Cerebral Cortex, 2020;00: 1–14

doi: 10.1093/cercor/bhaa029
Original Article

ORIGINAL ARTICLE

Left Inferior Frontal Gyrus Integrates Multisensory Information in Category Learning

You Li[1,2], Carol Seger[1,3], Qi Chen[1] and Lei Mo[1]

[1]School of Psychology and Center for Studies of Psychological Application, South China Normal University, Guangzhou 510631, Guangdong, China, [2]Shenzhen Institutes of Advanced Technology, Chinese Academy of Sciences, Shenzhen 518055, Guangdong, China, and [3]Department of Psychology, Colorado State University, Fort Collins, CO 80521 USA

Address correspondence to Lei Mo, School of Psychology, South China Normal University, Guangzhou 510631, China. Email: molei@scnu.edu.cn; Qi Chen, School of Psychology, South China Normal University, Guangzhou 510631, China. Email: qi.chen27@gmail.com.

Abstract

Humans are able to categorize things they encounter in the world (e.g., a cat) by integrating multisensory information from the auditory and visual modalities with ease and speed. However, how the brain learns multisensory categories remains elusive. The present study used functional magnetic resonance imaging to investigate, for the first time, the neural mechanisms underpinning multisensory information-integration (II) category learning. A sensory-modality-general network, including the left insula, right inferior frontal gyrus (IFG), supplementary motor area, left precentral gyrus, bilateral parietal cortex, and right caudate and globus pallidus, was recruited for II categorization, regardless of whether the information came from a single modality or from multiple modalities. Putamen activity was higher in correct categorization than incorrect categorization. Critically, the left IFG and left body and tail of the caudate were activated in multisensory II categorization but not in unisensory II categorization, which suggests this network plays a specific role in integrating multisensory information during category learning. The present results extend our understanding of the role of the left IFG in multisensory processing from the linguistic domain to a broader role in audiovisual learning.

Key words: basal ganglia, category learning, inferior frontal gyrus, multisensory, multimodal

Introduction

Across evolution, humans have developed the ability to quickly respond to an object in the environment based on what category it belongs to. For example, when hunting, one may have to quickly determine whether an animal is a desirable rabbit or a dangerous bear; and when gathering, one may have to skillfully distinguish whether a mushroom is edible or poisonous. The process through which humans acquire this ability to categorize is called category learning. Although much is known about how categories are learned within the visual modality, little is known about how multimodal categories are learned.

The neural mechanism of unisensory category learning, especially of visual categories, has been extensively studied (Ashby and Ell 2001; Ashby and Maddox 2005, 2011; Seger and Miller 2010). One predominant neurobiological account is the Competition between Verbal and Implicit Systems (COVIS) model, which proposed that 2 dissociable systems are responsible for category learning (Ashby et al. 1998; Ashby and Maddox 2005). These systems consist of an explicit hypothesis-testing system involved in rule-based (RB) categorization in which the decision boundary separating the categories can be easily verbalized, and an implicit procedural learning system involved in information-integration (II) categorization in which the optimal decision boundary is difficult or impossible to verbalize. II categorization requires combining information from 2 or more stimulus dimensions and depends on

stimulus-response association through feedback and reinforcement. Computational modeling and functional neuroimaging research have identified a corticostriatal network underlying II category learning (Ashby et al. 1998; Peterson and Seger 2013; Seger 2008). Within this network, information is sent from visual cortex to the body and tail of the striatum (Seger 2013), which in turn has a feedforward projection to SMA. The striatum learns via reinforcement learning mechanisms to associate visual input with appropriate motor responses. Extensions of this theory have allowed for the development of automaticity or habitual responding, in which control of behavior shifts to motor cortex and the putamen (Ashby et al. 1998; Ashby and Maddox 2011; Seger 2018), and the acquisition of flexible categorical representations via learning of separate mappings from perceptual information to category and from category to motor response (Cantwell et al. 2015).

One issue that remains unknown, however, is how the brain learns categories that require integration of information from multiple sensory modalities. In the real world, objects are associated with rich multisensory information, and across development humans form multimodal conceptual representations of the world. However, there is little research concerning the behavioral and neural processes underlying multisensory category learning. Two previous studies (Maddox et al. 2006; Smith et al. 2014) compared learning multimodal II categories (with one visual dimension such as angle of orientation and one auditory dimension such as pitch) with unimodal visual II categories (with 2 visual dimensions, such as spatial frequency and angle of orientation). Both studies found overall above chance performance for multimodal category learning, but worse performance on multimodal than unimodal categories. Closer examination of individual subject strategies revealed that some participants were able to effectively learn to integrate information across visual and auditory modalities but that a substantial subset failed to learn to integrate and instead relied on suboptimal unimodal RB or guessing strategies (Maddox et al. 2006; Smith et al. 2014).

At the neural level, it is unclear what networks might underlie multisensory II category learning. Originally proposed for visual category learning, the COVIS model hypothesized that the implicit procedural learning system responsible for II category was mediated largely within the caudate tail (Ashby et al. 1998; Ashby and Maddox 2005), which received projections from visual cortex (Seger 2013). For multimodal learning to share the same mechanism would require that information from auditory and visual cortical areas converge in this region. Although evidence is sparse, tracer studies in macaques and rodents find that auditory areas in the superior temporal lobe also project to the caudate and overlap with projections from visual areas in the inferior temporal lobe in some regions (see Seger, 2013 for review). Unimodal auditory category learning has been shown to recruit the body and tail of the caudate (Lim et al. 2014, 2019; Yi et al. 2016) in a manner similar to that found in studies of unimodal visual category learning. At the cellular level, electrophysiological evidence indicates that some individual caudate cells respond to stimulation from 2 (or more) sensory modalities (Nagy et al. 2006; Reig and Silberberg 2014). Multimodal information may also be integrated via cortical mechanisms. Previous research has found that the left inferior frontal lobe is sensitive to information accumulation in audiovisual perceptual decision-making tasks (Noppeney et al. 2010) and audiovisual II in speech perception (Fernandez et al. 2017; Miller and D'Esposito 2005; Tse et al. 2015), and multimodal statistical learning studies have associated multimodal learning with the inferior frontal lobe (Frost et al. 2015; Paraskevopoulos et al. 2018).

In the present study, using functional magnetic resonance imaging (fMRI), we systematically investigated the neural mechanisms underlying multisensory II category learning. In separate sessions a week apart (Fig. 1A), participants learned multisensory visuo-auditory categories and unisensory visual categories. Participants who were successfully able to learn and used an II strategy proceeded to the fMRI scanning portion of the study in which they performed the categorization task and a matched control task. To control for bottom-up inputs and motor responses, the control task adopted the identical stimuli as the categorization task but required a judgment on a categorization-irrelevant dimension. This design allowed us to identify a sensory-modality-general network active during II categorization (regardless of whether the information came from a single modality or from multiple modalities) by calculating the conjunction of the multisensory categorization activation and unisensory categorization activation. Furthermore, and most importantly, by comparing multisensory categorization activation to the unisensory categorization activation, this design allowed us to identify for the first time a network specific to multisensory II categorization.

Materials and Methods

Participants

Thirty students were recruited from South China Normal University. After the behavioral learning sessions, decision bound model analysis was performed on the behavioral data to determine the categorization strategy used by each participant (see Design and Procedure). Ten participants were classified as RB strategy users in their first behavioral learning session and were then excluded from the following imaging and behavioral sessions. The final sample in the present study included 20 participants (11 females, 9 males) who were classified as II strategy users in both the multisensory learning and unisensory learning sessions. The study was approved by Ethics Committee of School of Psychology, South China Normal University. All participants gave informed consent and were paid for participation. All participants were native speakers of Mandarin Chinese, right-handed, and had normal or corrected-to-normal vision.

Design

Each participant completed 4 sessions (Fig. 1A). Each fMRI session was preceded by a learning session in the behavioral laboratory completed one day before. Half of the participants completed 1) the multisensory category learning session on the first day and 2) the multisensory fMRI session on the next day. One week later, they completed 3) the unisensory category learning session and 4) the unisensory fMRI session on 2 consecutive days. To counterbalance the learning sequence of multisensory and unisensory stimuli, the other half of the participants completed the unisensory learning session and the unisensory fMRI session first, and then the multisensory sessions after one week. In each fMRI session, participants performed a categorization task and a control task in alternating blocks. The categorization task and the control task shared identical stimuli but required different judgments. In the categorization task, participants decided which category the stimulus belonged to,

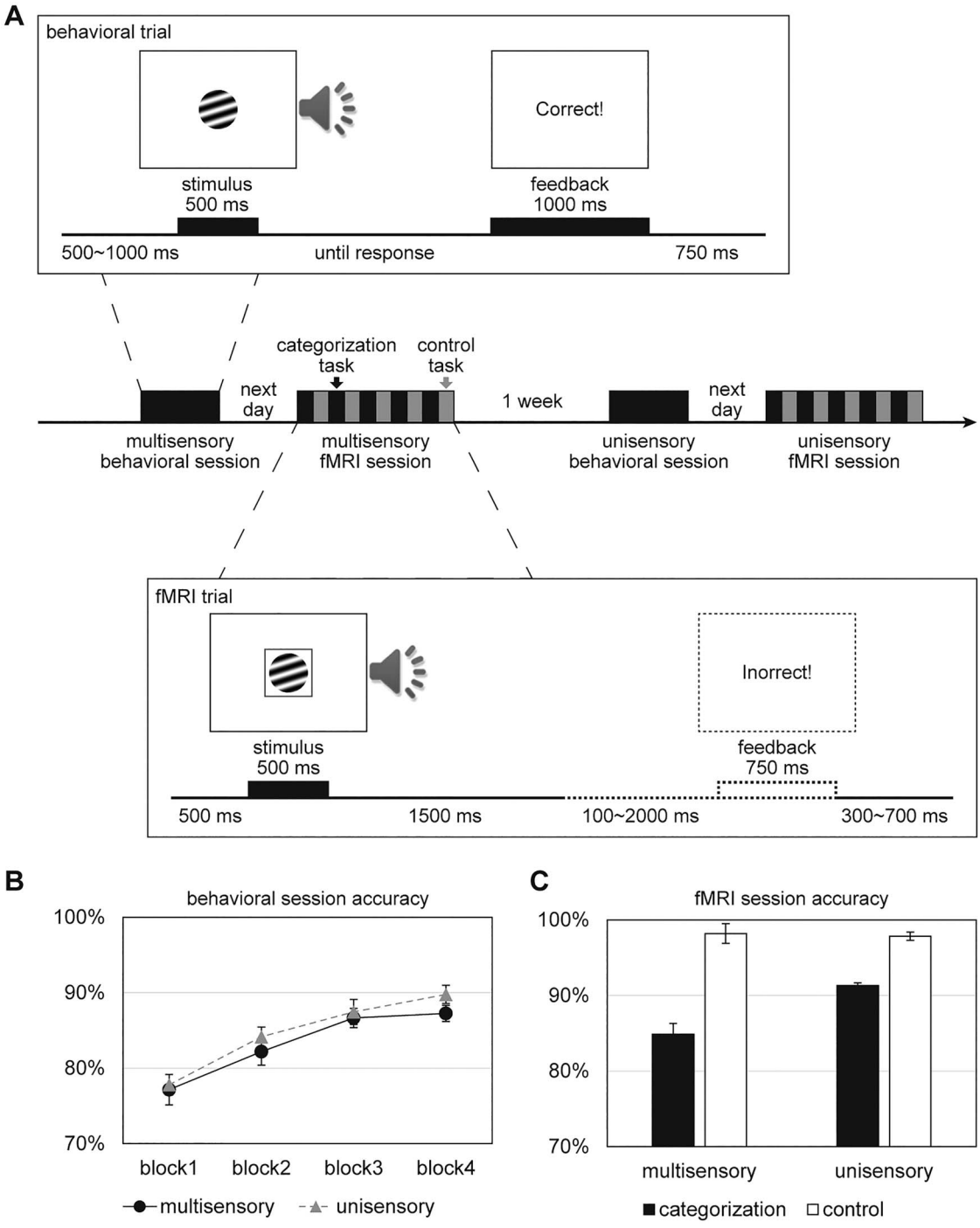

Figure 1. Experimental procedure and behavioral results. (*A*) Each participant completed 4 sessions. Each fMRI session was preceded by a learning session in the behavioral laboratory completed one day before. Shown in this figure is the procedure for half of the participants: 1) the multisensory category learning session on the first day, 2) the multisensory fMRI session on the next day, and one week later, 3) the unisensory category learning session, and 4) the unisensory fMRI session on 2 consecutive days. For the other half of the participants, not shown in the picture, the unisensory behavioral session and the unisensory fMRI session were completed first and the multisensory behavioral session and the multisensory fMRI session were completed one week later. For the behavioral learning sessions, multisensory stimuli were Gabor patches varying in spatial orientation (with the spatial frequency fixed) and pure tones that varied in auditory frequency. Unisensory stimuli were Gabor patches that varied in spatial orientation and frequency. Feedback was presented on every trial in the behavioral sessions. For each fMRI session, participants performed 2 tasks, the categorization task and the control task, in alternate blocks. Stimuli were the same as in the corresponding behavioral session, with the only exception being that a colored square was presented surrounding each Gabor. The color of the square was randomly chosen so that it was irrelevant to the category of the target stimulus. In each fMRI session, the categorization task and the control task shared the identical stimuli but required different judgments. In the categorization task, participants decided which category the target stimulus belonged to, whereas in the control task they decided which color the square was. Feedback was presented on half of the fMRI trials (indicated by the dashed line in the fMRI trial box). (*B*) Mean accuracy in the behavioral learning sessions. Data from each session were divided into 4 equal blocks to examine the learning curve. (*C*) Mean accuracy of the categorization and control tasks in the fMRI sessions. Error bars indicate standard error of means.

whereas in the control task, they decided the color of a square framing the stimulus, which was irrelevant to the category of stimulus (Fig. 1A).

Stimuli

Behavioral Sessions

In the multisensory session, stimuli were Gabor patches varying in spatial orientation (with the spatial frequency fixed) and pure tones varying in auditory frequency. In the unisensory session, stimuli were Gabor patches varying in spatial orientation and frequency. Stimuli were sampled from an II category structure generated using the randomization technique (Ashby and Gott 1988; Seger et al. 2015). First, in an arbitrary 100-by-100 2D exemplar space, 2 "pre-rotated" bivariate normal distributions were generated as a pair of RB categories with the optimal decision boundary perpendicular to the *X*-axis (category A: mean $X = 35$, standard deviation (SD) $X = 4$, mean $Y = 50$, SD $Y = 19$, covariate $XY = 0$; category B: mean $X = 65$, SD $X = 4$, mean $Y = 50$, SD $Y = 19$, covariate $XY = 0$). Second, the "pre-rotated" distributions were rotated 45° so that the decision boundary diagonally went through the origin with a 45° slope, which required the optimal categorization to integrate information from the 2 dimensions. Third, exemplars were randomly drawn from the "rotated" distributions. To reduce the number of outliers, exemplars exceeded 2 SDs from mean were dropped. Four, the arbitrary coordinates of the exemplars were converted to the physical parameters. For multisensory stimuli, the orientation of the Gabor patch (in degrees) was converted using $f(x) = 0.9x$, with spatial frequency fixed at 1.5 cycle/degree; the auditory frequency of the pure tones (in Hz) was converted using $f(y) = 220 \times 2^{(y/120)}$. For unisensory stimuli, the spatial orientation of the Gabor patch (in degrees) was converted using $f(x) = 0.9x$; the spatial frequency (cycle/degree) was converted using $f(y) = 0.0283y + 0.5$. All Gabor patches were 6° in radius and were presented with brightest points in white (RGB: 255, 255, 255) and darkest points in black (RGB: 0, 0, 0) against a gray background (RGB: 127, 127, 127). Visual stimuli were presented on an LCD screen at 60 cm viewing distance. Auditory stimuli were presented via stereo headphones. The loudness level was adjusted to the maximum within the comfortable range by each participant before the experiment. For each learning session, 150 exemplars were sampled (75 each category). Each exemplar was repeated 4 times, resulting in 600 learning trials.

fMRI Sessions

Stimuli for the fMRI sessions were generated in the same way as those in the learning sessions. In total, 200 exemplars were sampled (100 each category) from the same II category distribution and converted to physical parameters using the same functions. The only difference is that, in order to include a control task (the color judgment task) in the fMRI sessions, all Gabor patches were surrounded with a simultaneously presented colored square. The square was of 6° inner side length and of 0.3° edge width. The color of the square was chosen so that it was totally irrelevant to the category of the target stimulus. For half of the participants, the square was blue in the categorization task, red in a random half of the control trials, and yellow in the other half of the control trials. For the other half of the participants, the square was red in the categorization task, blue in a random half of the control trials, and green in the other half of the control trials. By this means, square color was counterbalanced between tasks across participants. Except for square color, the stimuli in the control task were identical to those in the categorization task. There were 400 trials in each fMRI session (200 categorization trials and 200 control trials). Visual stimuli were presented through an LCD projector onto a rear projection screen located behind the participant's head. Participants viewed the screen through an angled mirror on the head coil of the MRI setup. The size of the stimuli in terms of degree of visual angle was matched to the size during the behavioral sessions. Auditory stimuli were delivered via MR-compatible stereo headphones. Loudness was adjusted before the experiment to a level that auditory stimuli could be heard clearly during the fMRI scanning.

Procedure

Behavioral Sessions

Participants were tested individually while seated in a soundproof room in front of a computer screen. They were instructed to learn to categorize the stimuli through trial and error. Each trial started a central fixation cross with a random duration selected from 500, 750, and 1000 ms. Then, for the multisensory learning session, a Gabor patch and a pure tone were presented simultaneously for 500 ms, while for the unisensory learning session, only a Gabor patch was presented for 500 ms. After that, a central fixation cross was presented until a response was made. Participants indicated which category the stimulus belonged to by a button press with the right or left hand. Feedback (the word "Correct" or "Incorrect" in English) was then presented for 1000 ms. The next trial started after a 750-ms blank screen. The 600 learning trials were presented in random order. Participants had a rest break after every 30 trials. The mappings between response hands and categories were counterbalanced across participants.

After each behavioral session, decision bound model analysis (Ell and Ashby 2012; Seger et al. 2015) was performed for each participant. This analysis fitted the behavioral data from the last 150 trials of each session to several strategy models, and determined which strategy (unidimensional RB strategy, conjunctive RB strategy, II strategy, or random guessing response) was used by each participant. Only participants who used an II strategy in the learning session proceeded to the scanning session and ultimately were included in the analyses. Ten participants were classified as RB or guessing strategy users following the first behavioral session and were excluded. These results are consistent with previous studies of multimodal category learning, which found that some but not all participants adopted the optimal II strategy (Maddox et al. 2006; Smith et al. 2014). No subjects were eliminated after the second behavioral session, indicating that subjects who adopted an II strategy in the first session also adopted an II strategy in the second session. By these means, all the 20 participants in the present study were quantitatively classified as successful II category learners who used II strategy in both the multisensory learning and unisensory learning sessions.

fMRI Sessions

Each fMRI session included 2 tasks, the categorization task and the control task. In the categorization task, participants categorized the stimuli according to the categories learned one day before. Participants held individual response boxes inside the scanner in each hand and pressed the corresponding buttons with their left and right thumbs. To control bottom-up inputs and motor responses, the control task used stimuli that were

identical to those in the categorization task, but participants were instructed to decide which color the square surrounding the Gabor was. Each trial began with a 500-ms central fixation cross. Then the stimulus was presented for 500 ms. In the multisensory session, the stimulus consisted of simultaneous presentation of a Gabor patch, a surrounding colored square, and an auditory pure tone. In the unisensory session, the stimulus was a simultaneously presented Gabor patch and a surrounding colored square. After the stimulus, a blank screen was presented for 1500 ms during which the participant made a response. In half of the trials, feedback was presented for 750 ms according to the correctness of the response (categorization or color judgment task). The interval between the feedback and the preceding blank screen was jittered between 100, 200, 300, 400, and 2000 ms. The trial ended after the feedback. In the other half of the trials, no feedback was presented and the trial ended after the 1500-ms blank screen and response. The inter-trial interval (ITI) was jittered between 5 levels of 300, 400, 500, 600, and 700 ms. The stochastic presentation of the feedback (50% of the trials) and jittering between response and feedback was adopted from Seger et al. (2015), in order to dissociate the neural response to the stimulus and response from that to the feedback.

The 2 tasks were presented in a blocked design. The 400 trials (200 categorization trials and 200 control trials) were divided into 40 10-trial blocks. The 20 categorization blocks and the 20 control blocks were presented alternately. Each block started with a 2-s instruction that informed the participants of the upcoming task, that is, words "categorization task" or "color task" in Chinese. Within the 10 trials in each block, the number of exemplars from each category was matched (5 exemplars from one category and 5 from the other one). Each of the 5 ITI levels was repeated twice in a block. The 10 ITI lengths were presented in random order. Two null trials, which presented the central fixation cross alone for 2000 ms, were added into each block to further jitter the trial sequence. The 12 trials in each block (10 task trials and the 2 null trials) were presented in random order. Each block lasted about 43 s so that each fMRI session lasted about 30 min. Participants were instructed to maintain central fixation and not to move their heads throughout the scanning.

Imaging Data Acquisition

Images were obtained with a 3.0 Tesla Siemens Trio MRI scanner with a 12-channel head coil at the Brain Imaging Center at South China Normal University. Functional images were collected using a T2*—weighted echo planar sequence (matrix size = 64 × 64; voxel size = 3.4 × 3.4 × 3.5 mm^3; 36 transversal slices; slice thickness = 3.0 mm; slice gap = 0.75 mm; TR = 2200 ms; TE = 30 ms; FOV = 220 mm; flip angle = 90°). Each session included 799 EPI volumes, with the first 3 volumes discarded to allow for the magnetic field to reach a steady state. Structural images were collected using a T1-weighted magnetization-prepared rapid gradient echo sequence (matrix size = 256 × 256; voxel size = 1 × 1 × 1 mm^3; 176 sagittal slices; slice thickness = 1 mm; TR = 2300 ms; TE = 3.24 ms; FOV = 256 mm; flip angle = 9°).

Imaging Data Analysis

Data analysis was performed using SPM 12 (https://www.fil.ion.ucl.ac.uk/spm) (Wellcome Trust Centre for Neuroimaging, University College London). Preprocessing involved correction of slice time acquisition differences, motion correction of each volume to the first volume, and coregistration of the structural image to the mean functional image. The structural image was normalized to the Montreal Neurological Institute (MNI) space. The resultant deformations were subsequently applied to the functional images, and the functional images were resampled to 3 × 3 × 3 mm^3. Spatial smoothing was performed with a 6 mm full-width at half-maximum Gaussian kernel.

At the single-participant level, 4 separate event-related general linear models (GLM) were created to address different questions. Each GLM was created by combining the multisensory and the unisensory sessions from the same participant. In GLM1, for each session, 4 events of interest were included: the categorization task (onset of stimulus presentation in each categorization trial), the feedback in the categorization task, the control task (onset of stimulus presentation in each control trial), and the feedback in the control task. Note that, for both tasks, the feedback was presented stochastically, that is, only on half of the trials, and, if presented, the onset was jittered from the onset of the task. By this means, the task events and the feedback events were designed to be maximally uncorrelated so that the neural response during categorization processing (or color judgment) could be estimated separately from that during feedback processing. The null trials were not modeled and thus were included in the implicit baseline in the GLM. Each event was convolved with 2 basic functions, the canonical hemodynamic response function and its temporal derivative, with event duration of zero. The 6 head movement parameters derived from the realignment procedure were also included as covariates of no interest. Data were high-pass filtered with a cutoff frequency of 1/128 Hz to remove the low-frequency noise. Temporal autocorrelation was modeled using an AR (1) process. Simple main effects for each of the 4 tasks were computed by applying 1 on the regressor of interest and 0 on all the other regressors, that is, "task > implicit baseline" contrasts of categorization task in multisensory session (MultiCate), control task in multisensory session (MultiCon), categorization task in unisensory session (UniCate), and control task in unisensory session (UniCon). At the group level, the 4 first-level individual contrast images were input to a random effects within-subject analysis of variance (ANOVA) model. For GLM1, each participant's mean accuracy rate for each of the 4 tasks was entered into the group-level GLM as a covariate in order to further rule out the confound of task difficulty.

In GLM2, trials in the categorization task were modeled separately for correct categorization trials and incorrect categorization trials. The control tasks and feedback in each task were also modeled as events of no interest. To investigate the effect related to correct and incorrect categorization, the group-level ANOVA for GLM2 included 4 events, correct categorization in multisensory session (MultiCorr), incorrect categorization in multisensory session (MultiIncorr), correct categorization in unisensory session (UniCorr), and incorrect categorization in unisensory session (UniIncorr).

In addition, 3 control analyses were performed. To verify the effect found in GLM1 and rule out the possibility that our results were affected by idiosyncrasies on incorrect trials, GLM3 replicated the ANOVA from GLM1, but included only correct trials (in categorization and control) for each task. In order to rule out the possibility that the effects found in GLM1 and GLM2 were affected by feedback processing rather than purely categorization processing, in GLM4 and GLM5, only trials in which no feedback was provided (half of the trials) were modeled and entered into the group-level analysis. Otherwise, GLM4 and

GLM5 were performed in the same manner as GLM1 and GLM2, respectively.

For all GLMs, the activation threshold was $P < 0.001$ (uncorrected) at the voxel level, and $P < 0.05$ (FDR corrected) at the cluster level.

Additionally, region of interest (ROI) analyses were performed to further examine patterns of activity in specific regions using MRIcron (https://people.cas.sc.edu/rorden/mricron/) (McCausland Center for Brain Imaging, University of South Carolina) and the MarsBaR toolbox (http://marsbar.sourceforge.net/) (Brett et al. 2002). The Brodmann atlas was used to define Brodmann's area (BA 44 and 45) ROIs. The AAL (Automated Anatomical Labeling) atlas was used to define basal ganglia ROIs (i.e., the caudate, globus pallidus, and putamen, left and right side separately) (see Fig. 3). Within the caudate nucleus, the head of the caudate was divided from the body and tail of the caudate using the same methods as in a previous study (Seger and Cincotta 2005), with the border along an oblique plane angled at 45° from horizontal running between the lines defined by $y = 0$, $z = 14$ and $y = 10$, $z = 24$. Neural activity beta values for each task were extracted from the BA ROIs and the basal ganglia ROIs and were subjected to repeated-measures 2 (Modal: multisensory vs. unisensory) × 2 (Task: categorization vs. control) ANOVAs. For the basal ganglia ROIs, beta values for correct and incorrect categorization trials for each task were also extracted and were subjected to repeated-measures 2 (Modal: multisensory categorization vs. unisensory categorization) × 2 (Correctness: correct vs. incorrect) ANOVAs.

Results

Behavioral Results

In all behavioral learning sessions and fMRI sessions, participants were instructed to respond as accurately as possible regardless of response speed. Therefore, only accuracy was analyzed.

For the behavioral learning sessions, in order to examine the learning curves, data from each learning session were divided into 4 equal length consecutive blocks. Mean accuracy from each block of the 2 sessions were entered into a 2 (Modal: multisensory vs. unisensory) × 4 (Block: from 1 to 4) repeated measures ANOVA. The interaction between Modal and Block was not significant, $F(3, 57) = 0.45$, $P = 0.653$, and neither was the main effect of Modal, $F(1, 19) = 0.84$, $P = 0.369$, which indicated that there was no significant difference between the learning curves for multisensory category learning and unisensory category learning (Fig. 1B). This result was confirmed by t-tests, which showed that all comparisons between multisensory and unisensory learning across blocks were not significant: block 1, $t(19) = 0.31$, $P = 0.757$; block 2, $t(19) = 0.91$, $P = 0.370$; block 3, $t(19) = 0.39$, $P = 0.701$; block 4, $t(19) = 1.71$, $P = 0.102$. In addition, there was a significant main effect of Block, $F(3, 57) = 63.74$, $P < 0.001$. Because no Modal effect was found, data from the multisensory and unisensory sessions were pooled to reveal the effect of Block. Multiple comparisons with Bonferroni correction showed that accuracy on blocks 2, 3, and 4 was significantly higher than block 1, $Ps < 0.001$; accuracy on blocks 3 and 4 was significantly higher than block 2, $Ps < 0.001$; no significant difference was found between accuracy of block 3 and 4, $P = 0.528$. These results indicate that learning rate continually increased from block 1 to block 3, and reached asymptote during the second half of each learning session.

For the fMRI sessions, there were 2 tasks in each of the multisensory and unisensory sessions, which resulted in 4 conditions: categorization task in multisensory session (MultiCate), control task in multisensory session (MultiCon), categorization task in unisensory session (UniCate), and control task in unisensory session (UniCon). Mean accuracy for the 4 conditions was entered into a 2 (Modal: multisensory vs. unisensory) × 2 (Task: categorization vs. control) repeated measures ANOVA (Fig. 1C). The interaction between Modal and Task was significant, $F(1, 19) = 17.17$, $P = 0.001$, as well as the main effect of Modal, $F(1, 19) = 10.25$, $P = 0.005$, and the main effect of Task, $F(1, 19) = 105.82$, $P < 0.001$. Simple main effect analyses showed that, for the categorization tasks, the accuracy for unisensory categorization (91%) was significantly higher than multisensory categorization (85%), $t(19) = 3.77$, $P = 0.001$, while no significant difference was obtained between the accuracy of unisensory control (98%) and multisensory control (98%), $t(19) = 0.72$, $P = 0.477$. The accuracy of the control tasks was both significantly higher than the categorization tasks: for the multisensory session, $t(19) = 9.46$, $P < 0.001$; for the unisensory session, $t(19) = 5.92$, $P < 0.001$. This significant interaction, which mainly resulted from higher accuracy in unisensory categorization than in multisensory categorization, was unexpected given that accuracy in the 2 tasks was equivalent during the training session. To circumvent the confound of this unexpected accuracy difference between tasks, the accuracy for each task was entered as a covariate in fMRI analysis.

fMRI Results

Effects Revealed by Comparing Categorization Task to Control Task

To identify the neural correlates of II categorization and compare multisensory and unisensory conditions, our main analysis, GLM1, included 4 critical events (i.e., MultiCate, MultiCon, UniCate, and UniCon). Because we found higher accuracy for unisensory categorization than for multisensory categorization, we included the accuracy of each the 4 tasks as covariates in order to control for any confound of task difficulty.

First, we identified the network involved in both multisensory categorization and unisensory categorization by calculating the conjunction of "MultiCate > MultiCon" and "UniCate > UniCon". Conjunction analysis revealed a network including the left insula, right inferior frontal gyrus (IFG), supplementary motor area (SMA), left precentral gyrus, right superior parietal gyrus (SPG), and bilateral inferior parietal gyrus (IPG) (Fig. 2A blue regions and Table 1A).

Second, we identified brain regions specifically involved in multisensory categorization by calculating the interaction contrast "(MultiCate > MultiCon) > (UniCate > UniCon)". This contrast identified a region in the left IFG (−54, 15, 6), which was activated only in multisensory categorization task but not in unisensory categorization task (Fig. 2A red region). Activity in the left IFG is compared with activity in the right IFG, which was found to be sensitive to both multisensory and unisensory conditions, in Fig. 2B (note that neural activity was extracted from the activated regions only for illustrative purposes. Statistical analysis was not performed to avoid concerns about non-independence).

The left IFG area of activation spanned the opercular part (BA44) and triangular part (BA45) of the left IFG. ROI analysis extracted the neural activity from anatomically defined Brodmann areas across the 4 tasks and submitted it into 2 (Modal: multisensory vs. unisensory) × 2 (Task: categorization vs. control) repeated measures ANOVAs. The left BA44 and 45

Table 1 Results of GLM1. Effect of sensory modality in categorization

Regions	Side	t	voxels	x	y	z
(A) Sensory-modality-general regions "MultiCate > MultiCon" ∩ "UniCate > UniCon"						
Insula	L	5.13	40	−33	24	3
IFG	R	4.97	63	51	9	27
	R	3.77	23	36	33	18
Precentral gyrus	L	4.37	39	−42	3	39
SMA	M	4.42	35	−9	21	48
IPG	L	4.20	26	−39	−39	39
	R	3.97	37	42	−42	51
SPG	R	4.42	54	21	−63	51
(B) Multisensory-specific regions (MultiCate > MultiCon) > (UniCate > UniCon)						
IFG	L	5.33	76	−54	15	6

The coordinates (*x*, *y*, *z*) correspond to MNI coordinates. Clusters survived $P < 0.001$ at the voxel level and $P < 0.05$ with FDR correction at the cluster level. MultiCate: multisensory categorization task; MultiCon: multisensory control task; UniCate: unisensory categorization task; UniCon: unisensory categorization task.

both showed multisensory categorization specific activation (Supplementary Fig. S1). The interaction between Modal and Task was significant (left BA44, $F(1, 19) = 6.98$, $P = 0.016$; left BA45, $F(1, 19) = 7.09$, $P = 0.015$), which was driven by higher activity in multisensory categorization than unisensory categorization (left BA44, $t(19) = 2.92$, $P = 0.009$; left BA45, $t(19) = 2.61$, $P = 0.017$). In contrast, the same ROI analysis performed on the right homologous region showed that, the right BA44 and 45 were activated equivalently in multisensory and unisensory categorization (Supplementary Fig. S1): the interaction was not significant (right BA44, $F(1, 19) = 0.17$, $P = 0.685$; right BA45, $F(1, 19) = 1.92$, $P = 0.181$), while the main effect of Task was significant (right BA44, $F(1, 19) = 106.77$, $P < 0.001$; right BA45, $F(1, 19) = 21.95$, $P < 0.001$). These results demonstrated the specific role of the left IFG in integrating multisensory information during categorization.

To verify the main analyses, 2 control analyses were performed. First, in GLM1, all trials were included in each task, pooling the correct and incorrect responses together. Nevertheless, when including only the correct trials in each task (GLM3, see Materials and Methods), the above conjunction and interaction effect still held true (Supplementary Table S1). Second, although our stochastic presentation of feedback (i.e., feedback was presented only on half of the trials) and the jittering between feedback and stimulus maximally dissociated the neural estimation of categorization processing (or color judgment) from that of feedback processing, there is still a potential confound of feedback processing affecting measurement of activity during categorization. To rule out this possibility, a control analysis was performed by modeling only the trials in which no feedback was provided (half of all trials) (GLM4, see Materials and Methods). The results of the main analysis were replicated in this control analysis (Supplementary Table S2), which suggests that the effects in the main analysis were not due to feedback processing.

Third, since the basal ganglia has been long viewed as a key region in category learning and categorization (Seger 2008), we performed anatomically based ROI analyses on 8 subregions (right and left head of the caudate, body and tail of the caudate, putamen, and globus pallidus) (Fig. 3). Similar to GLM1 in the whole-brain analyses, we performed repeated measures ANOVAs with Modal and Task as factors. The left caudate body and tail (Fig. 4A) showed significant interaction between Modal and Task ($F(1, 19) = 4.86$, $P = 0.040$), with larger activation in multisensory categorization ($t(19) = 3.24$, $P = 0.004$) than in unisensory categorization ($t(19) = 2.10$, $P = 0.049$). No significant interaction was found in the right caudate body and tail ($F(1, 19) = 3.60$, $P = 0.073$), or other ROIs ($Fs < 1.51$, $Ps > 0.234$) (Fig. 4B–H). Although the left caudate body and tail did not survive in the whole-brain analysis of interaction effect with the standard threshold, we confirmed that when we used a more lenient threshold, an activation could be found in the tail of the caudate (peak at −12, −15, 21, $t = 2.74$, $P = 0.004$ uncorrected). In addition, a significant main effect of Task was found in the right caudate body/tail (Fig. 4B), right caudate head (Fig. 4D), and bilateral globus pallidus (Fig. 4E,F) (right caudate body/tail, $F(1, 19) = 19.03$, $P < 0.001$; right caudate head, $F(1, 19) = 7.03$, $P = 0.016$; left globus pallidus, $F(1, 19) = 21.66$, $P < 0.001$; right globus pallidus, $F(1, 19) = 13.65$, $P = 0.002$), which indicated equivalent activation in multisensory and unisensory categorization. For the left caudate head and bilateral putamen, neither main effect nor interaction was significant (Fig. 4C,G,H).

Effects Revealed by Comparing Correct Categorization to Incorrect Categorization

When contrasting correct and incorrect trials in the categorization task (GLM2, see Materials and Methods), the bilateral putamen showed greater activation in correct trials than in incorrect trials (Fig. 5G,H). Neural activity during categorization was submitted into 2 (Modal: multisensory categorization vs. unisensory categorization) × 2 (Correctness: correct vs. incorrect) repeated measures ANOVAs. Main effect of Correctness was significant for both the left and right putamen (left putamen, $F(1, 19) = 46.81$, $P < 0.001$; right putamen, $F(1, 19) = 56.45$, $P < 0.001$), with no significant interaction between Correctness and Modal nor main effect of Modal ($Fs < 1.14$, $Ps > 0.299$). These results demonstrated the different role of the basal ganglia nuclei: while the caudate and globus pallidus were involved in II categorization regardless of the correctness, putamen was related to the correctness of categorization.

Besides the putamen, the whole-brain analysis contrasting correct and incorrect trials in the categorization task identified areas related to correct categorization and incorrect categorization (GLM2, see Materials and Methods). Correct categorization was related to bilateral putamen and a medial region spanning the postcentral gyrus and the paracentral lobule (Fig. 6 upper panel, Table 2A). Incorrect categorization was related to bilateral

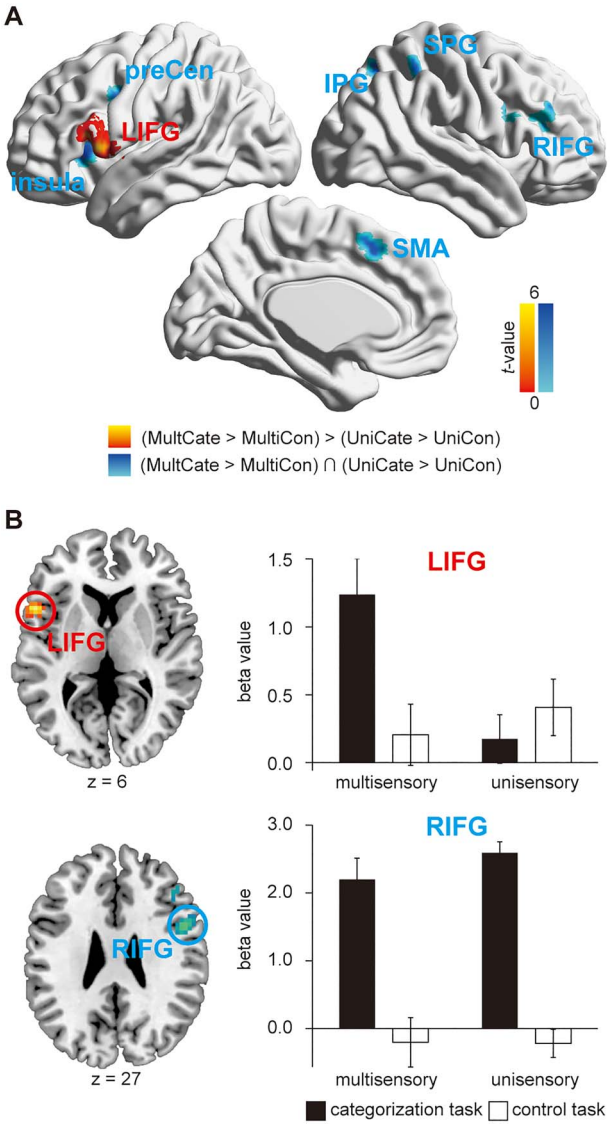

Figure 2. Brain activations revealed by comparing the categorization tasks with the control tasks. (*A*) Blue regions indicated the sensory-modality-general regions revealed by calculating the conjunction between the contrasts "MultiCate > MultiCon" and "UniCate > UniCon". Red region indicated the multisensory-specific region revealed by calculating interaction effect of (MultiCate > MultiCon) > (UniCate > UniCon). MultiCate: multisensory categorization task; MultiCon: multisensory control task; UniCate: unisensory categorization task; UniCon: unisensory categorization task; preCen: precentral gyrus. (*B*) Parameter estimates (beta values) of the 4 tasks were extracted from LIFG (the multisensory-specific region) and RIFG (an example of the sensory-modality-general regions) for illustrative purposes. Error bars indicate standard error of means.

insula and the anterior cingulate cortex (ACC) (Fig. 6 lower panel, Table 2*B*). No significant activation was found for the interaction effect between Modal and Correctness. These results were replicated in a control analysis that only modeled the trials in which no feedback was provided (GLM5, see Materials and Methods, Supplementary Table S3).

Discussion

In the present fMRI study, we systematically investigated for the first time the neural mechanisms underlying multimodal II category learning. We identified a corticostriatal network including the left inferior frontal gyrus (LEFG) and left body and tail of the caudate that was uniquely associated with multisensory categorization in comparison with unisensory categorization. We furthermore found that areas of the "core categorization network" identified in previous studies of unisensory visual categorization are also recruited for multisensory categorization, indicating that this network is independent of sensory modality.

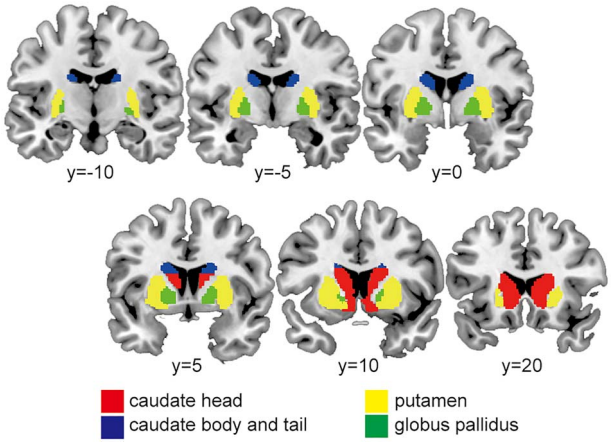

Figure 3. Anatomical ROIs. Red: head of caudate; blue: body and tail of caudate; yellow: putamen; green: globus pallidus. Numbers refer to the MNI coordinate of the slice along the Y-axis.

Left IFG-Caudate Network for Multisensory Categorization

In the present study, we identified a multisensory-specific region, the left IFG, which was specifically involved in multisensory II categorization (Fig. 2A red region, 2B upper panel). The left IFG has long been considered an important region related to audiovisual multisensory integration in studies of speech perception (Fernandez et al. 2017; McCormick et al. 2018; Miller and D'Esposito 2005; Tse et al. 2015) and in studies of working memory involving face and vocalization information (Hwang and Romanski 2015; Plakke et al. 2015). Recent research has extended the role of IFG in multisensory processing beyond the domain of language. Noppeney et al. (2010) found that the left IFG played a role in audiovisual information accumulation in a perceptual decision-making task. Statistical learning of sequential dependencies commonly recruits the IFG (Frost et al. 2015), and a recent human MEG study found that the left IFG also underlay the statistical learning of multisensory regularities (Paraskevopoulos et al. 2018). Moreover, the left IFG is anatomically and functionally connected with diverse cortical regions including the multimodal cortex and the motor cortex (Deacon 1992; Greenlee et al. 2007; Petrides 2005; Petrides and Pandya 2002; Skipper et al. 2007), which is consistent with the left IFG's function of associating multisensory representations with motor responses.

ROI analysis revealed that the left caudate body and tail also showed greater activation in multisensory categorization than in unisensory categorization (Fig. 4A). The pattern of activity matched the left frontal region in terms of sensitivity to task and interaction. The coactivation of these regions is consistent with the view that category learning relies on interacting corticostriatal loops (Seger and Miller 2010). The recruitment of

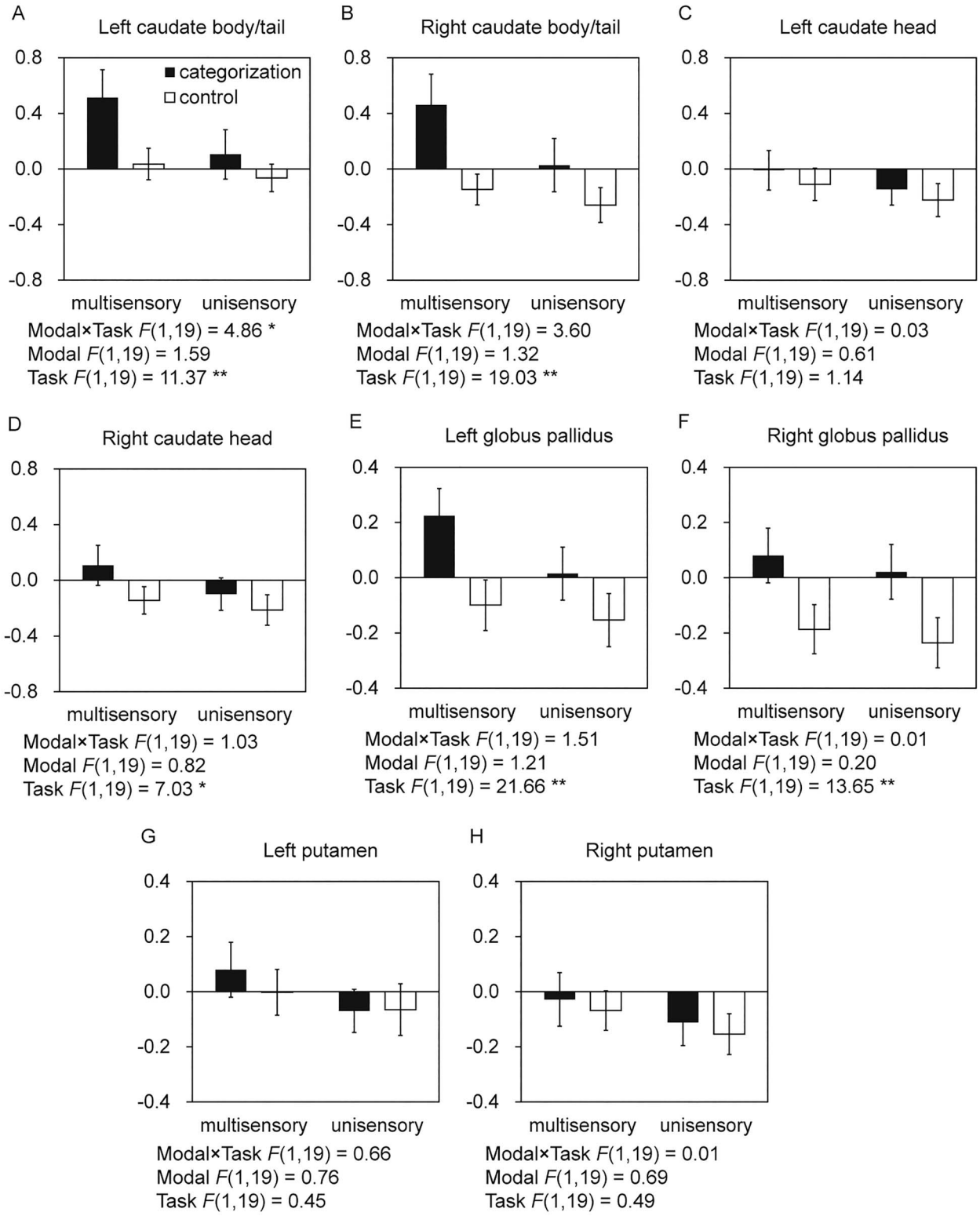

Figure 4. Basal ganglia ROI analysis of the effects of categorization versus control task and modality. Results are shown for the left and right caudate body and tail (*A*, *B*), caudate head (*C*, *D*), globus pallidus (*E*, *F*), and putamen (*G*, *H*). Parameter estimates (beta values) of the 4 tasks (multisensory categorization, multisensory control, unisensory categorization, and unisensory control) were extracted from the anatomical ROIs. Error bars indicate standard error of means. Results for repeated-measures 2 (Modal: multisensory vs. unisensory) ×2 (Task: categorization vs. control) ANOVAs are shown. Left caudate body and tail showed a significant interaction between Modal and Task, indicating that it was activated to a larger magnitude in multisensory categorization than in unisensory categorization. Right caudate body and tail, right caudate head, and bilateral globus pallidus showed significant main effect of Task, with non-significant interaction, indicating a role in integrating information in categorization, regardless of whether the information came from a single modality or from multiple modalities. No significant main effect or interaction was found for the putamen ROIs.

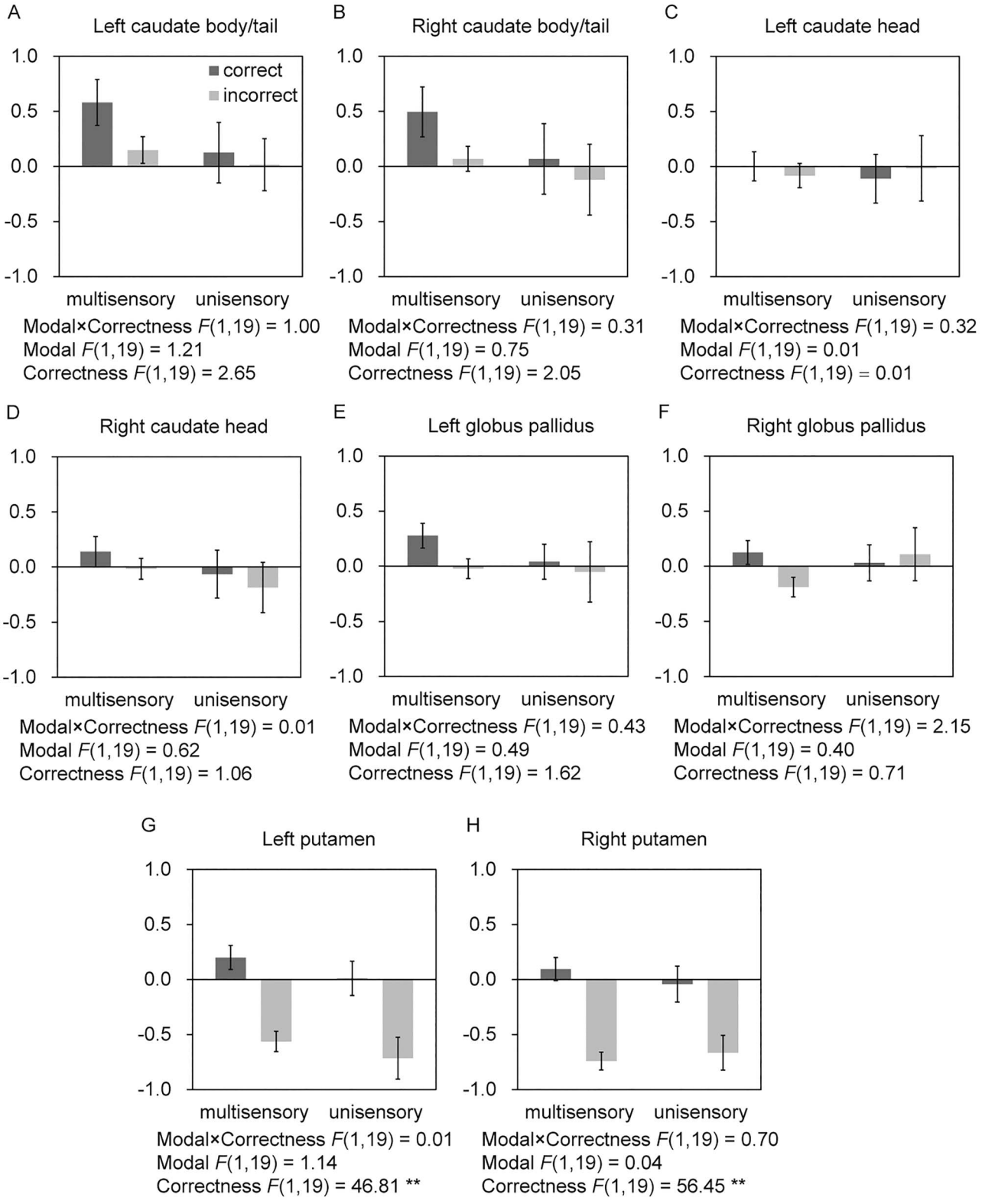

Figure 5. Basal ganglia ROI analysis of correct vs. incorrect categorization and modality. Results are shown for the left and right caudate body and tail (*A*, *B*), caudate head (*C*, *D*), globus pallidus (*E*, *F*), and putamen (*G*, *H*). Parameter estimates (beta values) of correct and incorrect trials in multisensory and unisensory categorization tasks were extracted from the anatomical ROIs. Error bars indicate standard error of means. Repeated-measures 2 (Modal: multisensory categorization vs. unisensory categorization) ×2 (Correctness: correct vs. incorrect) ANOVAs were performed. Bilateral putamen (*G*, *H*) showed significant main effects of Correctness, with non-significant interaction, indicating that putamen was related to the correctness of II categorization.

Table 2 Results of GLM2: correct versus incorrect categorization

Regions	Side	t	Voxels	x	y	z
(A) Correct related regions "MultiCorr > MultiIncorr" ∩ "UniCorr > UniIncorr"						
Putamen	L	243	5.95	−30	−15	6
	R	428	5.09	30	−9	9
Paracentral lobule	M	231	5.70	15	−39	72
(B) Incorrect related regions "MultiIncorr > MultiCorr" ∩ "UniIncorr > UniCorr"						
Insula	L	174	7.06	−30	24	−3
	R	259	6.76	39	21	6
ACC	M	286	5.27	9	24	36

The coordinates (x, y, z) correspond to MNI coordinates. Clusters survived $P < 0.001$ at the voxel level and $P < 0.05$ with FDR correction at the cluster level. MultiCorr: correct trials in multisensory categorization; MultiIncorr: incorrect trials in multisensory categorization; UniCorr: correct trials in unisensory categorization; UniIncorr: incorrect trials in unisensory categorization.

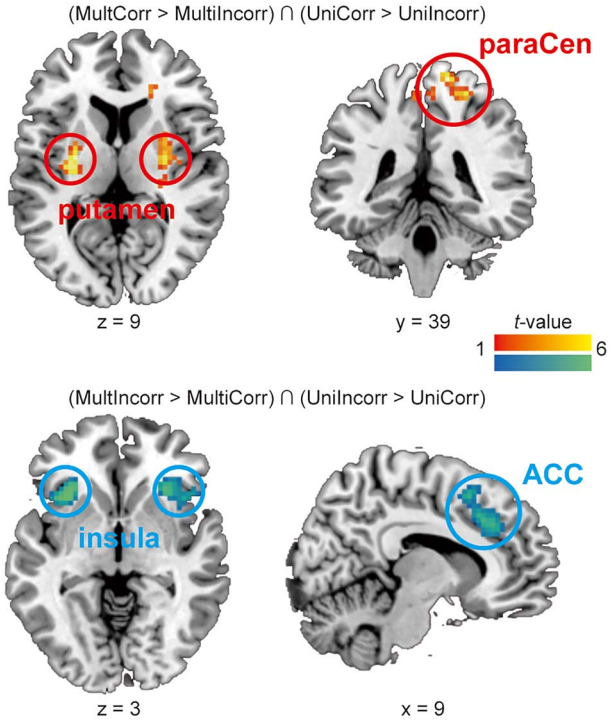

Figure 6. Brain activations revealed by contrasting correct and incorrect categorization trials. Upper red regions indicate correct categorization-related regions revealed by calculating the conjunction between the contrasts "MultiCorr > MultiIncorr" and "UniCorr > UniIncorr". Lower blue regions indicate incorrect categorization-related regions revealed by calculating the conjunction between the contrasts "MultiIncorr > MultiCorr" and "UniIncorr > UniCorr". MultiCorr: correct trials in multisensory categorization; MultiIncorr: incorrect trials in multisensory categorization; UniCorr: correct trials in unisensory categorization; UniIncorr: incorrect trials in unisensory categorization. ParaCen: paracentral gyrus.

the left body and tail of the caudate is consistent with known anatomical and functional connections of this region, including projections from the visual and auditory regions of the temporal lobe (Seger 2013), and interactions between the inferior frontal lobe and the body of the caudate (Jeon et al. 2014).

Sensory-System Independent Categorization Network

We found activity in areas previously associated with unimodal visual categorization were also active in multimodal audiovisual categorization. Previous neuroimaging studies examining unimodal II categorization also found activation in diverse cortical regions including insular, inferior frontal, anterior cingulate/SMA, and inferior and superior parietal cortices bordering the intraparietal sulcus (Carpenter et al. 2016; Cincotta and Seger 2007; Nomura et al. 2007; Seger et al. 2015; Seger and Cincotta 2002; Waldschmidt and Ashby 2011). Consistent with these findings, the present study identified a sensory-modality-general II network including the insula, IFG, SMA, precentral gyrus, and parietal cortex (Fig. 2A blue regions, 2B lower panel). Extending previous findings, the present study found that these regions were similarly activated in both multisensory II categorization and unisensory II categorization, which suggested that they were involved in categorization II not only when the information came from a single modality but also when they came from multiple modalities.

We were further able to identify categorization-related regions via multiple contrasts, and found different patterns of results depending on the conditions we compared. Some past studies compared categorization to a control baseline task (Cincotta and Seger 2007; Seger and Cincotta 2002), whereas other studies compared correct categorization with incorrect categorization (Nomura et al. 2007). In the present study we employed both approaches, which provided additional insights into the roles of these areas (Carpenter et al. 2016). The whole-brain analysis contrasting correct and incorrect trials showed that the putamen and postcentral gyrus were more active for correct categorization whereas the insula and ACC were related to incorrect categorization (Fig. 6). The putamen and sensorimotor cortical regions interact within the motor corticostriatal loop (Seger 2008), and this network is most strongly associated with selection of appropriate motor behaviors. Our results are consistent with previous research finding that the putamen is associated with a variety of measurements of response strength and response selection (Brovelli et al. 2011; McNamee et al. 2015; Peterson and Seger 2013). The insula and ACC are known to be functionally connected via the ventral attention or salience intrinsic connectivity network (Seeley et al. 2007). Greater activity for incorrect trials may reflect the known monitoring functions of this network, which are often reflected in increased activity when errors are detected (Ham et al. 2013), and/or increased attentional demands on difficult trials that are more likely to result in errors (Seger et al. 2015).

Role of the Basal Ganglia in Multisensory Category Learning

Different regions of the basal ganglia have been posited to contribute to different aspects of category learning, including

the head of the caudate, body and tail of the caudate, and putamen (Ashby et al. 1998; Ashby and Maddox 2005, 2011; Seger 2008). Anatomical ROI analyses showed that sub-regions of the basal ganglia were associated with different aspects of II categorization. As discussed above, the left body and tail of the caudate nucleus was uniquely associated with multisensory categorization in contrast with unisensory categorization. In addition, the right caudate body and tail, right caudate head, and bilateral globus pallidus were activated in both unimodal and multimodal II categorization tasks as compared to control tasks (Fig. 4B,D–F). In contrast, the putamen was not sensitive to categorization overall, but rather activity in the putamen was related to the correctness of II categorization with significantly higher activity in correct categorization than in incorrect categorization (Fig. 4G,H, 5G,H). As discussed above, the greater association of putamen with correct responding may reflect the role of this structure in response prediction and selection within stimulus-response learning (Brovelli et al. 2011; McNamee et al. 2015; Peterson and Seger 2013).

Caveats and Limitations

In this study, we chose to pretrain the participants before MRI scanning. Although some studies perform MRI scanning at the initial phase of category learning (Carpenter et al. 2016; Cincotta and Seger 2007; Nomura et al. 2007), other studies, as well as the present one, pretrained the participants before MRI scanning (Reber et al. 2003; Seger et al. 2015; Seger and Cincotta 2002; Turner et al. 2017). Pretraining had the advantage that, by screening participants based on their behavioral performance in the pretraining session, we could guarantee that every participant who was scanned was able to learn the artificial categories. Previous studies of multisensory II category learning found that participants showed substantial individual variation in performance accuracy and strategy used. For example, Smith et al. found that 50% and 36% of the participants in their experiment 1 and 2, respectively, used RB or guessing strategies when learning multisensory II categories (Smith et al. 2014). In the present fMRI study, this proportion was about 30%, that is, 10 participants were classified as RB strategy users in their first behavioral learning session and were excluded. All the 20 participants in the final sample were classified by the computational models as using an II strategy in both the multisensory learning and unisensory learning sessions. In addition, pretraining allowed us to avoid scanning during the early phases of acquisition, during which participants typically have widely varying accuracy levels and experiment with different categorization strategies including RB strategies (Ashby et al. 1998). Through the use of this pretraining and screening procedure, we intended to increase the homogeneity of cognitive processes across participants, which can significantly increase the efficiency and power of the fMRI experiments. However, one possible concern is that the pretraining might have resulted in our assessing performance that is past the learning phase and that has shifted into a habitual or automatic performance phase. Distinct differences have been recognized between well-learned categories (e.g., faces or houses) and early learning of new arbitrary categories (Mahon and Caramazza 2009; Martin 2007; Seger and Miller 2010). Although it is not easy to determine the exact transition point from the initial learning phase to the automatic categorization phase, recent studies have examined performance across extended training using objective measures of automaticity in category learning (Helie et al. 2010; Waldschmidt and Ashby 2011). Analysis of categorization performance for more than 11000 trials spread over 20 different training sessions showed that, in II categorization, automaticity occurred after about 5000 trials. In the present study, the 600-trial pretraining fell well short of this number of trials, and was similar in extent to other studies examining performance at relatively early stages of training (Seger et al. 2015; Turner et al. 2017).

An unexpected result observed in the present study was the accuracy difference between multisensory categorization and unisensory categorization in the MRI scanning sessions. In our pilot experiments, with the intention to match the learning performance of multisensory and unisensory categories, stimulus parameters were adjusted so that the multisensory and unisensory categories could be learned equivalently well. Our results in the behavioral sessions also confirmed that the learning rate for multisensory categories did not differ from that for unisensory categories (Fig. 1B, in the last block of the behavioral session, $87 \pm 5\%$ for multisensory and $90 \pm 5\%$ for unisensory). However, when performing categorization inside the MRI scanner on the next day after the behavioral learning session, the accuracy of multisensory categorization ($85 \pm 6\%$) significantly differed from that of unisensory categorization ($91 \pm 6\%$) (Fig. 1C). One possible cause of the decreasing multisensory accuracy might be the influence of the MRI scanning noise. Although the frequency range of the auditory stimulus was intentionally selected (200–392 Hz range) to be distinctive from the background EPI noise (~1500 Hz) (Huang et al. 2015; Ravicz et al. 2000), the MRI background noise might still have led to difficulties in processing the auditory stimulus. To control for this unexpected confound, we added the accuracy of each task into the group-level GLM as a covariate.

The present study investigated multisensory II category learning in a visuo-auditory paradigm. It is unknown whether multimodal category learning involving other sensory modalities would also recruit the IFG, so the interpretation of the present multisensory results should be limited to visuo-auditory II categorization. Unimodal category learning has been investigated in other sensory modalities, such as haptics (Gaissert et al. 2012; Schwarzer et al. 1999) and olfaction (Qu et al. 2016). Future studies should be conducted to directly examine whether the IFG is engaged in integrating information from those modalities during categorization (e.g., visuo-haptic and visuo-olfactory II category learning). Another potential limitation is the relatively small sample size of the present study (i.e., $N = 20$). This was mainly because we adopted a multiple-session design that required each participant to complete 4 sessions. To reduce the heterogeneity between participants, we used decision bound model analysis to ensure that all participants used the same adaptive strategy (i.e., the II categorization strategy). However, future studies might adopt a larger sample size and might test a broader range of participants, such as those who used the nonadaptive RB strategy in II categorization.

Conclusion

In conclusion, this study investigated for the first time how multisensory information is integrated in neural systems during multisensory II category learning. We found that multisensory II categorization was uniquely associated with activity in a corticostriatal network including the left IFG and the left body and tail of the caudate. These results support theories that the left IFG plays a broad role in audiovisual integration not

limited to the linguistic domain. We furthermore found that areas previously identified as important for unisensory visual categorization are also recruited for multisensory categorization, indicating that the core categorization network is modality independent and can integrate across sensory modalities.

Supplementary Material

Supplementary material can be found at *Cerebral Cortex* online.

Conflict of Interest

None declared.

Funding

The National Natural Science Foundation of China (31700975, 31871138, 31671108, 31960183); the National Social Science Foundation of China (19ZDA360); China Postdoctoral Science Foundation; Changjiang Scholars Program of Chinese Ministry of Education (to C.S., 2015; to Q.C., 2016); Program for New Century Excellent Talents in University of China (NCET-12-0645); Shenzhen Overseas Innovation Team Project (KQTD20140630180249366); International Partnership Program of Chinese Academy of Sciences (172644KYSB20160175); and The Scientific Instrument Innovation Team of Chinese Academy of Sciences (GJJSTD20180002).

References

Ashby FG, Alfonso-Reese LA, Turken AU, Waldron EM. 1998. A neuropsychological theory of multiple systems in category learning. *Psychol Rev*. 105(3):442–481.

Ashby FG, Ell SW. 2001. The neurobiology of human category learning. *Trends Cognit Sci*. 5(5):204–210.

Ashby FG, Gott RE. 1988. Decision rules in the perception and categorization of multidimensional stimuli. *J Exp Psychol Learn Mem Cognit*. 14(1):33–53.

Ashby FG, Maddox WT. 2005. Human category learning. *Annu Rev Psychol*. 56:149–178.

Ashby FG, Maddox WT. 2011. Human category learning 2.0. *Ann N Y Acad Sci*. 1224:147–161.

Brett M, Anton JL, Valabregue R, Poline JB. 2002. Region of interest analysis using an spm toolbox. *Paper presented at: the 8th International Conference on Functional Mapping of the Human Brain*. Sendai, Japan.

Brovelli A, Nazarian B, Meunier M, Boussaoud D. 2011. Differential roles of caudate nucleus and putamen during instrumental learning. *NeuroImage*. 57(4):1580–1590.

Cantwell G, Crossley MJ, Ashby FG. 2015. Multiple stages of learning in perceptual categorization: evidence and neurocomputational theory. *Psychon Bull Rev*. 22(6):1598–1613.

Carpenter KL, Wills AJ, Benattayallah A, Milton F. 2016. A comparison of the neural correlates that underlie rule-based and information-integration category learning. *Hum Brain Mapp*. 37(10):3557–3574.

Cincotta CM, Seger CA. 2007. Dissociation between striatal regions while learning to categorize via feedback and via observation. *J Cognit Neurosci*. 19(2):249–265.

Deacon TW. 1992. Cortical connections of the inferior arcuate sulcus cortex in the macaque brain. *Brain Res*. 573(1): 8–26.

Ell SW, Ashby FG. 2012. The impact of category separation on unsupervised categorization. *Atten Percept Psychophys*. 74(2):466–475.

Fernandez LM, Macaluso E, Soto-Faraco S. 2017. Audiovisual integration as conflict resolution: the conflict of the mcgurk illusion. *Hum Brain Mapp*. 38(11):5691–5705.

Frost R, Armstrong BC, Siegelman N, Christiansen MH. 2015. Domain generality versus modality specificity: the paradox of statistical learning. *Trends Cogn Sci*. 19(3):117–125.

Gaissert N, Waterkamp S, Fleming RW, Buelthoff I. 2012. Haptic categorical perception of shape. *PLoS One*. 7(8).

Greenlee JDW, Oya H, Kawasaki H, Volkov IO, Severson MA III, Howard MA III, Brugge JF. 2007. Functional connections within the human inferior frontal gyrus. *J Comp Neurol*. 503(4):550–559.

Ham T, Leff A, de Boissezon X, Joffe A, Sharp DJ. 2013. Cognitive control and the salience network: an investigation of error processing and effective connectivity. *J Neurosci*. 33(16):7091–7098.

Helie S, Waldschmidt JG, Ashby FG. 2010. Automaticity in rule-based and information-integration categorization. *Atten Percept Psychophys*. 72(4):1013–1031.

Huang S, Li Y, Zhang W, Zhang B, Liu X, Mo L, Chen Q. 2015. Multisensory competition is modulated by sensory pathway interactions with fronto-sensorimotor and default-mode network regions. *J Neurosci*. 35(24):9064–9077.

Hwang J, Romanski LM. 2015. Prefrontal neuronal responses during audiovisual mnemonic processing. *J Neurosci*. 35(3): 960–971.

Jeon H-A, Anwander A, Friederici AD. 2014. Functional network mirrored in the prefrontal cortex, caudate nucleus, and thalamus: high-resolution functional imaging and structural connectivity. *J Neurosci*. 34(28):9202–9212.

Lim S-J, Fiez JA, Holt LL. 2014. How may the basal ganglia contribute to auditory categorization and speech perception? *Front Neurosci*. 8:230.

Lim S-J, Fiez JA, Holt LL. 2019. Role of the striatum in incidental learning of sound categories. *Proc Natl Acad Sci USA*. 116(10):4671–4680.

Maddox WT, Ing AD, Lauritzen JS. 2006. Stimulus modality interacts with category structure in perceptual category learning. *Percept Psychophys*. 68(7):1176–1190.

Mahon BZ, Caramazza A. 2009. Concepts and categories: a cognitive neuropsychological perspective. *Annu Rev Psychol*. 60:27–51.

Martin A. 2007. The representation of object concepts in the brain. *Annu Rev Psychol*. 58:25–45.

McCormick K, Lacey S, Stilla R, Nygaard LC, Sathian K. 2018. Neural basis of the crossmodal correspondence between auditory pitch an visuospatial elevation. *Neuropsychologia*. 112: 19–30.

McNamee D, Liljeholm M, Zika O, O'Doherty JP. 2015. Characterizing the associative content of brain structures involved in habitual and goal-directed actions in humans: a multivariate fmri study. *J Neurosci*. 35(9):3764–3771.

Miller LM, D'Esposito M. 2005. Perceptual fusion and stimulus coincidence in the cross-modal integration of speech. *J Neurosci*. 25(25):5884–5893.

Nagy A, Eordegh G, Paroczy Z, Markus Z, Benedek G. 2006. Multisensory integration in the basal ganglia. *Eur J Neurosci*. 24(3):917–924.

Nomura EM, Maddox WT, Filoteo JV, Ing AD, Gitelman DR, Parrish TB, Mesulam MM, Reber PJ. 2007. Neural correlates of

rule-based and information-integration visual category learning. *Cereb Cortex*. 17(1):37–43.
Noppeney U, Ostwald D, Werner S. 2010. Perceptual decisions formed by accumulation of audiovisual evidence in prefrontal cortex. *J Neurosci*. 30(21):7434–7446.
Paraskevopoulos E, Chalas N, Kartsidis P, Wollbrink A, Bamidis P. 2018. Statistical learning of multisensory regularities is enhanced in musicians: an meg study. *NeuroImage*. 175:150–160.
Peterson EJ, Seger CA. 2013. Many hats: Intratrial and reward level-dependent bold activity in the striatum and premotor cortex. *J Neurophysiol*. 110(7):1689–1702.
Petrides M. 2005. Lateral prefrontal cortex: architectonic and functional organization. *Philos Trans R Soc B-Biol Sci*. 360(1456):781–795.
Petrides M, Pandya DN. 2002. Comparative cytoarchitectonic analysis of the human and the macaque ventrolateral prefrontal cortex and corticocortical connection patterns in the monkey. *Eur J Neurosci*. 16(2):291–310.
Plakke B, Hwang J, Romanski LM. 2015. Inactivation of primate prefrontal cortex impairs auditory and audiovisual working memory. *J Neurosci*. 35(26):9666–9675.
Qu LP, Kahnt T, Cole SM, Gottfried JA. 2016. De novo emergence of odor category representations in the human brain. *J Neurosci*. 36(2):468–478.
Ravicz ME, Melcher JR, Kiang NYS. 2000. Acoustic noise during functional magnetic resonance imaging. *J Acoust Soc Am*. 108(4):1683–1696.
Reber PJ, Gitelman DR, Parrish TB, Mesulam MM. 2003. Dissociating explicit and implicit category knowledge with fmri. *J Cognit Neurosci*. 15(4):574–583.
Reig R, Silberberg G. 2014. Multisensory integration in the mouse striatum. *Neuron*. 83(5):1200–1212.
Schwarzer G, Kufer I, Wilkening F. 1999. Learning categories by touch: on the development of holistic and analytic processing. *Memory & Cognition*. 27(5):868–877.
Seeley WW, Menon V, Schatzberg AF, Keller J, Glover GH, Kenna H, Reiss AL, Greicius MD. 2007. Dissociable intrinsic connectivity networks for salience processing and executive control. *J Neurosci*. 27(9):2349–2356.
Seger CA. 2008. How do the basal ganglia contribute to categorization? Their roles in generalization, response selection, and learning via feedback. *Neurosci Biobehav Rev*. 32(2):265–278.
Seger CA. 2013. The visual corticostriatal loop through the tail of the caudate: circuitry and function. *Front Syst Neurosci*. 7:104–104.
Seger CA. 2018. Corticostriatal foundations of habits. *Curr Opin Behav Sci*. 20:153–160.
Seger CA, Braunlich K, Wehe HS, Liu Z. 2015. Generalization in category learning: the roles of representational and decisional uncertainty. *J Neurosci*. 35(23):8802–8812.
Seger CA, Cincotta CM. 2002. Striatal activity in concept learning. *Cogn Affect Behav Neurosci*. 2(2):149–161.
Seger CA, Cincotta CM. 2005. The roles of the caudate nucleus in human classification learning. *J Neurosci*. 25(11):2941–2951.
Seger CA, Miller EK. 2010. Category learning in the brain. *Annu Rev Neurosci*. 33:203–219.
Skipper JI, Goldin-Meadow S, Nusbaum HC, Small SL. 2007. Speech-associated gestures, broca's area, and the human mirror system. *Brain Lang*. 101(3):260–277.
Smith JD, Johnston JJR, Musgrave RD, Zakrzewski AC, Boomer J, Church BA, Ashby FG. 2014. Cross-modal information integration in category learning. *Atten Percept Psychophys*. 76(5):1473–1484.
Tse C-Y, Gratton G, Garnsey SM, Novak MA, Fabiani M. 2015. Read my lips: brain dynamics associated with audiovisual integration and deviance detection. *J Cognit Neurosci*. 27(9):1723–1737.
Turner BO, Crossley MJ, Ashby FG. 2017. Hierarchical control of procedural and declarative category-learning systems. *NeuroImage*. 150:150–161.
Waldschmidt JG, Ashby FG. 2011. Cortical and striatal contributions to automaticity in information-integration categorization. *NeuroImage*. 56(3):1791–1802.
Yi H-G, Maddox WT, Mumford JA, Chandrasekaran B. 2016. The role of corticostriatal systems in speech category learning. *Cereb Cortex*. 26(4):1409–1420.