Brain & Language 127 (2013) 452–462

Contents lists available at ScienceDirect

Brain & Language

journal homepage: www.elsevier.com/locate/b&l

Two faces, two languages: An fMRI study of bilingual picture naming

Yunqing Li [a], Jing Yang [b,a], K. Suzanne Scherf [a], Ping Li [a,*]

[a] Department of Psychology and Center for Brain, Behavior, and Cognition, Pennsylvania State University, University Park, PA 16802, USA

[b] Guangwai Brain and Language Lab and National Key Research Center for Linguistics and Applied Linguistics, Guangdong University of Foreign Studies, Guangzhou 510420, China

ARTICLE INFO

Article history:
Accepted 13 September 2013
Available online 13 October 2013

Keywords:
Bilingual language processing
Picture naming with face cues
Cultural priming
Congruence effect
fMRI

ABSTRACT

This fMRI study explores how nonlinguistic cues modulate lexical activation in the bilingual brain. We examined the influence of face race on bilingual language production in a picture-naming paradigm. Chinese–English bilinguals were presented with pictures of objects and images of faces (Asian or Caucasian). Participants named the picture in their first or second language (Chinese or English) in separate blocks. Face race and naming language were either congruent (e.g., naming in Chinese when seeing an Asian face) or incongruent (e.g., naming in English when seeing an Asian face). Our results revealed that face cues facilitate naming when the socio-cultural identity of the face is congruent with the naming language. The congruence effects are reflected as effective integration of lexical and facial cues in key brain regions including IFG, MFG, ACC, and caudate. Implications of the findings in light of theories of language processing and cultural priming are discussed.

1. Introduction

Nonverbal cues, such as facial expressions and oral movements, are important for linguistic communications, especially in multilingual contexts where speakers and listeners are not from the same linguistic community. A large number of studies have indicated that listeners integrate both visual and auditory information in speech perception, as demonstrated by the well-known McGurk effect (Dick, Solodkin, & Small, 2010; McGurk & MacDonald, 1976; Skipper, Van Wassenhove, Nusbaum, & Small, 2007). Also, facial cues may facilitate language comprehension in bilingual or multilingual environments. For example, Sueyoshi and Hardison (2005) showed that advanced second language learners of English exhibit enhanced performance in an audiovisual English sentence comprehension task when the auditory information is presented together with faces. Facial cues can provide critical information regarding the linguistic and socio-cultural identity of the speaker or the listener (the interlocutor). Young children also use facial cues from an early age in language acquisition, as shown in recent studies of infants who discriminate between two languages on the basis of speaking faces in silent videos (e.g., Sebastián-Gallés, Albareda-Castellot, Weikum, & Werker, 2012; Weikum et al., 2007).

Throughout our human history, face cues like race have likely been a strong predictor of the language or languages that a speaker or listener knows (e.g., Latino face -> Spanish). However, with increasing globalization, face race may no longer be as reliable a predictor of the language of the speaker or listener, especially in a multi-cultural and multilingual society. In fact, face cues such as race could also potentially interfere with second language processing. In a recent study Zhang, Morris, Cheng, and Yap (2013) tested Chinese–English bilinguals in the United States. The experimenters asked bilingual participants to engage in a computer-mediated conversation while viewing a photograph of the imagined interlocutor, who was either Caucasian or Chinese. Interestingly, bilingual participants spoke their second language (English) less fluently after viewing an image of an imagined Chinese interlocutor than a Caucasian interlocutor. The authors interpreted this finding to reflect implicit cultural priming. That is, face cues and other culturally-laden icons or symbols can automatically prime activation of cultural and linguistic schemas, especially the socio-cultural and linguistic identity of the interlocutor. In so doing, the bilingual speaker may elevate the accessibility of the native language (in this case Chinese), which disrupts the fluency of production in the second language (English).

Understanding how facial cues facilitate or hinder speech processing in the bilingual brain is important in light of current evidence regarding bilingual language activation. There is a body of literature indicating that bilinguals simultaneously activate both languages while speaking or listening to only one language (e.g. Dijkstra & van Heuven, 1998; Marian & Spivey, 2003; see Grosjean & Li, 2013, Chapter 4 for review). Brain imaging studies have identified the neural substrates that support their ability to control the use of multiple languages during speech production, which include the inferior frontal gyrus, the inferior parietal cortex, the anterior cingulate cortex, and the basal ganglia (Abutelabi, 2008; Abutalebi & Green, 2007; Abutalebi et al., 2013; Crinion et al., 2006; Hernandez, 2009; Price, Green, & Von Studnitz, 1999). One way in which

* Corresponding author.
E-mail address: pul8@psu.edu (P. Li).

0093-934X/$ - see front matter © 2013 Elsevier Inc. All rights reserved.
http://dx.doi.org/10.1016/j.bandl.2013.09.005

bilingual speakers use these cognitive control networks to modulate access and use of their multiple languages may be explicit and/or implicit reliance on cues to help them avoid significant interference from the non-target language; after all, bilingual speakers only rarely make mistakes and produce words from the language that they did not intend to use. Such cues could be language specific (e.g., specific words in the language being processed) and/or language non-specific from the communicative context. We suggest that faces may serve as one of these powerful nonlinguistic cues to help bilinguals in this regard. However, given the limited evidence so far, we are only beginning to understand how face cues support or impair bilingual language processes, and there are no data on the underlying neurocognitive mechanisms.

In this study, we investigated the influence of face cues in facilitating or interfering with language production in bilingual participants, and the underlying neural architecture supporting this potential influence. According to the theory of ''audience design'' (Clark, 1996), speakers and listeners seek a shared knowledge base, a common ground in conversation, in order for successful communication to occur. To reach the common ground, speakers often adjust the form and style of speech so that the utterances are tailored to the particular knowledge of the addressee. For example, when talking to listeners in their non-native language, we quickly adapt our speech to a level that reflects the linguistic knowledge and competence of the listener (e.g., simplifying the use of certain words or structures we think may be used by the listener, Van Engen et al., 2010). Race of the interlocutor may be a useful cue to facilitate the mutual understanding for common ground in communication. Another way in which common grounds may be achieved is via ''interactive alignment'' in speech production, during which interlocutors mimic each others' language patterns, including lexical items, grammatical structures, and phonological patterns during conversation (Pickering & Garrod, 2004). The function of such alignment of vocabulary, structure, and phonology is to help interlocutors converge on the same ideas more easily. Given that the ''audience design'' and ''interactive alignment'' hypotheses highlight how inferred knowledge about the interlocutor might influence our language production, it is not surprising that such theories are resonant with hypotheses about ''cultural priming'' that are based on social psychology studies of inter-group dynamics (Fu, Chiu, Morris, & Young, 2007). These hypotheses predict that in-group biases (favoritism due to 'being the same') facilitate cultural communication, whereas out-group interactions introduce cross-group social anxiety that potentially hinders communication (Morris & Mok, 2011; Zhang et al., 2013). Importantly, face cues can be powerful indicators of in-group versus out-group memberships and, therefore, may serve to facilitate and/or hinder communication via cultural priming.

We designed a functional magnetic resonance imaging (fMRI) study to evaluate whether faces can function as cues to modulate bilingual lexical activation during language processing. We were specifically interested in the degree to which face race (Asian versus Caucasian) may facilitate activation of the intended language or inhibit activation of the unintended language. To do so, we tested Chinese–English bilingual participants performing picture-naming tasks in both their first (L1: Chinese) and second (L2: English) languages separately. Importantly, on each trial, prior to naming a picture of a nonliving object, participants saw either an Asian or a Caucasian face.

We predicted that the race of a face could facilitate naming in both the L1 and the L2 provided there was consistency in the expected mapping between the race of the face and the language to be spoken (i.e., congruence of the face–language pairings). For example, if Chinese–English bilinguals form strong expectations about hearing Chinese upon seeing an Asian (specifically Chinese) face and about hearing English upon seeing a Caucasian face (in the target language environment, the United States), we would expect to see faster naming times in their L1 (Chinese) when they are primed with a Chinese face. Similarly, we would expect to see facilitation of naming in the L2 (English) when the Chinese–English bilinguals are primed with a Caucasian face. Moreover, we predict reduced facilitation in naming in the absence of these face cues, and potential interference in naming when participants are primed with a face–language pairing that violates these strong expectations (e.g., being primed with a Caucasian face and having to name in Chinese). We can empirically differentiate facilitation from interference effects by using a ''no face'' condition as the baseline for comparison. Facilitation will be reflected in faster naming latencies compared to the no face condition, whereas interference will be reflected in slower naming latencies compared to the no face condition. In addition, we expected that these differential patterns of facilitation and/or interference would be reflected in the critical brain structures that modulate the control of multiple languages, such as the anterior cingulate gyrus, inferior frontal gyrus, and the basal ganglia (see Abutalebi, 2008; Abutalebi & Green, 2007; Abutalebi et al., 2013). Finally, in our design we also included a monolingual comparison group to help evaluate and interpret the bilingual results. We predicted that face cues would not be as relevant to the monolingual native speakers in our naming task, and as a result, we expected to see much less facilitation of naming responses when Caucasian faces preceded the naming in English, as well as less activation in the language control systems due to the monolingual group's lack of experience with multiple languages.

2. Methods

2.1. Participants

Fifteen Chinese–English bilinguals (eight females; mean age of 24.44 ± 3.43 years) and eleven monolingual English speakers (five females; mean age of 21.36 ± 3.08 years) from the Pennsylvania State University participated in the experiment and received payment for their participation. Both groups of participants were right-handed as judged by the Snyder and Harris's (1993) handedness questionnaire. Informed written consents were obtained from participants before the experiment. The study was approved by the Institutional Review Board of the Pennsylvania State University, and it followed the research and ethics protocols used by the Penn State Social, Life, and Engineering Sciences Imaging Center.

A language history questionnaire (Li, Sepanski, & Zhao, 2006) was used to measure self-reported language learning history and proficiency in the bilingual and monolingual participants. These bilingual participants started to learn English as their second language at an average age of 10.64 (±2.59) years. Their average exposure per day to Chinese (L1) was estimated at 52% and to English (L2) at 48% for their daily activities. They were also asked to rate their English language abilities on a scale of 1 (''not fluent at all'') to 7 (''very fluent''), and their self-reported scores for English reading, writing, speaking, and listening abilities were 5.14 (±0.86), 4.64 (±0.63), 4.5 (±0.94), and 4.86 (±0.86), respectively.

Both the bilingual and the monolingual participants received a standardized test of English receptive vocabulary, the Peabody Picture Vocabulary Test (PPVT-4) (Dunn & Dunn, 1997; Dunn & Dunn, 2007). The native English monolingual speakers differed significantly from the Chinese–English bilinguals in the size of the English vocabulary as measured by PPVT-4 (English monolinguals: mean score = 109 ± 14.99, Chinese–English bilinguals: mean score = 77 ± 7.1; $t_{24} = -7.33$, $p < .001$).

2.2. Design and materials

All participants performed a picture-naming task in a blocked functional MRI design experiment. The bilingual participants performed the naming task in Chinese and English in separate blocks, whereas the monolingual participants only performed the task in English. The pictures included 144 black and white drawings for concrete nonliving objects (e.g., key, strawberry, house), selected from the UCSD International Picture Naming Project picture database (http://crl.ucsd.edu/~aszekely/ipnp/; Bates et al., 2003). All the stimuli corresponded to high frequency words in Chinese (Liu, Shu, & Li, 2007; Liu, Hao, Li, & Shu, 2011) and English (Brysbaert & New, 2009). The stimuli from both languages were matched in naming agreement, word frequency, imageability, visual complexity, and age of acquisition measures (Bates et al., 2003; Liu et al., 2011).

During the experimental conditions, a picture of a blue or red frame was presented for 500 ms, followed by a picture of a person's face together with the frame for another 500 ms. Finally, a picture of an object appeared in the center of the frame while the face was still present and participants were asked to name the object as quickly as possible within 3 s (see Fig. 1). The total time of each trial within a block was 4 s. The color of the frame served as an explicit cue for the participant to name the target picture either in Chinese (C, L1) if the frame was red, or in English (E, L2) if the frame was blue. The face of the person holding the picture frame was either Asian or Caucasian (half male faces and half female faces in each block). The monolingual participants followed the same experimental procedure, except that they only viewed stimuli with blue frames and were told to name pictures in English. Participants were also told that their naming responses would be recorded; however, due to equipment limitations, the naming responses were not actually recorded in the scanner. Following previous studies (e.g. Zou et al., 2012), we collected participants' behavioral naming responses outside the scanner using the same procedure as in the fMRI study (see below).[1]

During scanning, participants were instructed to minimize head, jaw and tongue movements while naming inside the scanner in order to reduce artifacts due to movements. The baseline condition included a colored frame with no face prior to the presentation of the to-be-named object. There were six experimental conditions for the bilingual participants in a 2 (Language: Chinese vs. English) × 3 (Face: Asian vs. Caucasian vs. No Face) design: (1) naming the picture in Chinese while seeing an Asian face, Ca; (2) naming the picture in Chinese while seeing a Caucasian face, Cc; (3) naming the picture in Chinese while seeing no faces, Cn; (4) naming the picture in English while seeing an Asian face, Ea; (5) naming the picture in English while seeing an Caucasian face, Ec; and (6) naming the picture in English while seeing no faces, En. Since the monolingual controls only saw the blue picture frames, they were tested only in the conditions of English naming, that is, Ea, Ec, and En. Fig. 1 illustrates the design and procedure for the bilingual participants.

Both groups completed two runs of the task during a single scanning session. Each run included two blocks of experimental conditions (Ea, Ec, En, Ca, Cc, Cn for bilinguals; Ea, Ec, En for monolinguals), interleaved with 16-s fixation blocks. Each experimental block began with a 2-s instruction followed by six 4-s trials (26 s in total). The order of the blocks was counterbalanced in each run. During the scan, the stimuli were presented through an LCD projector onto a rear-projection screen mounted above the participant's heads. The screen was viewed with an angled mirror positioned on the head coil.

[1] Several researchers have used the overt naming paradigm successfully in the scanner, considering and correcting head motion problems (e.g., Guo, Liu, Misra, & Kroll, 2011; Hernandez et al., 2000; Parker Jones et al., 2012). We have similarly found this paradigm useful and reliable, as discussed below.

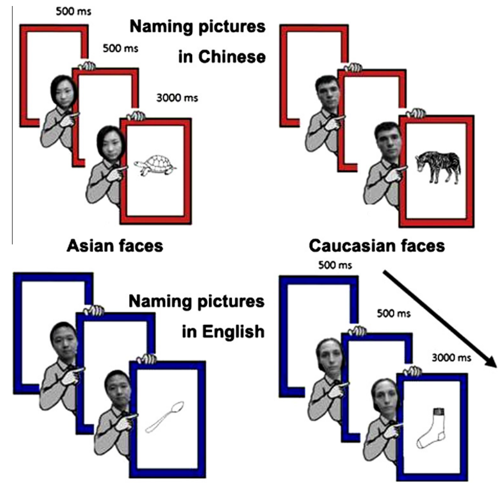

Fig. 1. Task paradigm in the picture naming experiment. There were four experimental conditions in a 2 (Language: Chinese vs. English) × 2 (Face: Asian vs. Caucasian) design: (1) naming the picture in Chinese while seeing an Asian face (Ca); (2) naming the picture in Chinese while seeing a Caucasian face (Cc); (3) naming the picture in English while seeing an Asian face (Ea); and (4) naming the picture in English while seeing an Caucasian face (Ec). A baseline condition of Chinese and English naming involved no faces (Cn and En).

2.3. MRI acquisition

Whole-brain imaging data were acquired on a Siemens Magnetom Trio 3T MRI scanner at the Penn State Social, Life, and Engineering Sciences Imaging Center, using a 12-channel phased array head coil. The functional MRI scans were acquired using a T2*-weighted gradient-echo echo-planar imaging (EPI) sequence (34 AC-PC aligned slices, TE = 30 ms, TR = 2 s, flip angle = 90°, FoV = 240 mm, matrix size = 80 × 80, voxel size = 3 × 3 × 4 mm^3). High-resolution, T1-weighted, anatomical images were also acquired using the 3D MPRAGE sequence with the following parameters: 160 slices, TE = 3.46 ms, TR = 1.8 s, flip angle = 9°, FoV = 220 mm, matrix size = 224 × 224, voxel size = 1 × 1 × 1 mm^3.

2.4. Data analysis

The functional images were analyzed with the Statistical Parametric Mapping software (SPM8, Wellcome Trust Centre for Neuroimaging, University College London; http://www.fil.ion.ucl.ac.uk/spm/software/spm8/) running under Matlab 7.7 (MathWorks, Natick, MA). The first three scans were excluded from data processing to minimize the transient effects of hemodynamic responses. Functional images were corrected for head motion by aligning all volumes to the first volume using a six-parameter rigid-body transformation. Then, the realigned time-series data from participant were normalized according to the MNI stereotactic space from SPM and then spatially smoothed by a 6-mm FWHM Gaussian kernel. Participants were excluded if their head movement during either run of the fMRI task exceeded 3 mm in translation or 0.5° rotation for any axis, and using this criterion, data from only one participant was discarded.

At the individual level, the preprocessed images were submitted to individual GLM-based analyses to estimate the statistical t-maps for each participant. There were 13 regressors in the General Linear Model for the bilinguals, with six (3 L1 and 3 L2) experimental conditions, a baseline condition, and six head-motion parameters. The GLM for the monolinguals included 10 regressors, including 3 L1 experimental conditions, a baseline condition, and six head-motion parameters.

These individual level t-maps were then submitted to separate voxel-wise analyses for each group. In the bilingual group, we examined the effects of Language (Chinese, L1 vs. English, L2) and Face (Asian vs. Caucasian, vs. No Face) in an ANOVA with subject as a random factor in the factorial design. Our analysis strategy was designed to evaluate whether face race functions as a cue for linguistic access primarily in the L1 or L2 of the bilingual participants. To identify potential effects of congruency in language–face pairings on neural activation for bilingual speakers in L2, we performed a mixed-factors ANOVA with group (bilingual vs. monolingual) and face condition (Asian, Caucasian, No Face) as the fixed factors and subject as the random factor for the blocks of *English naming*. To identify such effects in the L1, we conducted a similar ANOVA with group and face conditions when participants named pictures in their respective L1 (i.e., Chinese for the bilingual group vs. English for the monolingual group). All contrast maps were thresholded at $p < .05$ familywise error rate (FWER) corrected for multiple comparisons (35 contiguous voxels at a voxel-level threshold of $p < .005$) as determined by a Monte Carlo simulation implemented in Matlab.

2.5. ROI analyses

Regions-of-interest (ROI) analyses were also conducted on the bilingual data to examine potential differences across the combination of language and face conditions. We selected four regions of interests (ROI) on the basis of previous neuroimaging findings of bilingual language processing (e.g., Abutalebi, 2008; Abutalebi & Green, 2007; Abutalebi et al., 2013; Crinion et al., 2006; Hernandez, 2009; Luk, Green, Abutalebi, & Grady, 2012). These ROIs were all in the left hemisphere and included the middle frontal gyrus (MFG), inferior frontal gyrus (IFG), anterior cingulate gyrus (ACC), and caudate (marked in Table 1).

Time course data were averaged across all the voxels within a 6-mm sphere in each ROI for each bilingual participant. The magnitude of activation for each of the experimental conditions (Ca, Cc, Cn, Ea, Ec, En, and fixation) were averaged across all trials to be converted to percentage BOLD signal changes (PSC), determined by using the formula [(signal-control)/control $\times$ 100] for each time point, where the control condition for each language was the mean signal of the No Face condition (Cn for Ca and Cc conditions; En for Ea and Ec conditions). The averaged percent signal change (PSC) value for each experimental condition was then submitted to an ANOVA, separately for each ROI, with the factors of Language (Chinese L1 vs. English L2) and Face (Asian vs. Caucasian) for the bilingual participants.

3. Results

The accuracy rates and naming latencies for 13 Chinese–English bilingual and 10 English monolingual participants were analyzed and are presented in Fig. 2.[2]

3.1. Behavioral results: Accuracy

3.1.1. Bilingual naming

A two-way ANOVA of bilinguals' naming accuracy, treating Language and Face as independent variables, showed a main effect of Language: the bilinguals were significantly less accurate in English (L2) compared to Chinese (L1) naming across all three face conditions, $F_{(1, 12)} = 49.89$, $p < .001$ (Fig. 2a). This pattern was expected, given that the dominant language of the Chinese–English bilinguals was Chinese and they only had moderate proficiency in English. There was no main effect of Face, $F_{(2, 24)} = 1.66$, $p > .05$, nor a significant interaction effect between Language and Face, $F_{(2, 24)} = .87$, $p > .05$.

3.1.2. Bilingual vs. monolingual naming (in English)

The accuracy rates were comparable for the two groups in their respective L1 (over 95%). A 2 (Group: bilingual vs. monolingual) $\times$ 3 (Face: Ea, Ec, En) ANOVA showed that when naming pictures in English, monolinguals were significantly more accurate than bilinguals across all face conditions, $F_{(1, 21)} = 33.14$, $p < .001$, which is understandable given that English is the L2 for the bilinguals. The main effect of Face was not significant [$F_{(2, 42)} = 1.38$, $p > .05$] for the naming accuracy data, nor was the interaction between Group and Face, $F_{(2, 42)} = .79$, $p > .05$.

3.1.3. Bilingual vs. monolingual naming (in L1)

We conducted a 2 (Group: bilingual vs. monolingual) $\times$ 3 (Face: Asian, Caucasian, No Face) ANOVA to examine the effects of language and face race, but the Group contrast this time was focused on their respective L1 (i.e., naming in Chinese for the bilingual group vs. naming in English for the monolingual group). This analysis revealed no significant main effects or interactions of any type (Fs < 1, $p > .05$), suggesting that the face cues did not affect the accuracy of L1 picture naming in either group.

3.2. Behavioral results: Naming latency

3.2.1. Bilingual naming

As with the analysis of naming accuracy, we conducted an ANOVA of bilinguals' response time (RT), treating Language and Face as independent variables. The analyses revealed that there were no main effects of Language, $F_{(1, 12)} = 1.99$, $p > .05$, and only a trend for a main effect of Face, $F_{(2, 24)} = 3.07$, $p = 0.065$, with naming in the two face conditions generally faster than in the no-face condition (due to facilitation in the congruence situations; see below).

Interestingly, there was a significant effect of the interaction between Language and Face, $F_{(2, 24)} = 3.54$, $p < 0.05$: picture naming in the participants' L1 (Chinese) was significantly faster than picture naming in the participants' L2 (English) when the participants were presented with Asian faces ($t_{12} = -2.33$, $p < .05$), but not when presented with Caucasian faces ($t_{12} = 0.1$, $p > .05$) or with no faces ($t_{12} = -0.69$, $p > .05$). This interaction shows that participants' picture naming was facilitated only when the facial cues and the language used for naming were congruent, but not when the facial cues and the language used for naming was incongruent; the facilitation effect was most pronounced in the L1 (significant reduction of naming latency in the Ca condition), although a similar, but smaller facilitation effect was also observed when the bilingual participants named pictures in the L2 (English) when viewing Caucasian faces (the Ec condition) as compared with viewing no faces (the En condition), $t_{12} = -1.74$, $p = .11$).

3.2.2. Bilingual vs. monolingual naming (in English)

As with the analyses of naming accuracy, we conducted a 2 (Group: bilingual vs. monolingual) $\times$ 3 (Face: Ea, Ec, En) ANOVA. The analyses indicated that not surprisingly, the main effect of Group was significant, $F_{(1, 21)} = 12.28$, $p < .005$, as well that of Face, $F_{(2, 42)} = 4.13$, $p < .05$, but there was no significant interaction, $F_{(2, 42)} = 0.01$, $p > .05$. The lack of any interaction shows that the pattern of the influence of face cues on the two groups was similar: except an overall slower RT (by about 120 ms), bilingual naming in their L2 and monolingual naming in their L1 almost patterned identically. Further paired-samples *t*-tests revealed that for both

[2] The response time data from two bilingual participants and one monolingual participant were not recorded due to equipment failure (microphone malfunction) during the experiment.

Table 1
Significant peak activation within neural regions in bilingual participants during picture naming in each language in each condition (bilinguals)[a].

	L/R	BA	MNI coordinate			T	L/R	BA	MNI coordinate			T
			x	*y*	*z*				*x*	*y*	*z*	
Chinese (L1)	Asian face > Caucasian face						Caucasian face > Asian face					
Superior frontal gyrus	R	8	6	32	48	4.72		NA				
Middle frontal gyrus	L	9	−38	26	30	3.30						
	L	10	−36	60	2	4.20						
	R	9	52	18	34	4.41						
	R	46	58	30	20	3.48						
Medial frontal gyrus	L	8	−12	24	46	3.57						
Inferior frontal gyrus	L	47	−48	32	−10	3.71						
	L	45	−38	38	−6	3.35						
	L	44	−50	14	14	2.97						
Anterior cingulate gyrus[b]	L	32	−20	8	40	3.41						
Insula	L	–	−32	20	−6	2.94						
Inferior parietal lobe	L	40	−46	−56	50	3.64						
Precuneus gyrus	R	19	34	−74	46	3.37						
Angular gyrus	R	39	34	−68	38	3.65						
English (L2)	Asian face > Caucasian face						Caucasian face > Asian face					
Posterior cingulate gyrus		NA					R	30	2	−52	22	3.27
Asian face	Chinese (L1) > English (L2)						English (L2) > Chinese (L1)					
Superior frontal gyrus	R	8	8	46	44	3.00			NA			
Middle frontal gyrus[b]	L	10	−36	60	2	3.85						
	L	9	−38	16	42	3.36						
	L	6	−32	18	48	2.84						
	R	46	48	40	14	3.79						
	R	10	36	54	−12	3.74						
	R	9	48	20	34	3.40						
Medial frontal gyrus	L	11	−2	46	−22	3.68						
	R	8	8	34	50	3.49						
Inferior frontal gyrus[b]	L	47	−46	32	−10	3.59						
Anterior cingulate gyrus	L	24	−4	36	0	3.01						
	R	32	4	32	−14	3.18						
Middle temporal gyrus	L	21	−66	−36	−8	3.88						
	L	39	−50	−66	28	3.51						
	R	21	64	−32	−6	3.71						
	R	39	56	−62	24	3.89						
Inferior parietal lobe	L	40	−46	−58	52	3.56						
Supramarginal gyrus	R	40	48	−48	32	3.70						
Precuneus gyrus	L	7	−6	−50	62	3.65						
	R	7	8	−64	46	3.15						
Paracentral lobe	L	5	−2	−26	52	2.87						
	R	5	6	−22	54	2.87						
Caudate nucleus[b]	L	–	−10	18	−6	3.26						
Putamen	L	–	−14	10	−10	3.19						
Caucasian face	Chinese (L1) > English (L2)						English (L2) > Chinese (L1)					
Anterior cingulate gyrus		NA					R	32	8	32	38	3.00

[a] L, left hemisphere; R, right hemisphere. All activations reported were thresholded at cluster level $p < 0.05$.
[b] Averaged time course data from all voxels within a sphere of 6-mm radius in each region of interests (MFG, IFG, ACC, and Caudate) were extracted for comparisons of BOLD signal changes in experimental conditions versus the no-face control condition for each corresponding language.

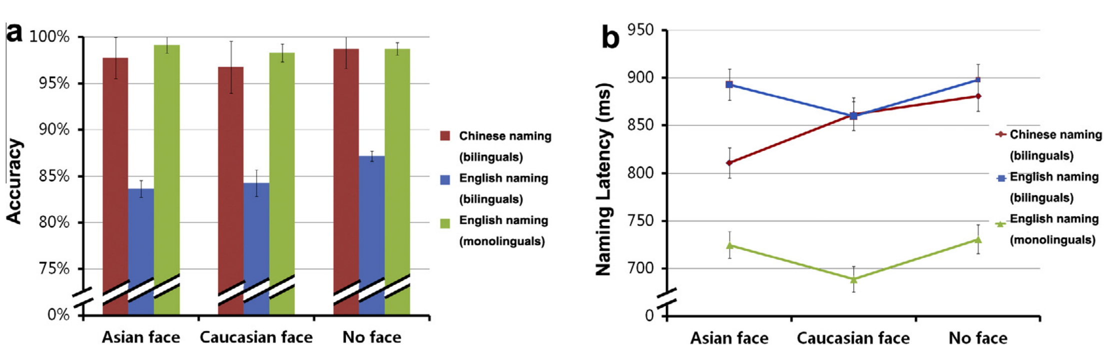

Fig. 2. (a) Accuracy (percent correct) and (b) latency (RT) in picture naming as a function of Language used for naming (Chinese naming, English naming) and Face (Asian, Caucasian, and No Face). Monolinguals are faster and more accurate than bilinguals in picture naming using English, but the pattern of facilitation is similar as a function of language–face congruency; bilinguals also receive a significant facilitation in Chinese naming with Asian faces.

bilinguals and monolinguals, naming in English when viewing Caucasian faces (i.e., congruent linguistic and socio-cultural cues) led to significant facilitation in picture naming, compared with viewing no faces, $t_{22} = 2.59$, $p < .05$, familywise error rate (FWER) corrected. For both groups, the no face condition and the Asian face condition did not yield significantly different naming latencies ($t_{22} = .51$, $p > .05$) when naming in English, showing that incongruent face cues did not specifically lead to interferences.

3.2.3. Bilingual vs. monolingual naming (in L1)

We conducted a 2 (Group: bilingual vs. monolingual) × 3 (Face: Asian, Caucasian, No Face) ANOVA, focusing on picture naming in the respective L1 for each group (i.e., Chinese for the bilingual group vs. English for the monolingual group). The analyses indicated a significant main effect of Group, $F_{(1,\ 21)} = 5.49$, $p < .05$, a marginally significant main effect of Face, $F_{(2,\ 42)} = 3.07$, $p = .06$, and a significant interaction effect between the two factors, $F_{(2,\ 42)} = 3.58$, $p < .05$. The main effect of Group was due to the faster naming speed in the monolinguals than the bilinguals, consistent with the literature that picture naming in English is in general faster than that in Chinese (see Liu et al., 2011 for analyses and discussion). The face cues, however, seemed to work similarly for the two groups when naming occurred in their respective L1: congruent face–language pairings (Ca and Ec conditions) led to facilitation of naming speed as compared with the no face condition, accounting for the significant interaction effects (further paired-samples *t*-tests: $t_{12} = 2.72$, $p < .05$ for bilinguals; $t_9 = 1.91$, $p = .09$ for monolinguals). For both groups, the no face condition and the incongruent face–language pairing conditions (i.e., Cc and Ea) did not yield significantly different naming latencies (ts < 1, $p > .05$) in their respective L1, suggesting that incongruent face cues did not interfere with L1 naming, consistent with the analyses in 3.2.2.

Taken together, these accuracy and RT data are highly consistent in suggesting that for both monolingual and bilingual participants, face cues facilitate naming when their socio-cultural identity is consistent with the naming language. Specifically, Caucasian faces facilitated picture naming in L1 English (monolingual) and L2 English (bilinguals) in the same way as Asian faces facilitated naming in Chinese for the bilinguals, although the latter had a much stronger effect in the RT data (see Fig. 2b and analysis above). However, we did not observe a clear interference or inhibition effect for the incongruent face–language trials with either the bilingual or the monolingual group; the incongruent trials were slower only when compared with the congruent trials but not when compared with the baseline (no face) trials, suggesting that it is mainly facilitation rather than inhibition that is playing a role here.

3.3. fMRI results: Group analysis

3.3.1. Bilingual naming

Table 1 presents a summary of the fMRI results for the bilingual participants. The ANOVA revealed that regardless of facial cues, the naming language influenced patterns of neural activation. Specifically, picture naming in Chinese (L1) elicited significantly greater activation than did naming in English in left orbito-frontal gyrus (BA 11), right anterior cingulate gyrus (BA 32), bilateral middle and superior temporal regions, as well as inferior and superior parietal areas. There were no regions that showed higher activation during English compared to Chinese naming in the bilingual participants. These results suggest a clear effect of language, which was also apparent in the naming accuracy data as shown in Fig. 2a.

With regard to the Face effects, our results indicated that when compared with the no face condition, face cues evoked stronger neural activation in the right ventral visual regions, including the fusiform gyrus (BA 37), occipital face area (BAs 19/37), middle occipital gyrus (BA 19), and lingual gyrus (BA 19). These areas have all been strongly implicated in face processing (e.g., Clark et al., 1996; Haxby, Hoffman, & Gobbini, 2000). Not surprisingly, the reverse comparison (i.e., no face > face conditions) failed to reveal neural responses in these regions.

We also evaluated a more specific contrast of the face effects by directly comparing activation in response to Asian and Caucasian faces for the bilingual participants (Table 1). This comparison showed significantly stronger neural responses in the right superior frontal gyrus (BA 8: 6, 32, 48) to Asian faces than to the Caucasian faces. This result is interesting, in that it shows an "own-race face effect", a finding in the literature indicating that the right superior frontal gyrus is involved in the processing of faces from one's own racial or cultural identity (Mathur, Harada, & Chiao, 2011).

To further identify the neural substrates for the effects of facial cues, we examined brain activation patterns in bilingual participants by directly comparing the Congruent (Ca and Ec) versus Incongruent (Cc and Ea) language–face pairing conditions. First, the congruent conditions elicited more activation in a larger number of regions than did the incongruent conditions. These regions included the right superior frontal gyrus (BAs 9/10), bilateral middle frontal and dorsolateral prefrontal gyri (BAs, 9/10/46), right frontal pole (BA 10), precuneus (BA 19), posterior cingulate gyrus (BA 23), and cerebellum. The reverse whole-brain comparison (Incongruent > Congruent conditions) resulted in no significant activation. These fMRI results are consistent with the behavioral data in pointing to the language–face congruency effect (rather than an interference or incongruence effect); in particular, naming in the L1 and seeing Asian faces (i.e., the Ca condition) led to the strongest facilitation effect, reflected in faster response times (behavioral data) and larger clusters of activation in frontal and parietal regions (fMRI data). The activation of the right superior frontal gyrus implicated in the "own-race face effect", as mentioned above, may have further contributed to facilitate picture naming in one's native language (see Section 4).

3.3.2. Bilingual vs. monolingual naming (in English)

To compare the bilingual and monolingual neural data, we conducted a whole-brain voxelwise ANOVA with the factors of Group (bilingual vs. monolingual) and Face (Ea, Ec, En). The results are presented in Table 2, and also in Fig. 3a as surface 3D images that indicate the brain regions responding more strongly under each of the three face conditions, for the bilingual (blue) versus the monolingual (red) participants. There were several regions showing a main effect of Group: When seeing Asian faces and naming pictures in English, bilinguals showed more neural activation in the left inferior frontal gyrus, bilateral cingulate gyrus, right caudate, and occipital regions, while English monolinguals had more brain activity in the right prefrontal cortex, right temporal regions, bilateral precuneus, and right supramarginal gyrus. When seeing Caucasian faces and naming pictures in English, bilinguals showed less activity in the prefrontal and parietal areas, but more activity in the right inferior posterior temporal areas, while English monolinguals showed similar patterns when presented with Asian faces (except the bilateral posterior temporal regions, which were more active in monolinguals than in bilinguals).

There were also several regions showing a main effect of Face. Both Asian and Caucasian faces evoked greater neural responses in the right hemisphere, specifically in the middle temporal gyrus (BAs 39/19), fusiform gyrus (BA 37), lingual gyrus, and cuneus (see Table 2) when seeing both Asian and Caucasian faces compared to no face. These activation patterns are consistent with the literature regarding key brain regions in face processing and recognition (see Fig. 4).

Our analyses also revealed a significant interaction between Face and Group. Face cues elicited different patterns of activation for the bilingual and monolingual participants, particularly in the Incongruent Ea condition (English naming while seeing Asian faces), as discussed above. Overall bilinguals activated more subcortical areas, such as caudate, parahippocampus and cerebellum, while monolinguals recruited a more extended network especially in temporal and parietal areas. This contrast between bilinguals and monolinguals supports the notion that subcortical regions play

Table 2
Significant peak activation of brain regions during picture naming (bilingual compared with monolingual participants)[a].

	L/R	BA	MNI coordinate x	y	z	T		L/R	BA	MNI coordinate x	y	z	T
Asian face		Bilingual > Monolingual						Monolingual > bilingual					
Inferior frontal gyrus	L	44	−58	6	20	3.44	Superior frontal gyrus	R	8	8	48	48	5.16
	L	47	−54	6	28	3.44		R	10	12	66	16	4.34
Supplementary motor area	L	6	−58	4	38	4.44		R	6	14	24	64	2.97
	R	6	62	−8	32	3.76	Middle frontal gyrus	R	8	44	22	40	3.96
Cingulate gyrus	L	24	−18	−2	36	3.32	Medial frontal gyrus	R	10	10	56	16	4.95
	R	24	22	4	34	4.06	Superior temporal gyrus	L	39	−52	−58	22	4.5
Middle occipital lobe	R	18	38	−92	2	2.83		R	22	60	−52	20	5.24
Inferior occipital lobe	L	18	−28	−92	0	2.94	Middle temporal gyrus	R	39	56	−68	24	3.9
Cuneus	L	17	−20	−86	8	3.43		R	21	48	6	−38	3.08
	R	18	26	−94	0	3.56		R	38	40	8	−36	2.83
Caudate	R	–	22	−42	14	4.43	Inferior temporal gyrus	R	20	50	−2	−38	4.11
Cerebellum	L	–	−4	−74	−22	2.94	Supramarginal gyrus	R	40	50	−48	34	4.11
	R	–	2	−70	−14	3.02	Precuneus	L	31	−10	−54	36	3.57
								R	31	2	−68	26	4.23
Caucasian face		Bilingual > Monolingual						Monolingual > bilingual					
Superior frontal gyrus	L	6	−22	6	66	3.41	Superior frontal gyrus	L	9	−6	54	20	3.21
Middle frontal gyrus	L	6	−22	−10	52	3.42		R	8	8	48	48	4.08
	R	6	22	−8	42	3.65	Medial frontal gyrus	L	8	−4	44	40	3.83
	R	6	22	−10	56	3.15		L	9	−8	42	30	3.75
Supplementary motor area	L	6	−58	4	38	4.43		R	9	8	54	16	3.51
Precentral gyrus	R	4	28	−18	38	3.16	Inferior frontal gyrus	R	47	50	40	−10	3.74
	R	6	40	−16	68	2.98	Superior temporal gyrus	L	39	−48	−60	20	3.78
Fusiform gyrus	R	37	54	−56	−16	3.54	Middle temporal gyrus	R	21	58	−30	−10	3.02
Cuneus	L	18	−20	−92	4	3.42	Inferior temporal gyrus	R	20	62	−22	−20	4.15
Middle occipital gyrus	R	18	28	−94	0	3.8	Supramarginal	L	39	−56	−66	32	3.68
Postcentral gyrus	L	7	−4	−52	72	3.94		R	40	58	−50	24	4.66
Caudate	R		22	−42	12	3.44	Inferior parietal lobe	L	40	−48	−64	50	3.78
Parahippocampal	L	30	−24	−38	6	3.7		R	40	46	−52	38	3.76
Cerebellum	R	–	6	−66	−28	3.6							
No face		Bilingual > Monolingual						Monolingual > Bilingual					
Supplementary motor area	L	6	−58	4	38	4.69	Superior frontal gyrus	R	8	14	46	44	2.91
Inferior frontal gyrus	L	44	−60	6	20	3.58	Middle frontal gyrus	R	9	40	24	36	3.82
Precentral gyrus	R	6	26	−16	72	3.68		R	8	42	28	48	3.26
Fusiform	L	37	−24	−56	−16	4.09	Medial frontal gyrus	L	8	−8	42	38	3.09
Middle occipital gyrus	R	18	38	−92	0	3.48		R	8	6	52	46	3.56
Postcentral gyrus	R	3	58	−12	24	2.91	Inferior frontal gyrus	R	47	36	26	−14	3.5
Cuneus	L	18	−20	−92	2	3.45		R	47	46	24	−6	
	R	18	28	−94	−2	4.59	Superior temporal gyrus	L	22	−62	−58	14	4.27
Cingulate gyrus	R	31	18	−44	22	2.98	Middle temporal gyrus	R	22	52	−38	2	3.86
Caudate	L	–	24	−36	12	4.09		R	21	54	−30	−8	3.38
	R	–	24	−40	16	4.61	Supramarginal	R	40	60	−52	24	4.78
							Inferior parietal lobe	L	40	−60	−46	46	5.04
								R	40	52	−60	48	4.2
							Precuneus	R	31	4	−70	28	3.19

[a] *Notes*: Same as in Table 1.

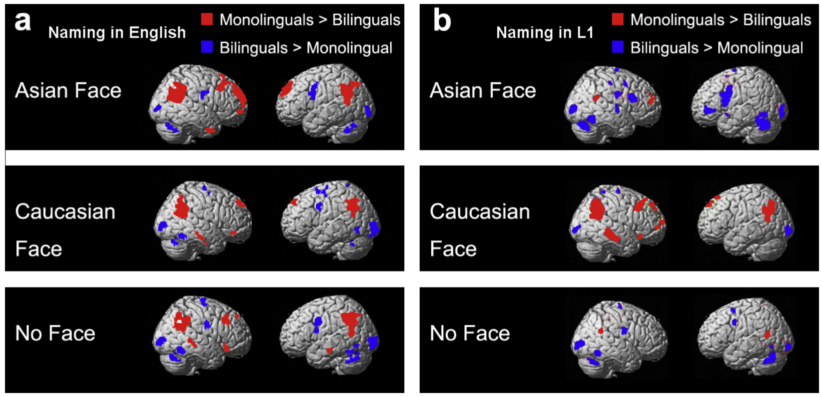

Fig. 3. Surface 3D images displaying brain regions that responded more strongly in the three face conditions for the monolinguals (red) or the bilinguals (blue) participants, (a) when they were asked to name pictures in English following Asian, Caucasian or No face cues; (b) when they were asked to name pictures in their respective L1 (Chinese for the bilinguals vs. English for the monolinguals) under the same face conditions.

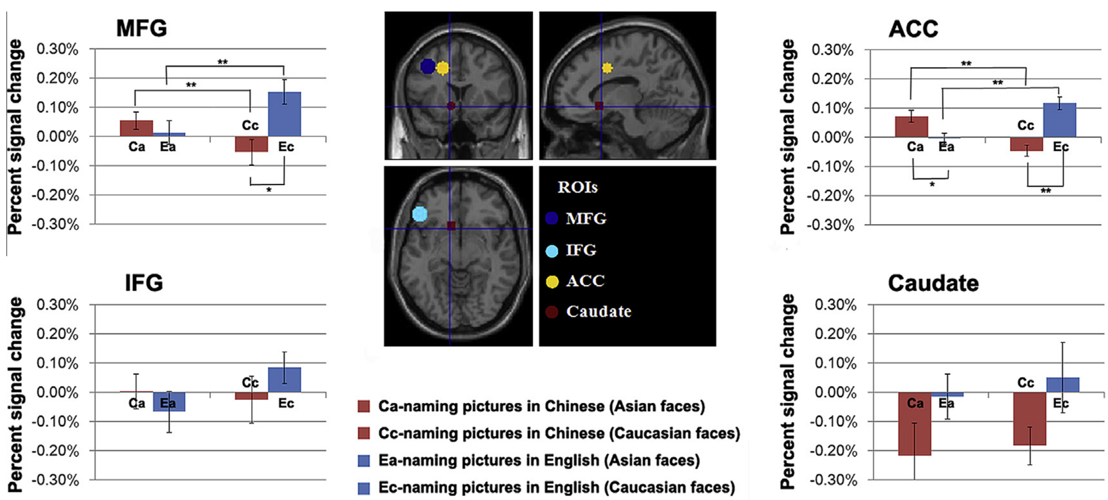

Fig. 4. Region of interest (ROI) analyses indicate significant main effect of Face at caudate, and interaction effects at left MFG and ACC ($^{*}p < .05$; $^{**}p < .01$).

a significant role in bilingual language selection and production (see Abutalebi, 2008; Abutalebi & Green, 2007, for review).

3.3.3. Bilingual vs. monolingual naming (in L1)

As with the behavioral data, we also conducted a 2 (Group: bilingual vs. monolingual) by 3 (Face: Asian, Caucasian, No face) ANOVA when both groups performed the picture naming in their respective L1. The results are presented in Fig. 3b, showing that when viewing Asian faces and naming pictures in their L1, bilinguals (naming in Chinese), as compared with monolinguals (naming in English), showed more neural activation in left inferior frontal gyrus, bilateral premotor and superior temporal regions, fusiform gyrus, middle occipital gyrus and cerebellum. In contrast, monolinguals exhibited stronger activation anteriorly in the right superior medial frontal gyrus and posteriorly in the right supramarginal gyrus. When viewing Caucasian faces and naming pictures in L1, monolinguals, as compared with bilinguals, showed more activity in the prefrontal, middle temporal and bilateral inferior parietal areas, but less activity in bilateral inferior occipital areas. In both cases, consistent with the behavioral data (see Sections 3.1.3 and 3.2.3) there were face–language congruency effects: the congruent face–language pairings elicited stronger activation in more brain regions than did the incongruent face–language conditions. We return to these differences in the General Discussion.

3.4. fMRI results: ROI analysis

To further understand the neural substrates underlying the observed facilitation effects of facial cues in bilingual picture naming, we included analyses of the magnitude of signal change in four *a priori* ROIs. We selected four regions of interest (ROIs) from the left hemisphere, the MFG (BA 9: −38, 16, 42), IFG (BA 47: −48, 32, −10), ACC (BA 32: −20, 8, 40) and Caudate (−10, 18, −6), based on previous neuroimaging findings of bilinguals' language control and processing (e.g., Abutalebi, 2008; Abutalebi & Green, 2007; Abutalebi et al., 2013; Crinion et al., 2006; Luk et al., 2012). Averaged time course data from each ROI were extracted from each bilingual participant's smoothed dataset and sorted according to the experimental conditions using the Marsbar toolbox (Brett, Anton, Valabregue, & Poline, 2002).

ANOVA and paired-samples *t*-tests were conducted to compare the mean percent signal changes (PSC) in each of the four bilingual experimental conditions (Ca, Cc, Ea, Ec) using the corresponding no face condition as the baseline (i.e., Cn as baseline for Ca and Cc; En as baseline for Ea and Ec). A significant main effect of Face was found in the left caudate, $F_{(1,\ 13)} = 5.36$, $p < .05$, but not in the other regions, and there was no main effect of Language in any of the ROIs. Significant interactions between Language and Face were observed in the left MFG [$F_{(1,\ 13)} = 13.69$, $p < .05$] and ACC

[$F_{(1,\ 13)} = 20.22$, $p < .01$]. Importantly, activation in these regions was higher when face cues were congruent with the language used for naming (Ca and Ec) compared to the no face condition. Also, both of these ROIs exhibited negative activation (or deactivation) when facial cues were incongruent with the naming language (especially Cc, naming in Chinese while viewing a Caucasian face). The left IFG showed a similar pattern, in that congruence between language and face cues led to stronger activations. These patterns are consistent with our earlier group analyses of the role that frontal regions play in bilingual production, and with the analysis that face cues facilitated rather than inhibited L1 and L2 speech (see also Section 4).

Finally, although there was no significant interaction between Language and Face in the left caudate, this area showed increased neural activation only when bilinguals named pictures in their L2 when presented with a Caucasian face (Ec). The enhanced neural responses in the left caudate in the English naming (L2) condition might reflect our bilingual participants' increased effort in word selection and articulation in second language production, in contrast to naming in their first language (Abutalebi, 2008; Abutalebi & Green, 2007). This finding is also highly consistent with our earlier comparison of the bilingual versus the monolingual fMRI data that revealed a stronger involvement of subcortical regions in general.

4. General discussion

Research on the neurocognitive mechanisms of bilingualism has flourished in the last decade (see Abutalebi, 2008; Abutalebi & Green, 2007; Abutalebi et al., 2013; Grosjean & Li, 2013; Li, 2013 for reviews). Currently there is strong evidence suggesting that bilingual speakers activate both languages even while speaking only one language (e.g. Dijkstra & Van Heuven, 1998; Marian & Spivey, 2003). Importantly, little is known about whether and how linguistic and/or nonlinguistic cues help bilinguals modulate (facilitate and/or inhibit) this activation of each language so as to consistently produce one of the multiple languages, in spite of the parallel activation. To date, there are only a handful of published studies that examined the degree to which activation of one versus the other language is constrained by linguistic context (e.g., Libben & Titone, 2009; Schwartz & Kroll, 2006), and even fewer that have examined the role of nonlinguistic cues, such as faces (see Zhang et al., 2013).

In this study, we evaluated whether and how one type of nonlinguistic cue, the race of a face, can be used to modulate activation of L1 and L2 in a picture-naming task. We designed an fMRI study in which faces from distinctive races (Asian vs. Caucasian) were orthogonally crossed with the language used for naming pictures (Chinese vs. English). We examined the behavioral and neural responses during picture naming by both bilingual and monolingual speakers.

At the outset of the study we made four predictions based on our research questions and the extant literature (see Introduction). The first was that face cues should significantly enhance the activation of the language to be spoken if the face–language pairings are congruent in terms of the face race and the linguistic identity (e.g., viewing a Chinese face while speaking Chinese). Our behavioral findings are consistent with this prediction in that there were significant facilitation effects in picture naming latency when the facial cue and the language used for naming were congruent, for both bilinguals and monolinguals, in both L1 and L2 (although the facilitation was greater in the L1 for the bilinguals). The second prediction was that there would be interference or disruption when face–language pairings are incongruent, which was not confirmed. Although we observed delay in picture naming when the language used for naming and the face race cues were incongruent (e.g., viewing a Chinese face while speaking English), we did not find evidence that incongruent face–language pairings interfered with or disrupted language production for bilingual participants since reduced facilitation in picture naming (compared with the congruent face–language pairings) was also observed when no face cues were provided (i.e., the baseline condition).

Our third prediction was that potential facilitation or interference in behavioral performance would be reflected in different neural responses, which was partially confirmed. Specifically, a network of frontal and parietal brain regions was more strongly activated when face–language pairings were congruent, in both the bilinguals (when viewing Asian faces and named pictures in Chinese) and the monolinguals (when viewing Caucasian faces and named pictures in English). Interestingly, this same network was not engaged when the face cues and the naming language were incongruent in either the bilingual or monolingual participants, which contradicts our original predictions. Based on the existing literature concerning the use of cognitive control networks during bilingual language production (e.g., Abutalebi, 2008; Abutalebi & Green, 2007; Abutalebi et al., 2013), we expected to observe significant activation in the bilingual language control network including the IFG, ACC, and basal ganglia specifically in the incongruent face–language pairing conditions, particularly in the bilingual participants. Surprisingly, we observed that these areas were more active in the *congruent* face–language pairing conditions, as revealed by the ROI analyses. While these findings at first glance appear to conflict with existing notions of the role of the cognitive control network in modulating bilingual language production, as we discuss below, they are in fact consistent with the notion that these same regions are playing a critical role in integrating the face and language cues, and there is evidence in the extant literature for this integrative role to occur in these regions.

Finally, we predicted that there would be differences in the degree to which face cues might impact bilingual speakers versus monolingual speakers, and this hypothesis was also partially confirmed. In both the behavioral and fMRI data we observed significant group differences between bilinguals and monolinguals, particularly with respect to L1 vs. L2 naming accuracy and speed (higher accuracy and faster RT for monolinguals than bilinguals when naming occurred in English). There was also a stronger involvement of the subcortical areas in bilingual than in monolingual production, particularly the left caudate (as revealed in our ROI analysis) when naming was in English. Interestingly, with additional analyses that contrasted the bilinguals with monolinguals when the two groups performed in their respective L1 (Chinese for the bilinguals and English for the monolinguals), we found that face race interacted with the naming language similarly for both groups as far as the congruence effects were concerned; in other words, face cues impacted L1 naming similarly, but L1 versus L2 naming differently. Consequently, contrary to our prediction that face cues would not be as relevant to native speakers (who lack experience with multiple languages and face races), the monolinguals also showed a facilitation effect of face cues, in that Caucasian faces enhanced picture naming performance for monolinguals (English) in the same way (though to a lesser extent) as Asian faces facilitated picture naming in bilinguals during production of their L1 (Chinese).

These findings have significant implications for understanding the cognitive and neural mechanisms of bilingual language processing. As discussed previously, the recent Zhang et al. (2013) study suggested that face cues, among other cultural icons, are *impediments* to successful L2 processing, due to implicit cultural priming that elevates the accessibility of L1 (which then necessarily hinders L2 production). Our results are inconsistent with those from the Zhang et al. study. We suggest that there are important methodological differences between our studies that may help interpret the discrepancy in findings. In the Zhang et al. (2013) study, the authors estimated the fluency of bilingual speakers using offline measures (words produced per minute and a subjective fluency rating). In contrast, we used online language production measures in which the integration of face cues and language cues occur on a millisecond timescale. It is possible that effective online integration of cues occurs in a fast manner, and offline measures such as those used by Zhang et al. may be insensitive to such integration effects. In addition, Zhang and colleagues only measured the effects of face cues on L2 production, while we measured the potential effects of face cues on the processing of both L1 and L2. This approach revealed that face cues facilitate production in the L1, even for monolinguals, suggesting that the integration between linguistic and nonlinguistic cues is relevant and useful for language production of all speakers. However, the face cues were used more strongly by the bilingual participants, which suggests that "cultural priming" is a more important modulator of bilingual language processing, and is potentially related to the "own-race face effect" seen only in our bilingual participants, as discussed below.

There is a large body of literature investigating the neural substrates of face processing (Haxby et al., 2000; Kanwisher, McDermott, & Chun, 1997). While our study does not address questions concerning the functional organization of networks supporting face-processing specifically, we evaluated whether and how the face-processing network might work collaboratively with the language-processing network in supporting bilingual language production. In fact, we found strong evidence that the core regions of the face-processing network, including the fusiform gyrus, were actively engaged in this task for both the bilingual and monolingual speakers. At the same time, we uncovered evidence of an "own-race face effect" (Mathur et al., 2011) in the neural basis of face processing in our bilingual participants. They evinced stronger activation in the right superior frontal gyrus for Asian compared to Caucasian faces. All of our Chinese–English bilingual participants were college students from China, which suggests that they developed disproportionate visual expertise individuating Asian compared to Caucasian faces, and identifying these faces as belonging to their own ethnic group. No such effects were observed with our monolingual participants when they viewed Caucasian faces, perhaps due to their life experience with a more varied set of faces in the United States (as compared with the bilinguals from China). This difference between the two participant groups accounts in general for the stronger impact that face cues have on bilingual language production, and specifically for the stronger face–language congruency effects in bilinguals than in monolinguals seen in our study.

Based on the recent literature, we initially predicted that there would be more activation in cognitive control areas in this task, particularly for bilinguals in the incongruent conditions. However, this prediction was not supported. Instead, we found more activation of the IFG, MFG/DLPFC, and ACC in *congruent* language–face pairing conditions (i.e., Ca and Ec) instead of the incongruent conditions (i.e., Cc and Ea). In the eyes of conflict monitoring and resolution, this seems rather counter-intuitive. However, as we have discussed, both the behavioral and fMRI data suggest that this activation may reflect facilitation rather than inhibition when compared with the absence of face cue conditions.

\We suggest that one interpretation of these findings is that when the face race served as a useful cue to prime the naming language, monolingual as well as bilingual participants experience more efficient integration of the relevant information for successful language production. Specifically, the activation of these regions in the cognitive control network may reflect a process in which the bilingual speaker is actively *integrating* multiple cues across domains and modalities. Such integration may include pairing the facial cue (Asian or Caucasian) and the language cue (color of picture frame) together with the articulatory motor program for speech production. This interpretation is consistent with a number of previous studies that have shown activation in the ACC and adjacent medial prefrontal areas under conditions when visual and auditory stimuli were contextually congruent (see Calvert, Campbell, & Brammer, 2000; Laurienti et al., 2003). Although our study did not involve auditory stimuli, the fact that participants have to plan and initiate an articulatory program to generate an overt speech output may have motivated the need for such integration. Interestingly, this integration is also implicated in the middle frontal gyrus (MFG), given the role that MFG (and the related prefrontal regions such as DLPFC; BAs 9/10/46) play in spatial visual working memory tasks, especially in the processing of faces and objects (e.g., Courtney, Petit, Maisog, Ungerleider, & Haxby, 1998; Leung, Gore, & Goldman-Rakic, 2002; McCarthy et al., 1996). Our experimental task may have placed high demands on the participants in using spatial working memory (seeing a face, holding it in memory, seeing a language cue, integrating with the speech motor program for target language production), and when effective integration occurs the MFG may become strongly activated as in the face–language congruent conditions.

The lack of enhanced cognitive control in the incongruent face–language conditions with our bilingual participants may also be considered in light of the specific design of our experiment. We used a blocked fMRI design, in which participants were not asked to switch between languages, which may have limited the amount of control and monitoring that our bilingual participants needed to engage. In previously published studies (e.g., Abutalebi, 2008; Abutalebi et al., 2013), event-related fMRI designs were used that likely required significantly more monitoring and cognitive control in bilingual participants as they switched language on a trial-by-trial basis. In the current study we did not use an er-fMRI design on grounds that such a design would introduce too many confounds and create too much confusion if switching were to occur between different languages, different faces, and different naming cues. The blocked design likely did not create a strong conflict situation for our participants to monitor. One consequence of this may have been that participants could only capitalize on the *benefits* of implicit cultural priming, that is, facilitation when the face and language pairings were congruent within a block (as in Ca, Ec). Additional work is needed to confirm this interpretation in which simplified er-fMRI or mixed blocked designs (with alternating face races within each block) are used.

More broadly, our findings suggest that the language–face congruency effects in our study reflect the effects of automatic priming from the visual perception of faces to the language system. When face cues and language cues are congruent, there is facilitation due to implicit spreading of activation from the linguistic/sociocultural identity of the perceived face to the identity of the language needed for production, which is reflected as faster speed in naming latency and as enhanced brain activities in the key language integration areas. When the two are incongruent, no such priming effects could be obtained (see Zhang et al., 2013). These effects of priming from nonlinguistic cues to language production are consistent with both the common ground and interactive alignment theoretical perspectives (Clark, 1996; Pickering & Garrod, 2004; see Introduction). Face race presents an important socio-cultural cue in the communicative context that can facilitate the pursuit of such alignments.

Finally, our picture naming procedure with distinctive faces of different interlocutors may also provide a new paradigm for testing bilingual language production. This paradigm is suitable for use with young children as well as adults, and with minor adaptations in the types of faces and in the specific language pairs used for naming, it can aid in the study of language processing in a variety of contexts. Researchers could also combine and compare face cues with other culturally-laden iconic symbols (see Zhang et al., 2013) to identify the impact of nonlinguistic, contextual cues, as well as linguistic cues on bilingual language processing. Future studies should also examine the context in which face cues are learned and used for the L2, for example, in immersed context as experienced by our participants vs. non-immersed context (see Kroll & McClain, 2013), and examine in further detail how face and language interactions are reflected in the bilingual brain's cognitive control network (see Abutalebi, 2008; Abutalebi & Green, 2007; Hernandez, 2013).

Acknowledgments

This work was supported by the US National Science Foundation's Perception, Action, and Cognition Program and the Linguistics Program (#BCS-1057855), and preparation of the revised manuscript was in part supported by the National Key Research Center for Linguistics and Applied Linguistics, Guangdong University of Foreign Studies, China. We thank Patrick Clark and the staff at the Penn State Social, Life, and Engineering Sciences Imaging Center for assisting with running the fMRI experiment. Y.L. and J. Y. contributed equally to this study. Y.L. is now with the Department of Radiology, College of Medicine, Pennsylvania State University, Hershey, Pennsylvania, USA.

References

Abutalebi, J. (2008). Neural aspects of second language representation and language control. *Acta Psychologica, 128*(3), 466–478.
Abutalebi, J., & Green, D. (2007). Bilingual language production: The neurocognition of language representation and control. *Journal of Neurolinguistics, 20*(3), 242–275.
Abutalebi, J., Rosa, P. A. D., Ding, G., Weekes, B., Costa, A. s., & Green, D. W. (2013). Language proficiency modulates the engagement of cognitive control areas in multilinguals. *Cortex, 49*(3), 905–911.
Bates, E., D'Amico, S., Jacobsen, T., Székely, A., Andonova, E., Devescovi, A., et al. (2003). Timed picture naming in seven languages. *Psychonomic Bulletin and Review, 10*(2), 344–380.
Brett, M., Anton, J. L., Valabregue, R., & Poline, J. B. (2002). Region of interest analysis using the MarsBar toolbox for SPM 99. *Neuroimage, 16*, S497.
Brysbaert, M., & New, B. (2009). Moving beyond Kučera and Francis: A critical evaluation of current word frequency norms and the introduction of a new and improved word frequency measure for American English. *Behavior Research Methods, 41*(4), 977–990.
Calvert, G. A., Campbell, R., & Brammer, M. J. (2000). Evidence from functional magnetic resonance imaging of crossmodal binding in the human heteromodal cortex. *Current Biology, 10*(1), 649–657.
Clark, H. H. (1996). *Using language*. Cambridge: Cambridge University Press.
Clark, V. P., Keil, K., Maisog, J. M., Courtney, S., Ungerleider, L. G., & Haxby, J. V. (1996). Functional magnetic resonance imaging of human visual cortex during face matching: A comparison with positron emission tomography. *Neuroimage, 4*(1), 1–15.
Courtney, S. M., Petit, L., Maisog, J. M., Ungerleider, L. G., & Haxby, J. V. (1998). An area specialized for spatial working memory in human frontal cortex. *Science, 279*, 1347.
Crinion, J., Turner, R., Grogan, A., Hanakawa, T., Noppeney, U., Devlin, J. T., et al. (2006). Language control in the bilingual brain. *Science, 312*(5779), 1537–1540.
Dick, A. S., Solodkin, A., & Small, S. L. (2010). Neural development of networks for audiovisual speech comprehension. *Brain and Language, 114*(2), 101–114.
Dijkstra, T., & Van Heuven, W. J. B. (1998). The BIA model and bilingual word recognition. *Localist Connectionist Approaches to Human Cognition*, 189–225.
Dunn, L., & Dunn, L. (1997). *Peabody Picture Vocabulary Test-III (PPVT-III)*. Circle Pines, MN: American Guidance Service. Inc.
Dunn, L. M., & Dunn, D. M. (2007). *PPVT-4 manual*. Bloomington, MN: NCS Pearson Inc.
Fu, J., Chiu, C., Morris, M., & Young, M. (2007). Spontaneous inferences from cultural cues: Varying responses of cultural insiders and outsiders. *Journal of Cross-Cultural Psychology, 38*, 58–75.
Grosjean, F., & Li, P. (2013). *The psycholinguistics of bilingualism*. New York: Wiley-Blackwell.
Guo, T., Liu, H., Misra, M., & Kroll, J. F. (2011). Local and global inhibition in bilingual word production: fMRI evidence from Chinese–English bilinguals. *NeuroImage, 56*(4), 2300–2309.
Haxby, J. V., Hoffman, E. A., & Gobbini, M. I. (2000). The distributed human neural system for face perception. *Trends in Cognitive Sciences, 4*(6), 223–233.
Hernandez, A. E. (2009). Language switching in the bilingual brain: What's next? *Brain and Language, 109*(2), 133–140.
Hernandez, A. E. (2013). *The bilingual brain*. Oxford, UK: Oxford University Press.
Hernandez, A. E., Martinez, A., & Kohnert, K. (2000). In search of the language switch: An fMRI study of picture naming in Spanish-English bilinguals. *Brain and Language, 73*(3), 421–431.
Kanwisher, N., McDermott, J., & Chun, M. M. (1997). The fusiform face area: A module in human extrastriate cortex specialized for face perception. *The Journal of Neuroscience, 17*(11), 4302–4311.
Kroll, J. F., & McClain, R. (2013). What bilinguals tell us about culture, cognition, and language. *Proceedings of the National Academy of Sciences, USA*. http://dx.doi.org/10.1073/pnas.1309472110.
Laurienti, P. J., Wallace, M. T., Maldjian, J. A., Susi, C. M., Stein, B. E., & Burdette, J. H. (2003). Cross-modal sensory processing in the anterior cingulate and medial prefrontal cortices. *Human Brain Mapping, 19*(4), 213–223.
Leung, H. C., Gore, J. C., & Goldman-Rakic, P. S. (2002). Sustained mnemonic response in the human middle frontal gyrus during on-line storage of spatial memoranda. *Journal of Cognitive Neuroscience, 14*(4), 659–671.
Li, P. (2013). Bilingualism as a dynamic process. In B. MacWhinney & W. O'Grady (Eds.), *Handbook of language emergence*. Wiley-Blackwell: Boston, MA.
Li, P., Sepanski, S., & Zhao, X. (2006). Language history questionnaire: A Web-based interface for bilingual research. *Behavior Research Methods, 38*(2), 202–210.
Libben, M. R., & Titone, D. A. (2009). Bilingual lexical access in context: Evidence from eye movements during reading. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 35*(2), 381–390.
Liu, Y., Hao, M., Li, P., & Shu, H. (2011). Timed picture naming norms for Mandarin Chinese. *PloS ONE, 6*(1), e16505.
Liu, Y., Shu, H., & Li, P. (2007). Word naming and psycholinguistic norms: Chinese. *Behavior Research Methods, 39*, 192–198.
Luk, G., Green, D. W., Abutalebi, J., & Grady, C. (2012). Cognitive control for language switching in bilinguals: A quantitative meta-analysis of functional neuroimaging studies. *Language and Cognitive Processes, 27*(10), 1479–1488.
Marian, V., & Spivey, M. (2003). Competing activation in bilingual language processing: Within-and between-language competition. *Bilingualism Language and Cognition, 6*(2), 97–116.
Mathur, V. A., Harada, T., & Chiao, J. Y. (2011). Racial identification modulates default network activity for same and other races. *Human Brain Mapping, 33*(8), 1883–1893.
McCarthy, G., Puce, A., Constable, T., Krystal, J. H., Gore, J. C., & Goldman-Rakic, P. (1996). Activation of human prefrontal cortex during spatial and nonspatial working memory tasks measured by functional MRI. *Cerebral Cortex, 6*(4), 600–611.
McGurk, H., & MacDonald, J. (1976). Hearing lips and seeing voices. *Nature, 264*, 746–748.
Morris, M. W., & Mok, A. (2011). Isolating effects of cultural schemas: Cultural priming shifts Asian-Americans' biases in social description and memory. *Journal of Experimental Social ology, 47*(1), 117–126.
Parker Jones, Ō., Green, D. W., Grogan, A., Pliatsikas, C., Filippopolitis, K., Ali, N., et al. (2012). Where, when and why brain activation differs for bilinguals and monolinguals during picture naming and reading aloud. *Cerebral Cortex, 22*(4), 892–902.
Pickering, M. J., & Garrod, S. (2004). Toward a mechanistic psychology of dialogue. *Behavioral and Brain Sciences, 27*(2), 169–189.
Price, C. J., Green, D. W., & Von Studnitz, R. (1999). A functional imaging study of translation and language switching. *Brain, 122*(12), 2221–2235.
Schwartz, A. I., & Kroll, J. F. (2006). Bilingual lexical activation in sentence context. *Journal of Memory and Language, 55*, 197–212.
Sebastián-Gallés, N., Albareda-Castellot, B., Weikum, W. M., & Werker, J. F. (2012). A bilingual advantage in visual language discrimination in infancy. *Psychological Science, 23*(9), 994–999.
Skipper, J. I., Van Wassenhove, V., Nusbaum, H. C., & Small, S. L. (2007). Hearing lips and seeing voices: How cortical areas supporting speech production mediate audiovisual speech perception. *Cerebral Cortex, 17*(10), 2387–2399.
Snyder, P. J., & Harris, L. J. (1993). Handedness, sex, and familial sinistrality effects on spatial tasks. Cortex: A journal devoted to the study of the nervous system and behavior. *Cortex: A Journal Devoted to the Study of the Nervous System and Behavior, 29*(1), 115–134.
Sueyoshi, A., & Hardison, D. M. (2005). The role of gestures and facial cues in second language listening comprehension. *Language Learning, 55*(4), 661–699.
Van Engen, K. J., Baese-Berk, M., Baker, R. E., Choi, A., Kim, M., & Bradlow, A. R. (2010). The wildcat corpus of native- and foreign-accented English: Communicative efficiency across conversational dyads with varying language alignment profiles. *Language and Speech, 53*(4), 510–540.
Weikum, W. M., Vouloumanos, A., Navarra, J., Soto-Faraco, S., Sebastián-Gallés, N., & Werker, J. F. (2007). Visual language discrimination in infancy. *Science, 316*, 1159.
Zhang, S., Morris, M. W., Cheng, C., & Yap, A. J. (2013). Heritage-culture images disrupt immigrants'second-language processing through triggering first-language interference. *Proceedings of the National Academy of Sciences, USA*. http://dx.doi.org/10.1073/pnas.1304435110.
Zou, L., Abutalebi, J., Zinszer, B., Yan, X., Shu, H., Peng, D., et al. (2012). Second language experience modulates functional brain network for the native language production in bimodal bilinguals. *NeuroImage, 62*(3), 1367–1375.