of the second language. *Brain* 121 (Pt 10), 1841–1852. doi: 10.1093/brain/121.10.1841

Poldrack, R. A., Sabb, F. W., Foerde, K., Tom, S. M., Asarnow, R. F., Bookheimer, S. Y., et al. (2005). The neural correlates of motor skill automaticity. *J. Neurosci.* 25, 5356–5364. doi: 10.1523/JNEUROSCI.3880-04.2005

Sakai, K. L., Miura, K., Narafu, N., and Muraishi, M. (2004). Correlated functional changes of the prefrontal cortex in twins induced by classroom education of second language. *Cereb. Cortex* 14, 1233–1239. doi: 10.1093/cercor/bhh084

Saur, D., Baumgaertner, A., Moehring, A., Büchel, C., Bonnesen, M., Rose, M., et al. (2009). Word order processing in the bilingual brain. *Neuropsychologia* 47, 158–168. doi: 10.1016/j.neuropsychologia.2008.08.007

Schmidt, S. (2009). Shall we really do it again? The powerful concept of replication is neglected in the social sciences. *Rev. Gen. Psychol.* 13, 90–100. doi: 10.1037/a0015108

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

Stein, M., Winkler, C., Kaiser, A., and Dierks, T. (2014). Structural brain changes related to bilingualism: does immersion make a difference? *Front. Psychol.* 5:1116. doi: 10.3389/fpsyg.2014.01116

Tomoschuk, B., Ferreira, V. S., and Gollan, T. H. (2018). When a seven is not a seven: self-ratings of bilingual language proficiency differ between and within language populations. *Bilingualism Language Cogn.* 22, 516–536. doi: 10.1017/S1366728918000421

Ullman, M. T. (2001). The declarative/procedural model of lexicon and grammar. *J. Psychol. Res.* 30. 37–69. doi: 10.1023/A:1005204207369

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

Ullman, M. T. (2005). "A cognitive neuroscience perspective on second language acquisition: the declarative/procedural model," in *Mind and Context in Adult Second Language Acquisition: Methods, Theory, and Practice*, eds C. Sanz (Washington, DC: Georgetown University Press), 141–178.

Ullman, M. T. (2016). "The declarative/procedural model: a neurobiological model of language learning, knowledge and use," in *The Neurobiology of Language*, eds G. Hickok and S. A. Small (Oxford: Elsevier), 953–968. doi: 10.1016/B978-0-12-407794-2.00076-6

Vissiennon, K., Friederici, A. D., Brauer, J., and Wu, C.-Y. (2017). Functional organization of the language network in three- and six-year-old children. *Neuropsychologia* 98, 24–33. doi: 10.1016/j.neuropsychologia.2016.08.014

Waldron, E. J., and Hernandez, A. E. (2013). The role of age of acquisition on past tense generation in Spanish-English bilinguals: an fMRI study. *Brain Language* 125, 28–37. doi: 10.1016/j.bandl.2013.01.002

Wartenburger, I., Heekeren, H. R., Abutalebi, J., Cappa, S. F., Villringer, A., and Perani, D. (2003). Early setting of grammatical processing in the bilingual brain. *Neuron* 37, 159–170. doi: 10.1016/S0896-6273(02)01150-9

Watkins, K. E., Klein, D., and Johnsrude, I. S. (2017). The neural basis of language learning: brief introduction to the special issue. *Neuropsychologia* 98, 1–3. doi: 10.1016/j.neuropsychologia.2017.03.019

Weber-Fox, C. M., and Neville, H. J. (1996). Maturational constraints on functional specializations for language processing: ERP and behavioral evidence in bilingual speakers. *J. Cogn. Neurosci.* 8, 231–256. doi: 10.1162/jocn.1996.8.3.231

Wong, B., Yin, B., and O'Brien, B. (2016). Neurolinguistics: structure, function, and connectivity in the bilingual brain. *Biomed Res. Int.* 2016, 1–22. doi: 10.1155/2016/7069274

Xue, G., Dong, Q., Jin, Z., Zhang, L., and Wang, Y. (2004). An fMRI study with semantic access in low proficiency second language learners. *Neuroreport* 15, 791–796. doi: 10.1097/00001756-200404090-00010

APPENDIX A

Full word list used in the current study, matched for CV structure and word frequency.

Regular			Irregular		
Verb	CV structure (past tense)	ICE past log frequency	Verb	CV structure (past tense)	ICE past log frequency
Spray	CCCVC	0.3010	Stride	CCCVC	0.6021
Glow	CCVC	0.0000	Bleed	CCVC	0.6990
Sway	CCVC	0.0000	Freeze	CCVC	0.3010
Flow	CCVC	0.0000	Cling	CCVC	0.3010
Free	CCVC	0.6990	Spin	CCVC	0.0000
Fry	CCVC	0.7782	Steal	CCVC	0.0000
Pray	CCVC	0.3010	Swear	CCVC	0.0000
Slow	CCVC	0.0000	Flee	CCVC	0.6990
View	CCVC	1.3617	Slide	CCVC	0.3010
Dry	CCVC	0.9542	Fling	CCVC	0.3010
Cry	CCVC	0.6990	Swing	CCVC	0.6990
Stay	CCVC	1.1139	Break	CCVC	1.2304
Play	CCVC	1.5911	Speak	CCVC	1.5682
Try	CCVC	1.6812	Breed	CCVC	0.0000
Crawl	CCVCC	0.0000	Creep	CCVCC	0.3010
Slap	CCVCC	0.4771	Shrink	CCVCC	0.6000
Train	CCVCC	1.1761	Sweep	CCVCC	0.4771
Drop	CCVCC	1.0792	Sleep	CCVCC	1.0414
Chew	CVC	0.3010	Bite	CVC	1.5441
Weigh	CVC	0.0000	Dig	CVC	0.6021
Sigh	CVC	0.4771	Ride	CVC	0.0000
Tie	CVC	0.8451	Shake	CVC	0.9031
Show	CVC	1.7243	Think	CVC	2.0334
Die	CVC	1.2553	Say	CVC	2.9217
Bake	CVCC	0.0000	Bind	CVCC	1.2553
Care	CVCC	0.6990	Weep	CVCC	0.0000
Guess	CVCC	0.0000	Cost	CVCC	1.9445
Wipe	CVCC	0.3010	Deal	CVCC	1.0414
Fear	CVCC	0.7782	Hurt	CVCC	1.3424
Gain	CVCC	1.1461	Leapt	CVCC	0.0000
Fill	CVCC	1.4771	Sell	CVCC	1.5441
Fail	CVCC	1.4314	Feel	CVCC	1.9395
Pull	CVCC	1.0414	Mean	CVCC	1.7634
Like	CVCC	1.2304	Keep	CVCC	1.8633
Live	CVCC	1.3617	Lose	CVCC	1.9395
Reach	CVCC	1.3617	Hold	CVCC	2.0253
Use	CVCC	2.5328	Heard	CVCC	2.0828
Seem	CVCC	1.8261	Find	CVCC	2.4425
Call	CVCC	2.1790	Tell	CVCC	2.0334
Link	CVCCC	1.2553	Burst	CVCCC	0.8451
Mean		0.8859			1.0297

CV, Consonant-Vowel. Word frequency counts were obtained from the Singapore International Corpus of English (ICE) database (http://ice-corpora.net/ice/icesin.htm); there was no significant difference between mean frequency for regular and irregular verbs (p > 0.05).