Cerebral Cortex, 2021;00: 1–21

https://doi.org/10.1093/cercor/bhab324
Original Article

ORIGINAL ARTICLE

Reversible Fronto-occipitotemporal Signaling Complements Task Encoding and Switching under Ambiguous Cues

Kaho Tsumura[1], Keita Kosugi[1], Yoshiki Hattori[1], Ryuta Aoki[2], Masaki Takeda[2], Junichi Chikazoe[3], Kiyoshi Nakahara[2] and Koji Jimura[1,2]

[1]Department of Biosciences and Informatics, Keio University, Yokohama 223-0061, Japan, [2]Research Center for Brain Communication, Kochi University of Technology, Kami 782-8502, Japan and [3]Supportive Center for Brain Research, National Institute for Physiological Sciences, Okazaki 444-8585, Japan

Address correspondence to Koji Jimura, Department of Biosciences and Informatics, Keio University, 3-14-1 Hiyoshi Kohoku-ku, Yokohama 223-0061, Japan. Email: koji.jimura@gmail.com

Abstract

Adaptation to changing environments involves the appropriate extraction of environmental information to achieve a behavioral goal. It remains unclear how behavioral flexibility is guided under situations where the relevant behavior is ambiguous. Using functional brain mapping of machine learning decoders and directional functional connectivity, we show that brain-wide reversible neural signaling underpins task encoding and behavioral flexibility in ambiguously changing environments. When relevant behavior is cued ambiguously during behavioral shifting, neural coding is attenuated in distributed cortical regions, but top-down signals from the prefrontal cortex complement the coding. When behavioral shifting is cued more explicitly, modality-specialized occipitotemporal regions implement distinct neural coding about relevant behavior, and bottom-up signals from the occipitotemporal region to the prefrontal cortex supplement the behavioral shift. These results suggest that our adaptation to an ever-changing world is orchestrated by the alternation of top-down and bottom-up signaling in the fronto-occipitotemporal circuit depending on the availability of environmental information.

Key words: behavioral flexibility, deep neural network, occipitotemporal cortex, prefrontal cortex, task cue

Introduction

Executive control guides flexible adaptation to changing environments and is most developed in humans throughout evolution (Miller and Cohen 2001; Stoet and Snyder 2009). Shifting between different types of behavior is one of the core executive control functions (Allport et al. 1994; Rogers and Monsell 1995), and task switching paradigms have been often used to investigate behavioral flexibility and its underlying neural mechanisms. Previous neuropsychological and neuroimaging studies of human and nonhuman animals suggest a critical role of the prefrontal cortex in switching tasks and rules (Dove et al. 2000; Rushworth et al. 2002; Bunge et al. 2005; Derrfuss et al. 2005; Crone et al. 2006; Yeung et al. 2006; Kim et al. 2012; Nee and D'Esposito 2016; Bissonette and Roesch 2017; Malagon-Vina et al. 2018; Fouragnan et al. 2019; Fig. 1A).

Importantly, executive control depends on perceived information of external environments, and relevant information is appropriately extracted from the external environment to achieve a behavioral goal. Perception of sensory information from the external environment guides the course of action, which is referred to as perceptual decision-making (Gold and Shadlen 2007; Hanks and Summerfield 2017). It involves the extraction of goal-relevant information, which is integrated to form a relevant decision. Studies of perceptual decision-

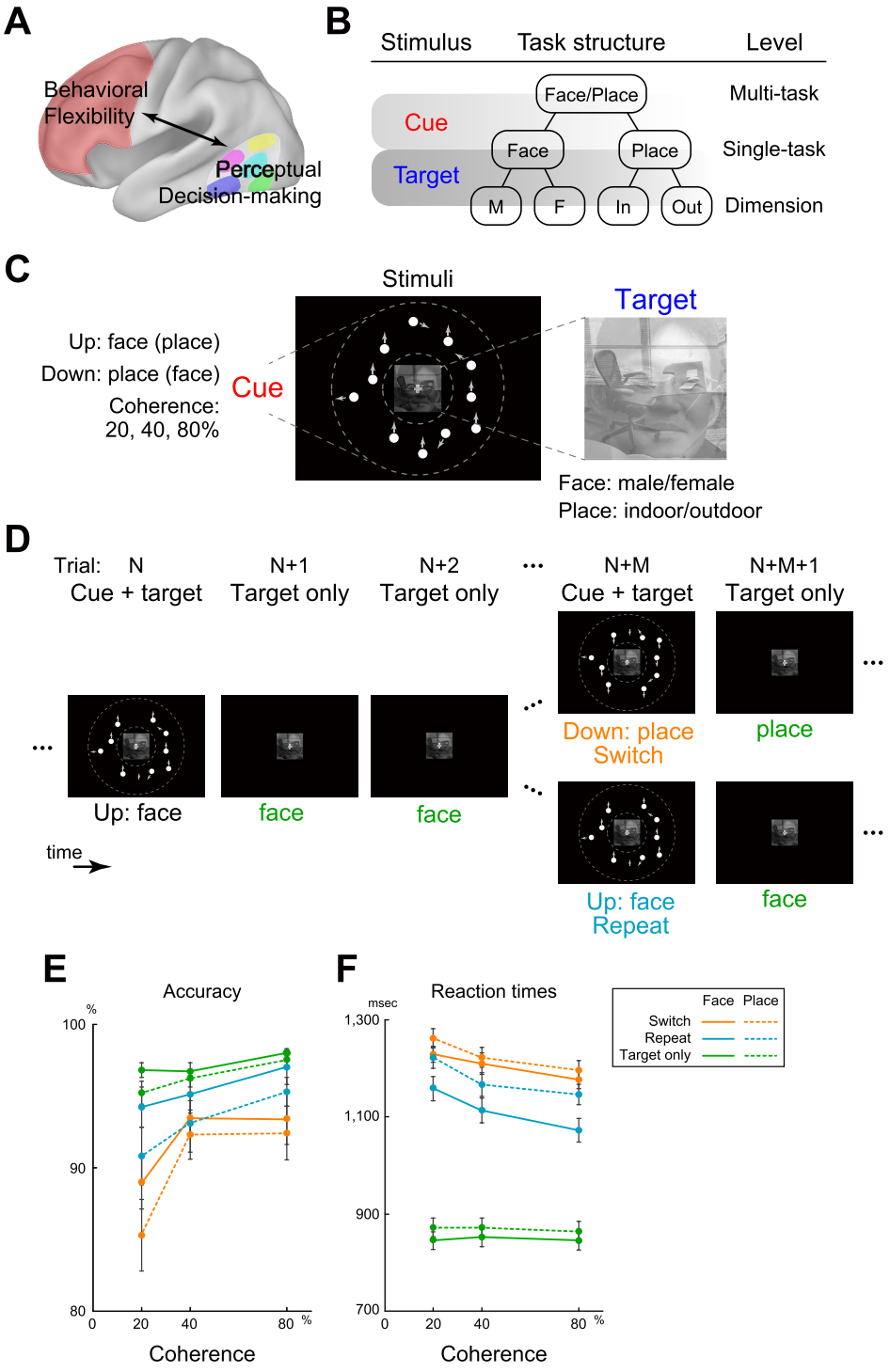

Figure 1. Experimental design and behavioral results. (*A*) Schematic illustration of a brain-wide model of interaction of behavioral flexibility and perceptual decision-making. Behavioral flexibility associated with prefrontal areas is subject to appropriate perception of the external world implemented in stimulus-modality-dependent occipitotemporal areas. (*B*) Hierarchical structure of task sets. Task sets are presented in a 3-tiered decision tree. Behavioral flexibility is guided by the cue stimulus, which indicates relevant behavior (judging faces or places) and bridges the upper layers in the hierarchical structure. Based on the task to be performed, target stimulus is judged [male (M) or female (F) in the face task; indoor (In) or outdoor (Out) in the place task]. (*C*, *D*) Behavioral task. (*C*) Stimuli. Cue stimulus indicates the task to be performed (face or place task) and is composed of a set of white moving dots presented within a donut-shaped display circle (indicated by dotted lines). The arrow indicates the motion direction of each dot, and overall motion was either upward or downward. Upward motion and downward motion indicate face and place tasks, respectively. Motion strength of the cue stimulus was manipulated by coherence of dot motion. The target image was superimposed picture of face and place, which was presented at the center of screen. Participants judged whether the face picture was male or female, or the place is indoor or outdoor, depending on the task to be performed. (*D*) Task procedure. Trials with simultaneous presentation of a dot cue and target (switch and repeat; N and N + M th trials) were followed by target-only trials (N + 1, N + 2 and N + M + 1 th trials). Switch and repeat trials were dependent on the performed task prior to the cue presentation. (*E*, *F*) Behavioral results. Accuracy (*E*) and RTs (*F*) as a function of motion coherence. Solid and dotted lines indicate face and place tasks, respectively, and orange, blue, and green lines indicate switch, repeat, and target-only trials, respectively. Error bars indicate standard error of the mean across participants.

making have used behavioral tasks that demand discrimination of sensory stimuli involving perceptual uncertainty (Newsome and Pare 1988; Corbetta et al. 1991; Kurikawa et al. 2018).

By manipulating perceptual uncertainty, neurophysiological and neuroimaging studies have examined cortical mechanisms of the perceptual decision-making. For example, the middle temporal (MT) region is known to play an important role in the perception of moving stimuli (Newsome and Pare 1988; Shadlen et al. 1996; Beauchamp et al. 1997; Huk et al. 2002; Kayser et al. 2010). It has been also suggested that the fusiform face area (FFA) and parahippocampal place area (PPA) are associated with the perception of face (Kanwisher et al. 1997; McCarthy et al. 1997; Ishai et al. 1999; Gazzaley et al. 2005; Freiwald and Tsao 2010) and place (Epstein and Kanwisher 1998; Ishai et al. 1999; Gazzaley et al. 2005) stimuli, respectively. These collective results suggest that temporal and occipital regions play important roles in perceptual decision-making and are functionally segmented depending on the modality of the stimulus (Fig. 1A).

In our daily life, goal-relevant information in external environments is not always evident. Such situation can be well illustrated by an incorporation of behavioral shifting and perceptual decision-making. Prior studies have explored neural mechanisms during task switching under situations where target stimuli involved perceptual uncertainty (Kayser et al. 2010; Mante et al. 2013; Zhang et al. 2013; Kumano et al. 2016). More recently, we showed that increased uncertainty of target stimulus engaged top-down signals from prefrontal to occipitotemporal cortices, which complemented task switching in such situations (Tsumura et al. 2021).

Notably, the task switching paradigm is composed of hierarchically structured configuration of a set of task rules, called task sets (Koechlin et al. 2003; Bunge et al. 2005; Badre 2008; Jimura and Braver 2010; Fig. 1B). Those prior studies above manipulated perceptual uncertainty of the target items within the lower layers of the task set hierarchy (Kayser et al. 2010; Mante et al. 2013; Zhang et al. 2013; Kumano et al. 2016; Tsumura et al. 2021). In other words, the target stimulus to be discriminated was presented ambiguously, but the task per se was cued without ambiguity. To our knowledge, it remains unclear how task switch is achieved when the relevant task is indicated by a cue involving perceptual uncertainty. As such, the uncertainty of relevant task information in the upper layers of the task-set hierarchy may provide a novel and important opportunity to examine the relationships between executive control and perceptual decision-making (Fig. 1A,B).

One potential approach to elucidate underlying neural mechanisms is to identify the signal contents of responsible brain regions; this has recently been demonstrated for perception by neural decoding techniques (e.g., Kamitani and Tong 2005; Haynes and Rees 2006; Norman et al. 2006). In particular, prior neuroimaging studies have shown that mental and behavioral states can be decoded from neural coding by machine learning techniques that classify distributed patterns of brain activity (Kamitani and Tong 2005; Haynes and Rees 2006; Norman et al. 2006; Nishimoto et al. 2011; Loose et al. 2017; Qiao et al. 2017; Chikazoe et al. 2019; Wang et al. 2020). One commonly used technique is the support vector machine (SVM; Vapnik 1998; Kamitani and Tong 2005; Misaki et al. 2010; Jimura and Poldrack 2012; Nakahara et al. 2016), which enables categorical discrimination by dividing multidimensional space composed of brain activation pattern using a linear hyperplane. Convolutional neural network (CNN) classifier (Krizhevsky et al. 2012; LeCun et al. 2015) is one of the deep neural network classifiers consisting of multiple feature-aggregating layers, enabling more robust classification. Importantly, recent technical advancements of CNN allow mapping that highlights image locations characterizing a classified image (Selvaraju et al. 2017). The mapping technique may provide novel information about neural coding and functional localization of the brain.

The current study aimed to elucidate relationships between behavioral flexibility and perceptual decision-making under cue uncertainty and to explore the underlying neural mechanisms (Fig. 1A,B). Functional magnetic resonance imaging (fMRI) was administered while human participants performed a task-switching paradigm with a cue involving perceptual uncertainty. Standard univariate analysis identified brain regions associated with task switching, motion strength, and task modality. In order to elucidate causal network dynamics during task switching under cue ambiguity, we examined effective connectivity among the task-related brain regions. Finally, whole-brain exploratory analyses based on machine learning techniques, CNN and SVM, were performed to identify brain regions that coded relevant task information.

Materials and Methods

Participants

Written informed consent was obtained from 30 healthy right-handed subjects (age range: 18–22; 11 females). Experimental procedures were approved by the institutional review board of Keio University and Kochi University of Technology. Participants received 2000 yen for each of the training and scanning sessions. One participant was excluded from analyses due to low behavioral performance; accuracy was lower than 30% in one of the experimental conditions. The number of participants was determined prior to the collection of the current data based on the effect sizes in pilot behavioral experiments and our previous relevant study (Tsumura et al. 2021).

Behavioral Procedures

The experiment consisted of 2 sessions administered on separate days. The first day was a training session, in which participants practiced discrimination tasks (random dot motion; Tsumura et al. 2021) and switching between 2 tasks (face and place tasks; see below for more details). On the second day, while fMRI scanning was administered, the participants performed the switching paradigm identical to those practiced in the training sessions.

Stimuli

All stimuli were generated in Matlab version 2012a, using the Psychophysics Toolbox (Brainard 1997) extension version 3.0.10, and were visually presented on a computer screen. The current task cue stimuli were randomly moving dot stimuli similar to those used in previous studies of perceptual decision-making (Chen et al. 2015; Tsumura et al. 2021). Each motion stimulus involved 60 dots moving inside a donut-shaped display patch with a white cross in the center of the patch on a black background (Fig. 1C). The display patch and cross were centered on the screen and extended from 6 to 12 degrees of visual angle (dva). Within the display patch, every dot moved at the speed of 10 dva/s. Some dots moved coherently toward one direction (upward or downward) while the others moved randomly. The

percentage of coherently moving dots determined the "motion coherence," which was set to 3 levels (20, 40, and 80%).

Dot presentation was controlled to remove local motion signals following a standard method for generating motion stimuli (Newsome and Pare 1988; Britten et al. 1993; Palmer et al. 2005; Chen et al. 2015; Tsumura et al. 2021). Namely, upon stimulus onset, the dots were presented at new random locations on each of the first 3 frames. They were relocated after 2 subsequent frames, such that the dots in frame 1 were repositioned in frame 4, and the dots in frame 2 were repositioned in frame 5, and so on. When repositioned, each dot was either randomly presented at the new location or aligned with the predetermined motion direction, depending on the predetermined motion strength on that trial. Each stimulus was composed of 18 video frames with a 60 Hz refresh rates (i.e., 300-ms presentation).

Within the center circle mask of the donut-shape motion stimulus, a face/place superimposed stimulus was presented simultaneously (Fig. 1C). The face image set consisted of an image of picture of 4 male and 4 female unfamiliar Japanese faces, and the place image set consisted of 4 indoor and 4 outdoor unfamiliar places; this resulted in 64 overlaid images.

Task Procedure

At the beginning of the task, a dot patch and face/place stimulus were simultaneously presented. The direction of the dot patch (up or down) indicated the task to be performed (discrimination of face or place). Depending on the motion direction, participants were required to judge whether the face was male or female, or whether the place was indoor or outdoor (Fig. 1C), and pressed the corresponding button with their right thumb. The simultaneous presentation of cue (motion dots) and target (face/place picture) stimuli was aimed to maximize the behavioral effect of switching and motion coherence by minimizing preparatory processes triggered by the presentation of cue stimulus (Sakai et al. 2008).

Participants made button responses using the left or right buttons in both tasks. We used this procedure to prevent an "action switch," derived from alternating different button sets depending on the tasks, which is known to involve separate prefrontal mechanisms (Kim et al. 2012). Stimulus–response (SR) mapping of the face and place task was counterbalanced across participants. Specifically, each participant was assigned to 1 of the 4 SR maps: 1) female: left, male: right, indoor: left, outdoor: right; 2) female: right, male: left, indoor: left, outdoor: right; 3) female: left, male: right, indoor: right, outdoor: left; and 4) female: right, male: left, indoor: right, outdoor: left.

Both of accuracy and speed were stressed. Stimulus was presented for 1.8 s, followed by a 0.7-s intertrial interval. If participants made an incorrect response or did not respond within 1.8 s from the stimulus onset, feedback stimulus indicating an error (X) was presented for 1.0 s, followed by high-coherence (80%) cue trials for the same task dimension. The cue trials immediately after the error were discarded from analyses. SR and cue-task associations for the 2 tasks were identical on days 1 and 2 and counterbalanced across participants.

The trial with simultaneous cue/target presentation was followed by trials with presentation of the face/place target stimulus without the dots cue stimulus (target only trials). The target-only trials were repeated for 3–5 times (Fig. 1D). In the target-only trials, participants were required to discriminate the center image stimulus along the same dimension until the next task cue (moving dots) was presented. One task block lasted for approximately for 90 s, and 20-sec fixation blocks were inserted between task blocks. Each functional run involved 3 task blocks and lasted for 305 s. The first trial at the beginning of each task block presented the dot cue with highest coherence (80%) and was discarded from analysis.

Practice Procedure

On the first day, participants practiced the tasks outside of the scanner. They first practiced a discrimination task for the moving dot stimulus and were required to judge the direction of overall motion (up or down) and to press the correct corresponding button as quickly as possible (Tsumura et al. 2021). The response window was 1050 ms. Each practice run involved 70 trials, and 5 runs were administered for each participant. The first 5 trials and last 5 trials in each run used the highest coherence level (80%). Thus, the middle 60 trials were composed of 20 trials for each of coherence levels (20, 40, or 80%).

The participants then practiced discrimination tasks for the face and place stimulus (see above). Across task switching practice runs, the switching frequency and coherence levels of the moving dot cue were manipulated such that the cue trials gradually became more difficult (i.e., more switch trials with low-coherence cue). Participants were instructed to first judge the motion direction and then to discriminate the picture based on the motion direction in the cue trials. They received 8 practice runs approximately for 40 min involving 120 cue trials.

Behavioral Procedure in Scanning Session

On the second day, after practicing task switching for 1 run, the participants performed 9 runs of task switching with identical procedure as for day 1 (see above) while functional MRI was administered. The frequency of switch and repeat trials and coherence level were approximately equivalent across runs.

Imaging Procedure

MRI scanning was administered by a 3T MRI scanner (Siemens Verio, Germany) with a 32-channel head coil. Functional images were acquired using a multiband acceleration echo-planar imaging sequence (Moeller et al. 2010) (repetition time [TR]: 0.8 s; echo time [TE]: 30 ms; flip angle [FA]: 45 degree; 80 slices; slice thickness: 2 mm; in-plane resolution: 3 × 3 mm; multiband factor: 8). Each functional run involved 385 volume acquisitions. The first 10 volumes were discarded from analysis to take into account the equilibrium of the longitudinal magnetization. High-resolution anatomical images were acquired using an magnetization-prepared rapid gradient echo T1-weighted sequence (TR: 2500 ms; TE = 4.32 ms; FA: 8 deg; 192 slices; slice thickness: 1 mm; in-plane resolution: 0.9 × 0.9 mm^2).

Behavioral Analysis

Trials were classified into 3 types in terms of switching: 1) task switch trials presenting the random dot cue and target stimuli simultaneously, where the task to be performed alternated (i.e., face to place or place to face); 2) task repeat trials presenting the random dot cue and target stimuli simultaneously, where the same task was repeated; and 3) target-only trials presented after the switch and repeat trials, and their subsequent trials without random dot cue stimuli presentation (Fig. 1D). These trial types were analyzed separately. Trials were also classified by task dimension (face or place), and switch and repeat trials

were examined at each coherence level (20, 40, or 80%). Accuracy and reaction times (RTs) were calculated for each trial condition and then compared. Statistical testing was performed based on repeated measures ANOVAs implemented in SPSS Statistics 24 (IBM Corporation, New York, NY).

Image Preprocessing

MRI data were analyzed using SPM12 software (http://fil.ion.ac.uk/spm/). All functional images were initially temporally realigned across volumes and runs, and the anatomical image was coregistered to a mean image of the functional images. The functional images were subsequently spatially normalized to a standard Montreal Neurological Institute (MNI) template with normalization parameters estimated based on the anatomical scans. The images were resampled into 2-mm isotropic voxels and spatially smoothed with a 6-mm full-width at half-maximum Gaussian kernel.

Imaging Analysis

Single-Level Analysis

A general linear model (GLM) approach (Worsley and Friston 1995) was used to estimate parameter values for task events. The events of interest were correct switch, repeat, and target-only trials. For switch and repeat trials, the normalized (z-scored) coherence level of the dot stimuli was also added as a parametrical effect of interest (Tsumura et al. 2021). Error trials in all conditions were separately coded in GLM as nuisance effects. Those task events were time-locked to the onset of target images and then convolved with canonical hemodynamic response function implemented in SPM. Additionally, 6-axis head movement parameters, white-matter signals, lateral-ventricle signals, and parametrical effect of RTs normalized across trials were also included in GLM as nuisance effects. The parameters were then estimated for each voxel across the whole brain.

Group-Level Analysis

Maps of parameter estimates were first contrasted within individual participants. Contrast maps were collected from all participants and subjected to a group-level paired *t*-test. For the coherence effect, the contrast maps were subjected to a one-sample group-mean test, with maps weighted and summed based on normalized coherence levels. Voxel clusters were identified using an uncorrected threshold of $P < 0.001$ based on voxel-wise *t*-statistics. The voxel clusters were tested for a significance with a threshold of $P < 0.05$ corrected by family-wise error (FWE) rate based on permutation methods (Nichols and Holmes 2001) (5000 permutations) implemented in *randomise* in FSL suite (http://fmrib.ox.ac.uk/fsl/). This group analysis procedure was validated to appropriately control false-positive rates in a prior study (Eklund et al. 2016). Peaks of significant clusters were then identified and listed on tables. If multiple peaks were identified within 12 mm in one cluster, the most significant peak was retained. When exploring brain regions associated with motion coherence, exploration was restricted within a mask obtained from Neurosynth (Yarkoni et al. 2011) (http://neurosynth.org/) for the search word "motion" ($z > 3.0$, for uniformity test), in order to ensure the extraction of motion-related regions, because the current cue trials simultaneously presented face/place stimuli in addition to cues indicating task switching tasks between face and place discrimination.

Effective Connectivity Analysis

The current analysis was designed to test the hypothesis that functional connectivity among brain regions associated with task switching, motion perception, face perception, and place perception identified in univariate analysis (Fig. 2A–C) is modulated by task manipulations and brain signals. Dynamic causal modeling (DCM; Friston et al. 2003) analysis implemented in SPM12 was performed in order to examine functional connectivity mechanisms associated with task switching under cue uncertainty (Fig. 2D,E and Supplementary Figs S1 and S2). DCM allows us to explore the effective connectivity among brain regions under the premise that the brain is a deterministic dynamic system that is subject to environmental inputs and produces outputs based on the space-state model. The model constructs a nonlinear system involving intrinsic connectivity, task-induced connectivity, and extrinsic inputs. Specifically, a model for neural activity was formulated as a linear time-invariant space-state dynamic system,

$$\begin{aligned}\frac{dx(t)}{dt} &= Ax(t) + u(t)Bx(t) + u(t)C \\ &= [A + u(t)B]\,x(t) + u(t)C\end{aligned} \quad (1)$$

where $x(t)$ denotes the states of neural activity in k brain regions ($k \times 1$ vector; k = 2, 3, or 4) at time t, $u(t)$ denotes inputs to the system from task events at time t (scalar value), A denotes intrinsic connectivity ($k \times k$ matrix), B denotes effective connectivity ($k \times k$ matrix), and C denotes the direct influence of the task variable on neural activity (direct extrinsic input; $k \times 1$ vector). Because the time derivative of neural activity (left side in eq. 1) is modulated by $[A + u(t)B]x(t)$, the directionality of connectivity is reflected in the A and B matrices. More specifically, the rows and columns of the A and B matrices indicate the target and source of the directionality. The A, B, and C matrices involved k^2, $k(k-1)$, and k parameters, respectively (only nondiagonal elements are parameters for matrix B). Thus, the model involved $2k^2$ parameters in total. The unit of the connectivity is arbitrary.

Next, the neural activity $x(t)$ in the model (eq. 1) was transformed as

$$y(t) = \lambda\,(x(t)) \quad (2)$$

where λ denotes nonlinear function providing fMRI signals from neural activity.

Then, parameters of the nonlinear system (A, B, and C) are estimated based on fMRI time series and task variable/events. The use of a high temporal resolution sequence for functional imaging enables us to collect more scan frames to increase the signal-to-noise ratio of the DCM analysis (Penny et al. 2004; Stephan et al. 2010; Tsumura et al. 2021).

Four regions of interest (ROIs) were first defined based on univariate analysis and prior studies of task switching and perceptual decision-making: 1) task switching (left lateral prefrontal cortex [lPFC]; Dove et al. 2000; Konishi et al. 2002; Rushworth et al. 2002; Bunge et al. 2005; Derrfuss et al. 2005; Crone et al. 2006; Yeung et al. 2006; Jimura and Braver 2010; Kim et al. 2012; Tsumura et al. 2021; Fig. 2A, see also Results); 2) motion perception (MT; Gold and Shadlen 2007; Newsome and pare 1988; Shadlen et al. 1996; Beauchamp et al. 1997; Huk et al. 2002; Kayser et al. 2010; Hanks and Summerfield 2017; Tsumura et al. 2021; Fig. 2B, see also Results); 3) face perception (FFA; Kanwisher et al. 1997; McCarthy et al. 1997; Ishai et al. 1999; Gazzaley et al. 2005; Freiwald and Tsao 2010; Fig. 2C, see also Results); 4) place perception (PPA; Epstein and Kanwisher 1998; Ishai et al. 1999;

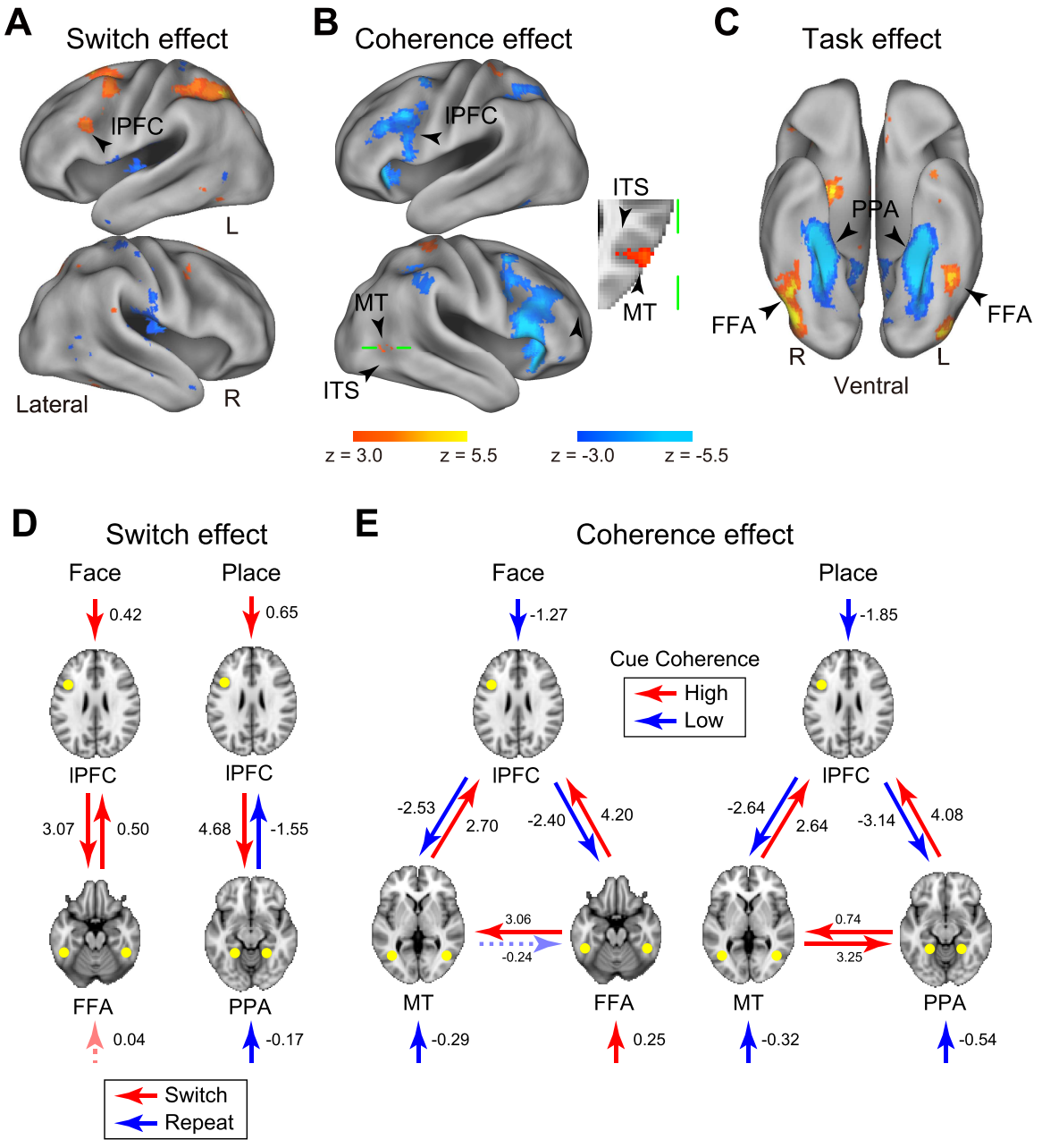

Figure 2. Whole-brain exploration activation and functional connectivity analyses (*A–C*). Statistical activation map of univariate analysis and functional connectivity analysis. Maps are overlaid onto the 3D surface of the brain. Hot and cool colors indicate positive and negative effects, respectively. (*A*) Switch effect (switch minus repeat trials). (*B*) Motion coherence effect (high- vs. low-coherence trials). Green solid line on the surface indicates the axial section on the right. (*C*) Task effect (face minus place tasks in target-only trials). ITS, inferior temporal sulcus. (*D*, *E*) Effective connectivity analysis. The interregional connectivity shows parameter estimates of the matrix B in eq. 1 (effective connectivity), and the direct extrinsic input to each region represents parameter estimates of the matrix C in eq. 1. (*D*) Effective connectivity and extrinsic inputs modulated by the contrast switch versus repeat trials during face (left) and place (right) tasks (coherence levels collapsed). Red arrows indicate greater connectivity or inputs in switch relative to repeat trials, and blue arrows indicate greater connectivity or inputs in repeat relative to switch trials. The values next to the arrows indicate the magnitude of connectivity enhancements (positive: switch > repeat; negative: repeat > positive). (*E*) Effective connectivity and extrinsic inputs modulated by motion coherence during face (left) and place (right) tasks (switch and repeat trials collapsed). Red arrows indicate greater connectivity or inputs in high-coherence relative to low-coherence trial and blue arrows indicate greater connectivity or inputs in low-coherence relative to high coherence trials. The values next to the arrows indicate the magnitude of connectivity (positive: high coherence > low coherence; negative: low coherence > high coherence). Arrows with solid line indicate statistically significant connectivity ($P < 0.05$, uncorrected). The arrows, circles, and labels in the panels indicate same locations.

Gazzaley et al. 2005; Fig. 2C, see also Results). More specifically, meta-analysis maps were obtained from Neurosynth (http://neurosynth.org/; Yarkoni et al. 2011) using a keyword search for "switching," "mt," "ffa," "place" to obtain the meta-analysis maps for lPFC, MT, FFA, and PPA ROIs, respectively. ROI images were then created with 6-mm radius spheres centered in the peak

coordinates in the meta-analysis activation maps thresholded above $z > 10$ (uniformity test).

Given these ROIs, we first tested whether the switch-related prefrontal region sends (or receives) a task-related signal toward (or from) the stimulus-modality-dependent occipitotemporal region of the target (i.e., FFA/PPA) during task switching (Fig. 2*D*). If this was the case, then we tested whether signaling between prefrontal and occipitotemporal regions changed depending on the uncertainty of cue stimuli (Fig. 2*E*).

Signal time courses (3375 scanning frames) of 4 ROIs and regressors in events of interest were extracted from first-level GLMs. The events of interest were cue (switch and repeat) trials and target-only trials of each task. For switch and repeat trials, the contrast of the 2 trials (switch: 1; repeat: −1) and normalized coherence level of the dot stimuli were added as parametrical effects of interest. Nuisance effects of head motion, white matter signal, ventricle signal, functional run, and contrast were subtracted out from the ROI timecourses. The input matrix was U mean-centered.

For each trial effect, causal models were defined as those that differed in external inputs and modulatory effects among ROIs. We were interested in strengths of effective connectivity between pairs of ROIs (i.e., lPFC, MT, and FFA/PPA), rather than exploration of a model that best fits to the data. Thus, we considered all theoretically possible models. As the current models involved 2 or 3 ROIs (Fig. 2*D*,*E*), the tested models included 16 or 512 types (i.e., 2^2 inputs and 2^2 connection effects or 2^3 inputs and 2^6 connection effects). Connectivity matrices reflecting 1) first-order connectivity, 2) effective change in coupling induced by the inputs, and 3) extrinsic inputs on MRI signal in ROIs were estimated for each of the 16 or 512 models based on DCM analysis implemented in SPM12. Parametric regressor (switch vs. repeat/coherence) was used as an extrinsic effect for effective connectivity between ROIs and ROI inputs.

In order to estimate the strength of effective connectivity, a Bayesian model reduction method (Friston et al. 2016) was used. The reduction method reduces the number of models based on model evidence (free energy; Penny. 2012) and calculates posterior densities for all reduced models, which were then inverted to a fully connected model. In the current analysis for the switch effect, out of the 16 possible models, 10 and 8 models were selected for parameter estimations of the face task and place task, respectively (minimum free energy: −639.1 [face], −677.8 [place]). For the coherence effect, out of the 512 possible models, 203 and 61 models were selected for parameter estimations of the face task and place task, respectively (minimum free energy: −737.6 [face], −742.8 [place]).

The reduced models were then supplemented by second-level parametric empirical Bayes (Friston et al. 2016) to apply empirical priors that remove subjects' variability from each model.

Next, the parameters of these models were estimated based on Bayesian model averaging (Friston et al. 2003) to estimate group-level statistics. Because the current analysis aimed to identify effective connectivity observed as an average across participants, we used a fixed effect (FFX) estimation assuming that every participant uses the same model. This is in contrast to using a random effect (RFX) estimation assuming different participants use different models, which is often used to test group differences in effective connectivity (Penny et al. 2010). The significance of connectivity was then tested by thresholding at a posterior probability at the 95% confidence interval. We used the uncorrected threshold, because the current analysis aimed to test if connectivity between 2 specific brain regions was enhanced depending on task manipulation and brain activity, not to explore one model involving connectivity among multiple brain regions that best fits to the imaging and behavioral data (Tsumura et al. 2021).

Additionally, in order to test the robustness of the functional connectivity (c.f. Smith et al. 2011), we performed supplemental analyses. We estimated model parameters 1) without an empirical prior (Supplementary Fig. S2C,D); 2) changing the number of the ROIs in the models (Supplementary Fig. S2E,F), and 3) changing the definition of the ROIs (Supplementary Fig. S2G,H). When changing the ROI definition, we used a leave-one-out procedure; the centers of ROIs of 1 participant were determined based on group-level univariate activation maps of corresponding contrasts (i.e., lPFC: switch vs. repeat trials; MT: high vs. low coherence trials; FFA/PPA: face vs. place of target-only trial) without the participant in order to circumvent circular analysis. The ROIs were created as spheres with 6 mm radius for individual participants. Complete results can be provided upon request.

Even with these analysis procedures and supplementary analyses above, we acknowledge that better fitting models would be possible if appropriate sets of ROIs were defined.

CNN Classifier

In order to explore brain regions involving task-related neural representation, a CNN classifier (Krizhevsky et al. 2012; LeCun et al. 2015) was used. The layer weights of CNN represent convolutional filters that extract physical features of images, such as edges, shapes, and spatial frequency. Serial multiplication of the convolutional layers enables multilevel abstraction for image classification (Krizhevsky et al. 2012; LeCun et al. 2015).

The current CNN model was based on VGG16 (Simonyan and Zisserman 2015), with 5 convolution layers for extracting image features and 2 fully connected layers for binary classification. Initial parameters of convolution layers were set to parameters pretrained with concrete object images provided from ImageNet (http://www.image-net.org/; Deng et al. 2009; Supplementary Fig. S3A).

The VGG16/ImageNet model is capable of classifying concrete object images into 1000 item categories. Importantly, it has been demonstrated that the pretrained model can learn novel image sets more efficiently than the nontrained model by tuning convolution and fully connected layers and fully connected layers only (Donahue et al. 2014; Pan and Yang 2010; Fig. S3C). Thus, the current analysis retrained the pretrained VGG16-ImageNet model to classify brain activation maps.

Training data were single-subject second-level z-maps during the N-back working memory (WM) task from the S1200 release of the Human Connectome Project ($N = 992$; HCP; http://www.humanconnectomeproject.org/; Barch et al. 2013; Glasser, Smith, et al. 2016a). From each participant, statistical z-maps for activation contrasts for face versus fixation and place versus fixation (2-back and 0-back corrupted) were collected. We used grayscaled flat 2D cortical maps (Glasser, Coalson, et al. 2016b) provided from HCP (992 images; face: 496, place: 496; Supplementary Fig. S3B) for dimensional compatibility of images between VGG16-ImageNet and activation maps. The training data set was divided into 10 subsets, and 9 subsets were used for retraining and the remaining 1 set was used for validation, enabling a 10-fold cross-validation test. Then, the pretrained VGG16 model was retrained by the activation

maps such that the model classifies face and place trials. Model training and testing were implemented using Keras (https://keras.io/) under Tensorflow backend (https://www.tensorflow.org/) (input image size: 480 × 1280 pixels; batch size: 10; epoch: 50; learning rate: 0.0001; optimizer: Stochastic Gradient Descent; Supplementary Fig. S3C top).

After retraining of the HCP WM maps, the model with highest classification accuracy was further retrained to classify activity maps for face and place tasks during target-only trials of the current dataset (Supplementary Fig. S3C bottom). For each functional run of each participant, a single-level GLM estimation was performed with regressors identical to those in the univariate analysis as described above. The GLM estimations were performed within standard MNI space. Activation z-maps for the contrasts for face versus fixation and place versus fixation during correct target-only trials were collected from each functional run. Activity maps for the contrast for face versus fixation and place versus fixation were then grayscaled and flattened such that these maps were anatomically and geometrically identical to those from the HCP WM task using Connectome Workbench (https://www.humanconnectome.org/software/connectome-workbench/). The training dataset consisting of 522 images (261 face and 261 place maps from each of 9 runs of 29 participants) was divided into 10 subsets, enabling a 10-fold cross-validation test.

Given the limited number of images available from the current experiment, this 2-step retraining of the model was found to be effective when classifying current tasks, because 1) training randomly initialized models failed in classifying the current target-only trials (Supplementary Fig. S4A) and in classifying the HCP WM conditions (Supplementary Fig. S4B); 2) retraining VGG16-ImageNet was successful in classifying the HCP WM conditions (Supplementary Fig. S4B); 3) retraining VGG16-ImageNet model failed in classifying the current target-only trials (Supplementary Fig. S4A); and 4) retraining VGG16-ImageNet/HCP was successful when classifying the current face/place tasks (Supplementary Fig. S4A).

After learning of the target-only trials from the current dataset, the retrained 10 models were tested to classify activation maps during task switching and repeat trials where the dot cue stimulus was presented to indicate the task to be performed. Testing data were created based on a GLM analysis where switch and repeat trials at differential coherence levels were coded separately. For each functional run of each participant, a single-level GLM estimation was performed, and activation contrast z-maps for face versus fixation and place versus fixation were collected during those correct switch and repeat trials. Grayscale 2D activation maps were created for the contrasts for face versus fixation and place versus fixation during 6 types of cue presentation trials (switch/repeat × high/middle/low coherence). Importantly, the testing data were independent of the 2 sets of retraining data (HCP WM and current target-only trials). The maps were tested, and accuracy was averaged across cross-validation models within each participant. A statistical test of classification accuracy was performed based on repeated measures ANOVAs implemented in SPSS Statistics 24 (IBM Corporation, New York, NY).

In a separate supplemental analysis, in order to examine whether the current results were biased by subject-specific characteristics of image data, we used a leave-one-subject-out procedure to retrain the CNN classifier to classify activation maps of the target only trials and then tested the remaining subject (Supplementary Fig. S6, see also Results).

A recent study demonstrated that a CNN model successfully learned and classified task-related fMRI images without flattening the images (Wang et al. 2020). Although retraining of the classifier was also effective for small data sets, this technique is available only for blocked-design experiments as the model was trained and tested based on fMRI time series, which increased the number of the training data. The model also normalizes and convolves the time series along the temporal axis. Because the current study used event-related design, only activation maps estimated by single-level GLM analyses were available for training and testing. Because of the nature of GLM analysis, the number of available images was limited in comparison with fMRI time series. Thus, retraining of pretrained model based on flattened 2D activation maps is effective for small size dataset of event-related fMRI.

Visualization of Activations of Convolution Layers

When an image is given to CNN, the CNN extracts image tensor information through convolutional layers, which is reflected in the activation of the layers (Krizhevsky et al. 2012; LeCun et al. 2015). Thus, the magnitudes of the layer activations involve critical information to classify performed tasks (i.e., face or place task).

To identify brain regions involving such critical information, the current study used gradient-weighted class activation mapping (Grad-CAM; Selvaraju et al. 2017). Grad-CAM aggregates activations of convolution layers and creates maps of the aggregated activations, enabling us to highlight image locations critical for classification (Fig. 3C). Specifically, a greater layer activation in an image location indicates that the location involves more critical information to classify the image. The map can be created for each image because the layer activations are calculated for each image.

Aggregated layer activations were visualized onto 2D brain surface maps for each tested map. For each participant, the aggregated layer activation maps were averaged across maps within each of the 6 cue trial conditions (switch/repeat × high/middle/low coherence cue). The averaged maps were contrasted between switch and repeat trials to explore brain regions showing differential activations between switch and repeat trials within participants. For coherence the effect, 3 maps for coherence level trials were weighted and summed based on behavioral accuracy estimated by sigmoid fitting within participants. For each contrast, maps were collected from all participants, and pixel-wise z-values were calculated treating participants as a random effect, and the z-values were mapped on the 2D surfaces of the cortical areas. Because we made activation contrasts, sign and magnitude of the activation contrasts (Fig. 3*D*,*E* and Supplementary Fig. S7B,C) indicate relative difference in the aggregated layer activations.

In order to statistically test dissociable activation patterns during cue trials in the FFA and PPA, ROI analyses were performed (Fig. 3F). ROIs were defined based on Neurosynth, which were identical to those used in the DCM analyses. From each ROI, aggregated layer activation magnitudes were collected for each of the trial conditions: task (face/place), switching (switch/repeat), and cue coherence (high/middle/low: 80/40/20%) and then averaged within ROIs for each participant. Statistical tests were then performed based on repeated measures ANOVA.

In order to test the robustness of the ROI analysis against the definition of ROIs, we redefined ROIs as regions showing greater activation in the contrast of face minus place tasks

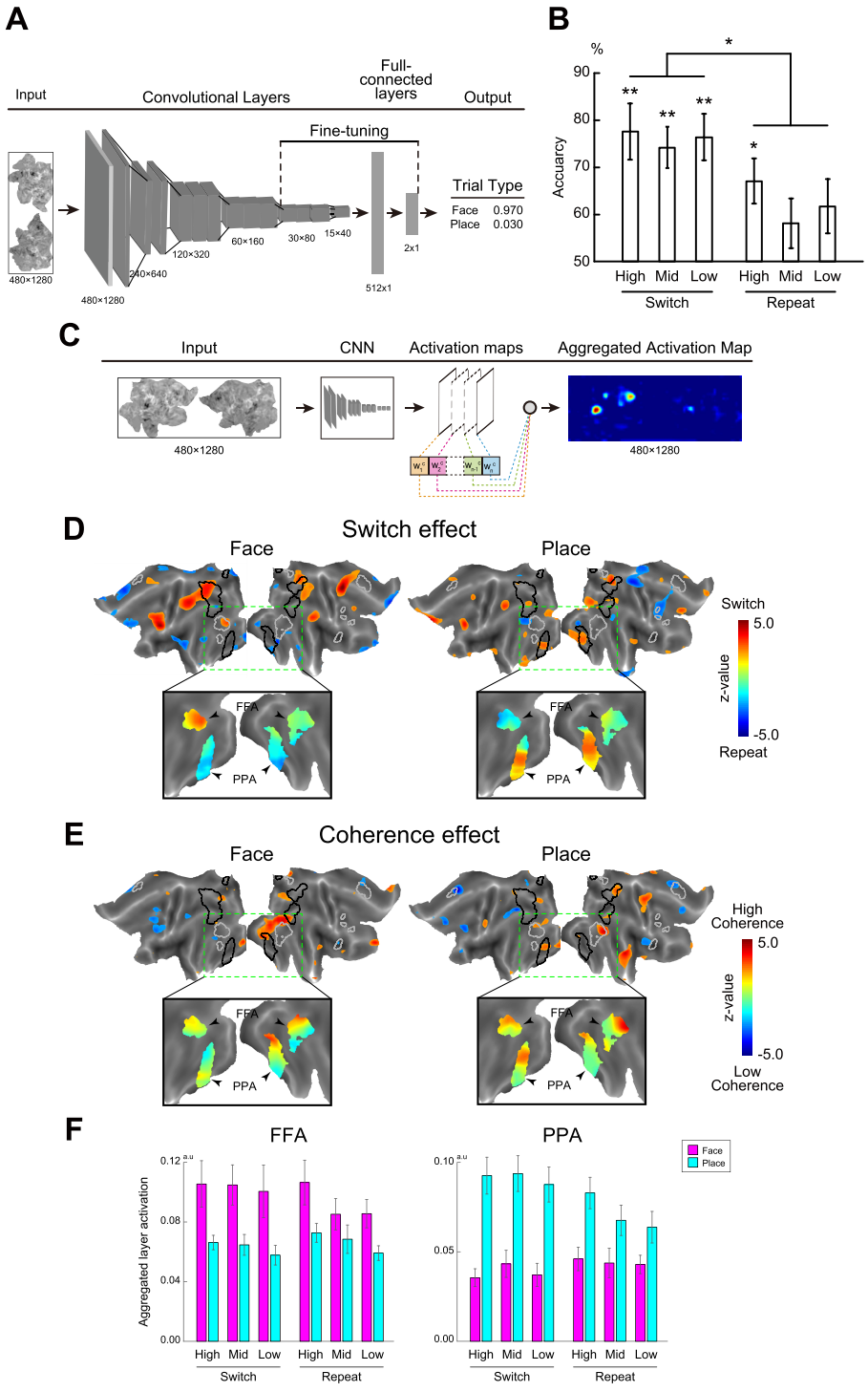

Figure 3. CNN classifier mapping. (*A*) CNN model was based on VGG16 pretrained by ImageNet data. The model was retrained to classify performed task (face or place). Retraining of brain maps was based on fine-tuning from the fifth convolution layer to the full-connected layers. (*B*) Classification accuracy for each task condition. $^{*}P < 0.005$; $^{**}P < 0.001$. High: high-coherence (80%) trials; Mid: middle-coherence (40%) trials; Low: low-coherence (20%) trials. (*C*) Activations of convolution layers were aggregated across layers and then visualized to identify cortical regions involving more information to classify tasks. (*D*) Visualization of aggregated layer activation contrasts for switch versus repeat trials in face task (left) and place task (right). Statistical maps are overlaid onto flat cortical anatomical images with a statistical threshold of $|z| > 2.0$ (top). Positive z-values indicate greater activations in switch relative to repeat trials, and negative z-values indicate greater activations in repeat relative to switch trials. Gray and black closed lines overlaid on flat map indicate brain regions significantly activated during face and place tasks in univariate analysis, respectively (Fig. 2C and Supplementary Fig. S5). Occipitotemporal regions in rectangular boxes with green broken lines were expanded below. Maps were overlaid onto flat maps masked by the univariate activation contrast face versus place tasks for target-only trials (gray and black closed lines in the top panels). The FFA and PPA are indicated by arrow heads and the same locations in Fig. 2A–E. (*E*) Visualization of aggregated layer activation contrast for coherence effect in face task (left) and place task (right). Positive z-values indicate greater activations in high-coherence relative to low-coherence trials, and negative z-values indicate greater activations in low-coherence relative to high-coherence trials. The formats are similar to those in panel (*D*). (*F*) ROI analysis. Aggregated layer activation magnitudes were collected for each cue trial conditions and ROIs. All error bars in the figure indicate standard errors of the means across participants.

or place minus face tasks (Fig. 2C; Supplementary Fig. S5 and Supplementary Table S3), independently of the cue trials.

SVM Analysis for Whole-Brain Cortical Regions

In order to supplement the CNN classifier analysis, multivariate pattern analyses (MVPA) based on SVM were performed. The SVM classifier was trained to perform bivariate classification for face and place tasks. Training and testing were implemented using scikit-learn package (https://scikit-learn.org/stable/) with a Tensorflow backend (https://www.tensorflow.org/). We used a linear kernel and adjusted *C* parameters ($C = 0.1, 1.0, 10.0$). As the overall results were maintained with the *C* adjustment, then we reported the results with the default parameter ($C = 1.0$).

For classifier training, we used the image set of single-subject second-level z-maps during the N-back WM task (face vs. fixation and place vs. fixation) obtained from the S1200 HCP ($N = 992$), which was identical to those used in the CNN classifier training. The classifier was trained by the activation images, such that it classified face and place tasks during the HCP N-back WM task. Weights of the trained classifier were mapped on the 2D cortical surface.

The testing dataset was also identical to those used in the CNN classifier analysis: 2D z-maps for activation contrast of switch and repeat trials (switch vs. fixation and repeat vs. fixation) at each coherence level (20, 40, 80%) of the current experiment (Supplementary Fig. S3B). Testing activation images were subject to the trained classifier for each cue trial condition of each participant, and classification accuracy was averaged for each trial condition across participants (Fig. 4A). In a separate analysis, the SVM classifier was trained based on single-level z-maps during target-only trials from the current dataset ($N = 29$), and the identical image set was tested (Supplementary Fig. S8).

Searchlight SVM

In order to explore brain regions in which local activity patterns involve information about a performed task (face/place), searchlight MVPA (Kriegeskorte et al. 2006) was conducted. Bivariate classification based on SVM was used to decode the performed task (face or place). A searchlight procedure with a 5-voxel radius was used to provide a measure of decoding accuracy in the neighborhood of each voxel. Training and testing were performed based on the Decoding Toolbox (TDT; version 3.95; https://sites.google.com/site/tdtdecodingtoolbox/). Again, training and testing data were independent, based on different behavioral tasks and data sets (training: HCP WM; testing: current task switching with male/female or indoor/outdoor judgments).

Training data were 3D single-subject second-level z-maps during N-back WM task ($N = 1000$) from HCP S1200 release (Barch et al. 2013; Glasser, Smith, et al. 2016a). Similar to the CNN analysis, we used activation contrasts for face task versus fixation and for place task versus fixation (2-back and 0-back corrupted). These activation contrasts were collected from each participant, and the whole dataset was divided into 10 subsets ($N = 100$ each). For each subset of the training data, a classifier in each searchlight was trained based on these z-maps such that it classified face and place conditions during the HCP N-back WM task.

Test data were single-subject z-maps during switch, repeat, and target-only trials of the current experiment. For each functional run of each participant, single-level GLM estimation was performed with regressors identical to those in the univariate analysis as described above. Activation maps for face versus fixation and place versus fixation during switch and repeat correct trials were collected from each functional run. These GLM analyses were performed within standard MNI space.

Another set of testing data was created based on a separate GLM analysis with correct cue (switch and repeat) trials separately coded at each coherence level (20/40/80%). For each functional run of each participant, a single-level GLM estimation was performed, and activation maps for face versus fixation and place versus fixation were collected for each coherence level.

For each training subset, the classifier was tested on whether it correctly classified the performed task (face or place task) for each trial condition (i.e., switch/repeat/target only and coherence level). Classification performance was then collected from all functional runs and averaged within participants for each searchlight. The performance of classification was calculated as the accuracy minus chance level for bivariate classification. Accuracy maps were then averaged across testing data sets within participants.

Accuracy maps for switch, repeat, and target-only trials were first averaged across training subset models within participants, and averaged accuracy maps were collected from all participants. Voxel-wise one-sample group-mean test was performed for each trial condition, with a procedure similar to that in the univariate analysis as stated above. In order to explore brain regions showing differential classification accuracy between switch and repeat trials, voxel-wise group-level paired test was performed, and significance was tested similarly.

Accuracy maps for each coherence level trials were also averaged across training subset models within participants, and the averaged accuracy maps were collected from all participants. Voxel-wise one-sample group-mean test was performed for each coherence level, and significance was tested similarly. In order to examine coherence effect, the maps for 3 coherence levels were weighted and summed based on behavioral accuracy estimated by sigmoid fitting. Voxel-wise one-sample group-mean test was performed, and significance was tested similarly.

In separate analyses, all group-level tests above were also performed for each of the 10 training subset models, and we confirmed that overall results were consistent.

ROI analysis was performed to compare classification accuracy across experimental conditions in FFA and PPA. ROIs were identical to those used in DCM analyses defined based on Neurosynth.

Results

Behavioral Results

Human participants performed a task-switching paradigm (Koechlin et al. 2003; Bunge et al. 2005; Badre 2008; Jimura and Braver 2010; Tsumura et al. 2021), in which they alternated discrimination tasks for face and place stimuli (Fig. 1C,D). The relevant task was indicated by a cue stimulus involving perceptual uncertainty, which was manipulated by the motion strength of randomly moving dots.

In the face task, participants made correct responses for 30.0 ± 4.8 (mean $\pm$ SD) switch trials, for 29.9 ± 4.4 repeat trials, for 21.3 ± 3.1 high-coherence trials, for 19.8 ± 3.2 middle-coherence trials, and for 18.7 ± 3.3 low-coherence trials. In the place task, they made correct responses for 29.5 ± 4.4 switch trials,

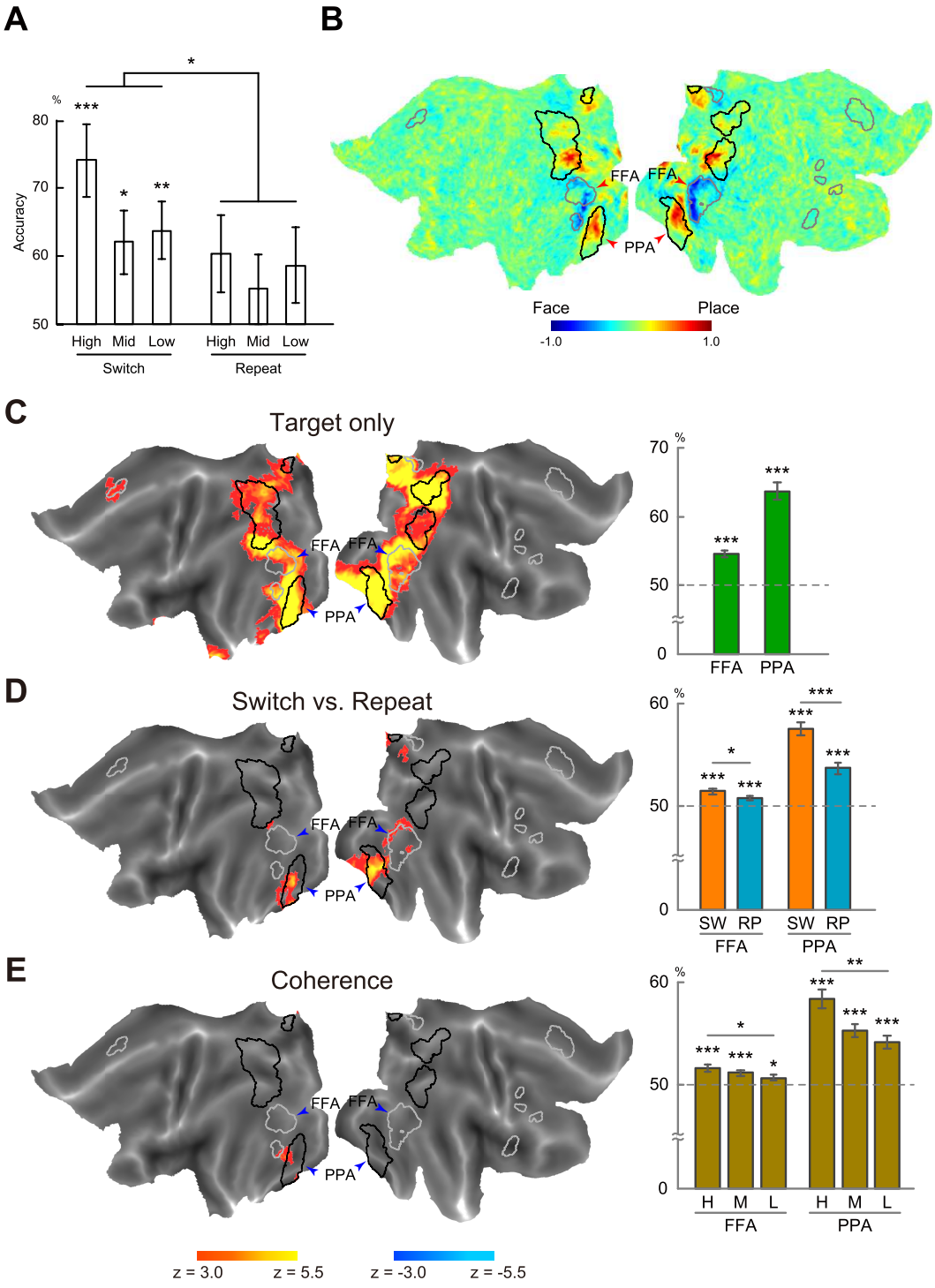

Figure 4. Multivariate pattern analysis (MVPA) with SVM. (*A*) Classification accuracy for each task condition with SVM classification in the whole brain. The model was trained for classification of face and place tasks with the working memory task in the HCP and tested with each condition of cue trials in the current study. Error bars indicate standard error of the mean across participants. $^{*}P < 0.05$; $^{**}P < 0.005$; $^{***}P < 0.001$. (*B*) Visualization of weight assigned to the pixels for classification of face and place tasks in the working memory task in the HCP. Formats are similar to those in Figure 3*D*,*E*. (*C*–*E*) Statistical significance maps for searchlight MVPA (left). Classifiers were trained to classify the performed task. Maps are overlaid onto a 2D flat map of the brain and displayed from a ventral view. White and black closed lines overlaid onto the 2D flat map of the brain indicate significant clusters for contrast face versus place tasks in univariate analysis, respectively (Fig. 2*C* and Supplementary Fig. S5). The FFA and PPA are indicated by blue arrow heads and the same locations in Figs 2*A*–*E* and 3*D*,*E*. ROI analysis (right). Voxel-wise classification accuracy was averaged within FFA and PPA and compared across experimental condition. Error bars indicate SEM across participants. (*C*) Target-only trial. Hot and cool colors indicate statistical level for classification accuracy relative to chance level. (*D*) Accuracy difference between switch and repeat trials. Hot and cool colors indicate higher accuracy in switch and repeat trials, respectively. (*E*) Differential classification accuracy depending on the coherence effect. Hot and cool colors indicate higher accuracy in high- and low-coherence trials, respectively.

for 29.2 ± 4.9 repeat trials, 20.7 ± 4.0 high-coherence trials, for 19.3 ± 3.6 middle-coherence trials, and for 18.7 ± 2.6 low-coherence trials. The number of the trials is comparable to those in our previous study (Tsumura et al. 2021).

Accuracy was lower in low-coherence (i.e., more uncertain) trials compared to high-coherence trials [$F(1, 28) = 12.7$; $P < 0.005$; Fig. 1E] and became lower in switch trials than in repeat trials [$F(1, 28) = 11.8$; $P < 0.005$]. Likewise, RTs were longer in low-coherence trials compared to high-coherence trials [$F(1, 28) = 69.9$; $P < 0.001$; Fig. 1F] and were longer in switch trials than in repeat trials [$F(1, 28) = 43.2$; $P < 0.001$]. These behavioral results suggest that the current behavioral task successfully manipulated task switching (Allport et al. 1994; Rogers and Monsell 1995; Dove et al. 2000; Rushworth et al. 2002; Crone et al. 2006; Yeung et al. 2006; Jimura and Braver 2010; Tsumura et al. 2021) and perceptual decision-making (Newsome and Pare 1988; Shadlen et al. 1996; Palmer et al. 2005; Kayser et al. 2010; Hanks and Summerfield 2017; Tsumura et al. 2021). The interaction effects of trial type (switch/repeat) and coherence levels did not reach statistical significance [Accuracy: $F(1, 28) = 1.8$; $P = 0.19$. RT: $F(1, 28) = 2.1$; $P = 0.16$].

In trials without cue presentation, occurring after the switch and repeat trials until the next cue trials (target only trial; Fig. 1D), accuracy was lower after lower coherence cue trials [$F(1, 28) = 16.3$; $P < 0.001$; Fig. 1E], but RTs remained unchanged after lower coherence cue trials [$F(1, 28) = 1.0$; $P = 0.32$]. Explicit presentation of the cue stimulus involves encoding of the cue stimulus and implementation of the task set even in repeat trials (Sakai 2008; Tsumura et al. 2021). On the other hand, when the cue stimulus was presented with lower coherence dots in switch and repeat trials, the accumulation of visual evidence about cue information became slower (Palmer et al. 2005). The slower accumulation of the cue information may delay the completion of task set encoding and implementation, and task preparation may not be sufficient for the next target-only trial (Sakai 2008). RTs were longer in place than face tasks [$F(1, 28) = 5.0$; $P < 0.05$; Fig. 1F].

Exploration of Switch-Related and Stimulus-Modality-Dependent Brain Regions

We first explored brain regions associated with task switching, motion coherence, and the perception of face and place based on univariate GLM analysis. Figure 2A shows brain regions showing significant increases and decreases in univariate brain activity during switch relative to repeat trials ($P < 0.05$ corrected with cluster-wise FWE rate based on nonparametric permutation tests; see Materials and Methods). Robust activation increases were observed in the left frontal regions, including the inferior frontal cortex (IFC), dorsolateral prefrontal cortex (DLPFC), inferior frontal junction (IFJ), and pre-supplementary motor area (pre-SMA), and in left parietal regions, including posterior parietal cortex (PPC), consistent with prior studies (Dove et al. 2000; Konishi et al. 2002; Rushworth et al. 2002; Bunge et al. 2005; Derrfuss et al. 2005; Crone et al. 2006; Yeung et al. 2006; Jimura and Braver 2010; Kim et al. 2012; Tsumura et al. 2021). A full list of brain regions is shown in Table S1.

We then explored brain regions associated with motion coherence. Figure 2B shows brain regions showing significant modulation of brain activity in relation to motion coherence during the cue (i.e., switch and repeat) trials. In low-coherence trials, activation was increased in multiple frontoparietal regions including IFC, DLPFC, IFJ, pre-SMA, and PPC (Fig. 2B and Supplementary Table S2), consistent with prior studies (Kayser et al. 2010; Tsumura et al. 2021). In contrast, activation was greater in high-coherence trials in the MT region (Fig 2B right; Supplementary Table S2), which is also consistent with prior studies of perceptual decision-making for motion (Newsome and Pare 1988; Shadlen et al. 1996; Beauchamp et al. 1997; Huk et al. 2002; Kayser et al. 2010; Tsumura et al. 2021).

We next explored brain regions associated with face and place tasks (Fig. 2C and Supplementary Table S3). Consistent with prior studies of perception of face (Kanwisher et al. 1997; McCarthy et al. 1997; Ishai et al. 1999; Gazzaley et al. 2005; Freiwald and Tsao 2010) and place (Epstein and Kanwisher 1998; Ishai et al. 1999; Gazzaley et al. 2005), in the face task, activity was greater in the FFA, whereas in the place task, increased activity was observed in the PPA.

We also performed whole-brain exploratory univariate activation analyses for the interactions between 1) switch and coherence effects, 2) switch and task effects, and 3) coherence and task effect and found no significant regions showing the interactions.

These collective univariate activation results suggest that the current univariate activation analysis successfully identified brain regions associated with task switching in addition to the perception of face, place, and motion stimulus and that those regions were cooperatively engaged in task switching when the cue stimulus involved perceptual uncertainty.

Reversal of Functional Connectivity Depending on Cue Uncertainty

The whole-brain exploratory analyses of univariate activation identified 3 types of brain regions: 1) left PFC associated with task switching and perceptual uncertainty (Fig. 2A,B); 2) MT region associated with motion coherence of task cue (Fig. 2B); and 3) FFA and PPA associated with discrimination of face and place stimuli, respectively (Fig. 2C). One possible mechanism to explain these regions playing differential roles in task switching with cue uncertainty is that lPFC, MT, and FFA/PPA mutually received or sent task-related signals during switching, which was modulated by cue coherence, such that the task-related signal complemented the engagement of these regions depending on cue uncertainty and the task to be performed.

In order to test this hypothesis, we performed an interregional effective connectivity analysis based on DCM that allows the examination of directionality of task-related functional connectivity based on the state-space model (see Materials and Methods). The ROIs in lPFC, MT, and FFA/PPA were defined independently of the current results. The lPFC ROI showed a joint effect of switching (greater activity in switch relative to repeat trials) [$t(28) = 3.70$, $P < 0.001$] and negative coherence (greater activity in low-coherence relative to high-coherence trials) [switch: $t(28) = -2.47$, $P < 0.05$; repeat: $t(28) = -5.2$, $P < 0.001$]. The MT ROI showed a significant positive coherence effect (greater activity in high coherence trials) in both hemispheres [left: $t(28) = 2.49$; $P < 0.05$; right: $t(28) = 3.92$; $P < 0.001$]. The FFA ROI showed greater activity during face relative to place task [$t(28) = 6.54$; $P < 0.001$], and the PPA ROIs showed greater activity during place relative to place task [$t(28) = 12.9$; $P < 0.001$].

We first examined the task-related effective connectivity during task switching and found that the connectivity was enhanced from the lPFC toward the FFA during switch relative to repeat trials of the face task (i.e., switch-to-face vs. repeat-face trials; Fig. 2D left and S1A top) and also enhanced from the lPFC

toward the PPA in switch relative to repeat trials of the place task (Fig. 2D right and Supplementary Fig. S1A bottom). These results are in line with the well-known role of the left lPFC in behavioral flexibility (Konishi et al. 2002; Derrfuss et al. 2005; Crone et al. 2006; Yeung et al. 2006; Jimura and Braver 2010; Kim et al. 2012; Tsumura et al. 2021) and suggest top-down signaling from the prefrontal cortex to stimulus-modality-dependent occipitotemporal regions during task switching (Tsumura et al. 2021).

We then asked a critical question whether the top-down signaling from the lPFC to the stimulus-modality-dependent regions is modulated depending on the uncertainty of the relevant task (i.e., multitask level) that was manipulated by the motion coherence of task cue. During face tasks with high-coherence cue, the task-related effective connectivity was enhanced from the MT and FFA regions to the lPFC, and the directionality of connectivity between these regions was reversed in low-coherence trials (Fig. 2E left and Supplementary Fig. S1B top). Likewise, the effective connectivity was enhanced from the MT and PPA regions to the lPFC during the place task with the high coherent cue, and the directionality of the connectivity was also reversed in low-coherence trials (Fig. 2E right and Supplementary Fig. S1B bottom). The term "reverse" here refers to the connectivity in the reverse direction between a pair of brain regions (e.g., lPFC and MT) showing an opposite task-related effect (e.g., connectivity from MT to lPFC became greater in high-coherence relative to low-coherence trials, whereas connectivity from lPFC to MT became greater in low-coherence relative to high-coherence trials).

In order to test the reliability and robustness of the results above, we estimated the effective connectivity by 1) using an alternative estimation method (Supplementary Fig. S2A,B), 2) changing the number of the ROIs in the models (Supplementary Fig. S2C,D), and 3) changing the definition of the ROIs (Supplementary Fig. S2E,F) (see Materials and Methods). Overall results were maintained, confirming that the connectivity results were robust against estimation procedures, model structures, and model parameters. Task-unrelated intrinsic connectivity is shown in Supplementary Fig. S2G,H.

These collective results of effective connectivity suggest that, when relevant task was indicated ambiguously, top-down signal from prefrontal regions become stronger toward stimulus-modality-dependent occipitotemporal regions (i.e., MT and FFA/PPA for face/place tasks). On the other hand, when task cue information is more evident, bottom-up signals from the occipitotemporal regions to the prefrontal regions become stronger.

Whole-Brain Decoding by a CNN Classifier

Given the brain regions associated with tasks, switching tasks, and perceptual decision-making identified by univariate activation analysis and their effective connectivity mechanisms, we explored brain regions that code relevant task information using a CNN classifier (LeCun et al. 2015). More specifically, we examined whether brain activity patterns involve discriminable information about face and place tasks during task switching with cue uncertainty and then identified brain regions involving critical information about the relevant task.

The current analysis used VGG16 (Simonyan and Zisserman 2015) that was trained to classify a concrete object image dataset provided by ImageNet (http://www.image-net.org/; Krizhevsky et al. 2012; Supplementary Fig. S3A). We retrained the VGG16-ImageNet model using flat whole cortical activation maps (Fig. 3A and Supplementary Fig. S3B) such that it classified face and place tasks based on fine-tuning (Donahue et al. 2014; Supplementary Fig. S3C). The retraining was performed based on cortical maps that were independent of the tested maps. We retrained the model using flat maps during a WM task for face and place stimuli obtained from the HCP (Supplementary Fig. S3C top; see also Materials and Methods), followed by additional retraining based on flat activation maps during target-only trials of the current task in which only the face/place target stimulus was presented without the dot cue stimulus (Fig. 1D and Supplementary Fig. S3C bottom).

Classification accuracy for target-only trials was $82.1 \pm 5.0\%$ (mean $\pm$ SD with 10-fold cross validation), which was significantly greater than chance level ($P < 0.001$) (Supplementary Fig S4A; see Materials and Methods). Interestingly, direct retraining of VGG16-ImageNet model to classify the current target-only trial maps showed little increase in accuracy (Supplementary Fig. S4A). Training of a randomly initialized VGG model to classify HCP WM maps was also not successful (Supplementary Fig. S4B). Thus, these results demonstrate that the current 2-step retraining of the VGG16-ImageNet model was sufficient for the CNN model to learn from small sample data sets of brain images.

Given that the CNN model successfully classified face and place tasks during the target-only trials with high accuracy, we then examined the classification accuracy for cue trials (Supplementary Fig. S3D). Accuracy was higher than chance level in switch trials at all coherence levels [80% switch: $t(28) = 4.7$, $P < 0.001$; 40% switch: $t(28) = 5.5$, $P < 0.001$; 20% switch: $t(28) = 5.4$, $P < 0.001$] and 80% repeat trials [$t(28) = 3.6$, $P < 0.005$] (Fig. 3B), which ensured those maps contained information about performed tasks. More importantly, classification accuracy was higher in switch than in repeat trials [$F(1, 28) = 10.9$; $P < 0.005$], suggesting that cortical activation patterns involve more information about task dimension in switch than in repeat trials, although the coherence effect was absent [$F(1, 28) = 0.3$; $P = 0.6$]. The interaction effect of switch and coherence was insignificant [$F(1, 28) = 0.8$; $P = 0.4$].

Because participants made button responses using the left or right buttons in both of the place and face tasks (Materials and Methods), correct responses could be made even if participants perform an incorrect alternative task. It is thus possible that the classification performance of the CNN classifier is affected by the mislabeling of activation images due to this button press procedure. However, switch effect on classification accuracy (switch vs. repeat: 0.137) was greater than that on behavioral accuracy (0.033) [$t(28) = 2.53$; $P < 0.05$], suggesting that the switch effect on classification accuracy cannot be fully explained by the image mislabeling due to the current button press procedure.

Visualization of Activation of CNN Convolution Layers

When an image was given to CNN, the CNN classifier extracts critical information through convolutional layers, which is reflected in the activations of the layers. We thus visualized activations of convolution layers to identify brain regions involving critical information to classify the face and place tasks. We used Grad-CAM (Selvaraju et al. 2017) that aggregates layer activations and highlights image locations with greater activations when important information to classify the image are involved (Fig. 3C; Materials and Methods). Because the layer activations were available for each classification, the aggregated

layer activation map was created for each of the tested images. Then, the aggregated maps were collected for each of the tasks (face/place), switching conditions (switch/repeat), and coherence levels (high/mid/low).

We first contrasted the aggregated layer activation maps between switch and repeat trials and calculated pixel-wise group-level z-statistics with participants treated as a random effect (Fig. 3D). Prefrontal, parietal, and occipitotemporal areas including FFA and PPA (see Fig. 2C and Supplementary Fig. S5 for references in 3D surface and 2D flat maps) showed greater layer activations in switch trials than in repeat trials (Fig. 3D). In particular, the layer activations became greater in trials switching to face task but not to place task in the FFA (Fig. 3D left). On the other hand, greater layer activations were observed in trials switching to the place task but not to face task in the PPA (Fig. 3D right). These results suggest that modality-dependent FFA and PPA encode task-relevant information to a greater degree during the switch to the task that demands stimulus discrimination of optimal modality. Next we examined the coherence effect during cue trials by calculating the weighted sum of layer activations for each pixel (Fig. 3E). The FFA and PPA also showed greater layer activations in high-coherence trials than low-coherence trials in both of the face and place tasks.

In order to statistically test dissociated layer activation patterns in the FFA and PPA during cue trials, ROI analysis was performed (Fig. 3F). ROIs were identical to those used in the DCM analyses (see Materials and Methods), and layer activation magnitudes were extracted from the ROIs. In the FFA, the layer activations were greater during the face task than the place task [$F(1, 28) = 15.1$, $P < 0.01$], and in the PPA, layer activations were greater during the place task than the face task [$F(1, 28) = 75.4$, $P < 0.001$]. For the layer activation magnitudes with optimal task–region relationships (i.e., face task in FFA and place task in PPA), activations became greater in switch relative to repeat trials [$F(1, 28) = 6.0$, $P < 0.05$] and in high-coherence relative to low-coherence trials [$F(1, 28) = 12.5$, $P < 0.01$ with linear contrast]; however, there was no switching-by-coherence interaction [$F(1, 28) = 0.9$, $P = 0.36$].

To examine the robustness of these results against the definition of ROIs, FFA and PPA ROIs were redefined based on target-only trials, independently of the cue trials (see Materials and Methods), and the aggregated layer activation magnitudes were calculated similarly (Supplementary Fig. S6). Again, in the FFA, layer activation was greater during the face task than the place task in the FFA [$F(1, 28) = 15.2$, $P < 0.01$] and was greater during the place task than the face task in the PPA [$F(1, 28) = 66.5$, $P < 0.001$]. For layer activation magnitudes with optimal task–region relation (i.e., face task in FFA and place task in PPA), activations became greater in switch relative to repeat trials [$F(1, 28) = 6.4$, $P < 0.05$] and in high relative to low-coherence trials [$F(1, 28) = 13.4$, $P < 0.01$ with linear contrast]. The switch-by-coherence interaction was insignificant [$F(1, 28) = 0.1$, $P = 0.75$].

Additionally, in a separate analysis, we used a leave-one-subject-out procedure when retraining the classifier to classify the activation maps of the target only trials and then tested the remaining subject (Supplementary Fig. S7). Overall results are consistent, suggesting that subject-specific noises are not dominant in our results.

These results suggest that the FFA and PPA involve more modality specialized task-related pattern information in high-coherence trials. Thus, modality-dependent occipitotemporal regions may encode relevant task information (i.e., FFA for face task and PPA for place task), which is enhanced in switch trials with a high-coherence task cue.

Whole-Brain Decoding by SVM

In order to complement decoding and mapping by the CNN classifier, we performed another decoding analysis using an SVM classifier. Similar to the CNN classifier analysis above, the classifier was trained based on HCP WM task such that the classifier discriminates the dimension of the tasks (face or place) (see Materials and Methods). We then tested the experimental data to examine classification performance for face and place tasks of the cue trials.

We found that accuracy was higher than chance level in switch trials at all coherence levels [80% switch: $t(28) = 4.5$, $P < 0.001$; 40% switch: $t(28) = 2.5$, $P < 0.05$; 20% switch: $t(28) = 3.3$, $P < 0.005$; Fig. 4A] and higher in switch than repeat trials [$F(1, 28) = 5.0$; $P < 0.05$], which is consistent with the CNN classifier results. Main effect of coherence and the interaction of switch and coherence effects were insignificant [coherence: $F(1, 28) = 1.4$, $P = 0.26$; interaction: $F(1, 28) = 1.2$, $P = 0.28$]. Switch effect on SVM accuracy (0.086) was greater than that on behavior (0.033) [$t(28) = 2.4$, $P < 0.05$], consistent with CNN classification.

Weights of the SVM classifier were mapped onto 2D cortical surface of the brain in order to identify brain regions with greater weights to classify the face and place tasks. Occipitotemporal regions including the FFA and PPA showed prominent reverse directed weights (Fig. 4B), indicating that these regions involve important information to classify the 2 tasks. Notably, these maps are consistent with the CNN-based mapping, especially in the FFA and PPA (Fig. 3D,E). We also trained another SVM based on target-only trials in the current task and found that the classification accuracy of cue trials (Supplementary Fig. S8A) and weight maps (Supplementary Fig. S8B) was consistent to those with the CNN classifier (Fig. 3B,D,E) and whole-brain SVM (Fig. 4A,B). We note that SVM weight maps (Fig. 4B and Supplementary Fig. S8B) reflect a hyperplane calculated by training data (i.e., maps for HCP WM or current target-only trials), whereas CNN activation maps reflect degree of contribution to classify tested image (i.e., current cue trial maps) (see also Discussion).

Decoding Mapping by Searchlight SVM

The above CNN and SVM classifiers were based on pattern information of whole-brain cortical regions. Another SVM analysis was also performed using the searchlight procedure (see Materials and Methods). By exploring across the whole brain, searchlight was used to identify brain regions where local image voxels involved pattern information about the performed task to classify face and place tasks. Again, the classifier in each searchlight was trained using HCP datasets, and thus the training and testing datasets were independent.

For target-only trials, classification accuracy was significantly higher in the FFA and PPA regions (Fig. 4C and Supplementary Table S4; see Supplementary Fig. S9A for 3D surface maps), suggesting that modality-dependent occipitotemporal regions involve relevant task information; this result is consistent with the univariate analysis (Fig. 2C). We also performed ROI analysis to examine accuracy in FFA and PPA. ROIs were identical to those defined in DCM analyses, and both of the PPA and FFA ROIs showed significant higher accuracy than

chance level to classify target-only trial [FFA: $t(28) = 8.8$, $P < 0.001$; PPA: $t(28) = 10.7$, $P < 0.001$] (Fig. 4C right).

Voxel-wise classification accuracy maps were contrasted between switch versus repeat trials for each participant, and group-level statistical tests were performed in order to identify brain regions where discriminable pattern information is greater in switch trials than in repeat trials. Occipitotemporal regions showed a significant effect of switching (Fig. 4D and Supplementary Table S5; see Supplementary Fig. S9B for 3D surface maps), indicating that, in these regions, classification accuracy is higher in switch relative to repeat trials, consistent with our previous study Tsumura et al. 2021). Interestingly, these regions were spatially located in-between the FFA and PPA, where pattern information of searchlight classifiers may modestly involve both of face-related and place-related information, possibly in a balanced manner (Tsumura et al. 2021). In order to examine whether PPA and FFA show differential classification accuracy between the switch and repeat trial, ROI analysis was performed. The classification accuracy for the switch and repeat trials was significantly higher than chance level in both FFA and PPA [ts(28) > 4.4; Ps < 0.001; Fig. 4D right]. More importantly, the classification accuracy for the switch trial was significantly higher than that for the repeat trial in both FFA and PPA [FFA: $t(28) = 2.1$, $P < 0.05$; PPA: $t(28) = 5.2$, $P < 0.001$].

These regions also showed a coherence effect with higher accuracy in high-coherence trials (Fig. 4E and Supplementary Table S6; see Supplementary Fig. S9C for 3D surface maps). In FFA and PPA ROIs, the coherence effect was also significant [FFA: $t(28) = 2.4$, $P < 0.05$; PPA: $t(28) = 3.7$, $P < 0.01$; Fig. 4E right]. Notably, these results were consistent with layer activation mapping of the CNN classifier (Fig. 3D,E).

It is important that these regions showed significantly higher classification accuracy than chance level in switch and repeat trials (Supplementary Fig. S10A,B and Supplementary Tables S7 and S8), and cue trials at each coherence level (Supplementary Fig. S10C–E and Supplementary Tables S9–S11). This assures that the differential accuracies between switch and repeat trials, and across coherence levels were attributable to accuracy enhancement in switch trials with the high-coherence cue.

Whole-brain exploratory analysis for the infarction effect between switching and coherence did not reveal significant brain regions.

These collective results suggest that occipitotemporal regions adjacent to stimulus-modality-dependent FFA/PPA areas involve information about ongoing task and that the information amount is increased during task switching with more coherent cue presentation. These results are also consistent with those of the classification performance and mapping based on whole-brain CNN and SVM classifiers. Such differential classification accuracy was not observed in frontoparietal regions well known to be involved in executive control (Supplementary Fig. S10F,G and Supplementary Tables S5 and S6), even when the classifier was trained by target-only trials in the current experiment (Supplementary Fig. S10H,I).

Discussion

The current study examined neural mechanisms during task switching under situation where task cue involved uncertainty. Task-related neural coding in FFA/PPA became more evident during task switching and also when the relevant task was cued more explicitly. When task cue was distinct, the lPFC received task-related signals from the MT region and PPA/FFA, and the direction of the signal was reversed when the task cue involved more ambiguity. These results suggest a distributed cortical network of fronto-occipitotemporal regions for behavioral flexibility where task-related signal among these regions helps to implement task representation depending on the ambiguities of external cue (Fig. 5).

Neural Mechanism for Task Switching and Perceptual Decision-Making

Prior work of task switching has examined switch-related neural mechanisms under situations where perceptual uncertainties were applied to target stimulus (Kayser et al. 2010; Mante et al. 2013; Zhang et al. 2013; Kumano et al. 2016; Tsumura et al. 2021). In contrast, the current study manipulated uncertainty of the task cue that involves multitask information (Figs 1C,D and 5). Notably, the task cue indicates relevant task dimension at the upper layers of the hierarchical task set structure (Koechlin et al. 2003; Bunge et al. 2005; Badre 2008; Jimura and Braver 2010; Fig. 1B), and thus, the current study allowed us to elucidate higher-level cognitive functions governing task switching and perceptual decision-making.

In the cue trials, dot cue and face/place target stimuli were simultaneously presented, as we aimed to maximize the behavioral switch and coherence effects (see Materials and Methods). Due to this presentation procedure, it is possible that the coherence effect involved general task difficulty derived from the perception of dual stimulus. However, in the cue trials, it was impossible to correctly discriminate the target picture without appropriate perception of motion direction, because the motion direction indicated the dimension of the discrimination (i.e., face or place). This is because the current task constitutes the hierarchical structure of a task set (Fig. 1B), where the upper layer representation (i.e., task to be performed) governs the lower layer representation (i.e., male/female or outdoor/indoor discrimination). Participants were also instructed to first judge the motion direction and then to discriminate the face/place picture based on the motion direction to prevent them from simultaneously perceiving the motion dot stimulus and face/place picture (see Materials and Methods). Given the nature of the current task design and training procedures, it seems less likely that the dual perception in low coherent trials dominate the negative coherence effects.

For putative mechanisms to achieve task switching under cue uncertainty, 3 hypotheses are possible: 1) a unitary mechanism implements task switching under perceptual uncertainty in a task cue; 2) distinct mechanisms for perceptual decision-making and task switching interactively guide successful task switching; and 3) a hub-like region links the 2 distinct mechanisms. Behaviorally, an interaction between switching and coherence levels was absent (Fig. 1E,F; see also Results); this is consistent with the univariate imaging analysis and CNN classifier mapping showing no interaction effect. The absence of an interaction effect, together with distinct brain regions associated with cue/target perception and task switching (Figs 2A–C and 3D–F), suggests distributed mechanisms for task switching under cue ambiguities, which supports the second hypothesis. This is consistent with our previous study, in which the interaction effect of switching and coherence of the target stimuli was absent (Tsumura et al. 2021). Thus, by modulating effective connectivity and interregional signaling depending on cue ambiguity, these regions cooperatively guide behavioral flexibility (Fig. 5). This interpretation is also compatible with the central role of frontal

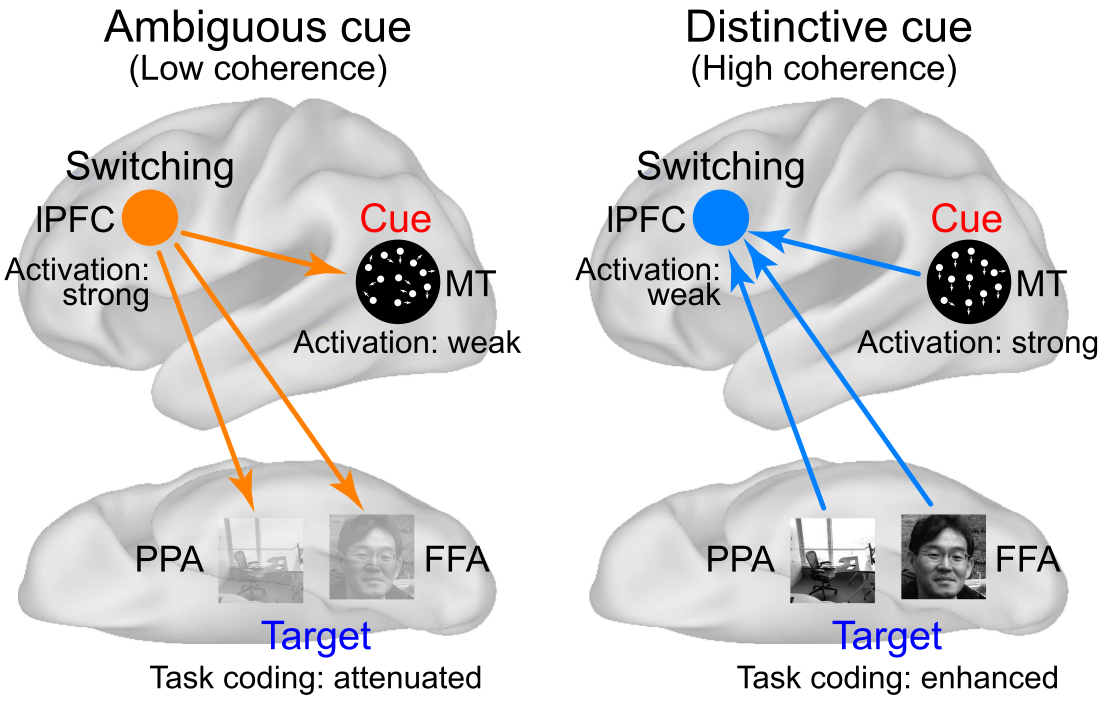

Figure 5. Putative model of behavioral flexibility under perceptual uncertainty. Schematic diagrams for functional mechanism among the MT region, FFA, PPA, and lPFC during task switching with cue uncertainty. The arrows indicate signal directions.

regions for flexible task control (Egner and Hirsch 2005; Kayser et al. 2010; Cole et al. 2013; Waskom et al. 2014).

The current results and interpretations are consistent with our recent study (Tsumura et al. 2021) in that complementary cortical mechanisms are engaged in task switching under situations where goal-relevant information is ambiguous, but extend it by showing homologous and more elaborated mechanisms help to achieve task switching when perceptual uncertainty is involved in the task cue instead of the target stimuli.

Specifically, in our previous study (Tsumura et al. 2021), the target stimulus involved perceptual uncertainty of dot motion and color, and the cue stimulus was presented without perceptual uncertainty. Top-down signals from the prefrontal to occipitotemporal cortices guide task switching, which is supplemented by the contralateral prefrontal regions when the target stimulus involved more uncertainty.

On the other hand, in the current study, the cue stimuli involved the perceptual uncertainty of the motion dot stimulus. The dot stimulus was identical to that used in the previous study (Tsumura et al. 2021). The signaling between the prefrontal and occipitotemporal cortices was reversed depending on the perceptual uncertainty of the cue stimulus, which was not observed in the previous study that manipulated the perceptual uncertainty of the target stimulus.

In reference to the task set hierarchy (Fig. 1B), our previous study applied perceptual uncertainty of dot motion to the lower layer of the hierarchy (i.e., target), while the current study applied an identical perceptual uncertainty to the upper layer of the hierarchy (i.e., cue). Because task-relevant information in the lower layer is under the control of upper layer information, task-related signals flow from the upper to lower layers (Koechlin et al. 2003; Badre 2008), which our previous study demonstrated in the prefrontal and occipitotemporal cortices (Tsumura et al. 2021).

On the other hand, in the current study, when the cue was presented more distinctively, bottom-up signals from the occipitotemporal to prefrontal cortices helped to guide task switching, because the upper layer information about the relevant task requires an appropriate perception of the dot motion stimulus.

One common, important observation in our current and previous studies is that when perceptual uncertainty was increased, activation was enhanced in the prefrontal cortex, and top-down signals from the prefrontal to occipitotemporal regions became more prominent. This signaling may complement the accumulation of stimulus evidence in the occipitotemporal cortex when the visual stimulus involved more uncertainty (Gold and Shadlen 2007).

The complexity and sophistication of the mechanisms are attributable to the increased number of occipitotemporal regions: In the current study, 3 regions, MT, FFA, and PPA, are involved, whereas 2 regions, MT and ventral visual complex, are involved in our prior study (Tsumura et al. 2021).

Fronto-occipitotemporal Network Mechanisms

DCM analyses revealed top-down signal from the lPFC to the FFA or PPA, depending on the task to be switched, and this top-down signaling is enhanced when task cue involved ambiguity during switching (Fig. 5). The top-down mechanisms may reflect supplemental attention to visual stimulus required to collect task cue information about the tasks to be performed (Zanto et al. 2011; Lee and D'Esposito 2012). The supplemental attention involving frontal engagement may complement stimulus-modality-dependent activation in occipitotemporal regions (Desimone and Duncan 1995; Kastner and Ungerleider 2000; Corbetta and Shulman 2002; Lewis-Peacock and Postle 2008).

In contrast, bottom-up signaling with increased cue information may reflect the conversion of sensory information to

behavioral information through an information stream from the visual sensory area to executive control areas (Desimone and Duncan 1995; Kastner and Ungerleider 2000). Thus, when the task cue was apparent, the bottom-up signal was strengthened because cue-related information is more available, which may help to enhance task switching performance.

On the other hand, such signal reversal did not occur in our prior study, and the top-down signals from prefrontal to occipitotemporal regions complement switching independently of the uncertainty of the target stimulus (Tsumura et al. 2021). It is possible that the signal reversal is characteristic of the uncertainty of the cue stimulus, as the cue stimulus indicates more abstract information in the upper layers of the task-set hierarchy (Fig. 1B). Thus, the uncertainty of the abstract information may implicate more complex mechanisms to implement task sets to be performed (Koechlin et al. 2003; Bunge et al. 2005; Badre 2008; Jimura and Braver 2010).

The stronger connectivity from FFA/PPA to lPFC in high-coherence trials may reflect the efficient generation of response for the face/place task derived from quicker judgment about the motion direction. On the other hand, in low coherence trials, because the direction judgment was slower, the stronger top-down signal from lPFC and FFA/PPA may complement the slower judgment to make a correct response for the face/place stimulus.

Additionally, from the overall inspection of the entire set of connectivity results (Fig. 2E and Supplementary Fig. S2B,D,F), there seems to be a small tendency for the connectivity to be stronger between MT and FFA in the high-coherence face trials and between MT and PPA in the high-coherence place trials.

In the cue trial, after the judgment for dot motion direction, the dot motion stimulus became unnecessary to make a correct response, but participants then needed to perceive face or place in the picture in accordance with the judgment. Such an attention shift from dot motion to face/place picture may occur faster in high-coherence trials than in low-coherence trial. Additionally, stimulus evidence of motion direction in MT was accumulated faster in high-coherence trials (Gold and Shadlen 2007), and moreover, the MT, FFA, and PPA engagements in the cue trial reflect the perception of goal-relevant information rather than a simple response to visual stimulus (Kayser et al. 2010; Tsumura et al. 2021; Fig. 2C and Supplementary Fig. S5). Thus, the direct connectivity between occipitotemporal regions in high-coherence trials may reflect enhanced signaling, reflecting faster accumulation of task-relevant stimulus evidence in MT and a swift attention shift from the dot motion to the face/place picture.

Comparisons of Classification and Mapping among Machine Learning Techniques

In the current study, whole-brain exploration of task-related neural representation was performed by 3 approaches based on 3 machine learning techniques, 1) CNN classifier, 2) whole-brain cortical SVM, and 3) searchlight SVM.

CNN classified activation maps along task dimensions based on all pixels across whole cortical regions, and the classification accuracy was higher in switch trials than repeat trials. One novel signature of the current CNN classifier approach is that brain regions involving critical information to classify cue trials were mapped by aggregated activation across convolution layers based on the Grad-CAM technique. It is notable that this CNN activation mapping is available on image-by-image basis for testing data, which is not the case for the SVM mapping using whole cortical images. Then, the aggregated layer activation mapping of CNN revealed that task representation in the FFA during the face task and in the PPA during the place task was enhanced in switch and more coherent trials than in repeat and less coherent trials. Increased task-related activation (Fig. 2C and Supplementary Fig. S5) may be associated with higher classification accuracy and enhanced task representation.

Standard voxel-wise univariate GLM analysis identifies brain regions where the MRI signal is differentiated between task conditions but does not necessarily indicate that identified brain regions are critical for task performance; this makes it hard to identify brain regions playing an important role in cognitive functions (i.e., reverse inference). In contrast, the current CNN classifier demonstrated that the visualization of convolution layers of the classifier for brain activation is useful in identifying brain regions that characterize task performance.

The analysis based on the CNN classifier was complemented by a standard SVM analysis for whole-brain cortical maps that tested identical map images. The classification accuracy of cue trials was highly consistent between the 2 classifiers. Additionally, SVM weight maps also showed differential weights in the FFA and PPA, which is also consistent with the CNN classifier. One notable technical limitation of the SVM is that weight mapping for testing classification for cue trials was unavailable, unlike Grad-CAM of CNN classifier. Thus, the SVM weight map indicates that the FFA and PPA are critical to classify tasks for the HCP N-back WM task or target-only trials in the current task (i.e., training data), but not necessarily for cue trials of the current task (i.e., testing data). Nonetheless, together with differential classification accuracy among cue conditions, the SVM weight maps suggest that pattern information in the FFA and PPA is distinct (i.e., distant from separating hyperplane) in switch trials than in repeat trials.

Another approach to identify brain regions that characterize task performance is the searchlight SVM, which also allows whole-brain exploration of activation patterns, but individual classifications were restricted in local brain regions (Fig. 4C–E; Supplementary Figs S9 and S10). This is in contrast to the CNN classifier and whole-brain cortical SVM that are trained and classified based on a whole-brain image. Nonetheless, results were complementary to those whole brain-based classifiers in that 1) occipitotemporal regions adjacent to the FFA and PPA were capable of task classification, and 2) the classification performance in these regions became higher in switch and more coherent trials.

The higher classification accuracy in switch trial is consistent with our recent study (Tsumura et al. 2021) whereas previous MVPA studies of task switching suggested that task coding in frontoparietal regions is attenuated in a switch trial (Qiao et al. 2017), and task coding is independent of task switching (Loose et al. 2017). Interestingly, those previous studies used task-cueing paradigms, where a task cue was presented in each trial (Loose et al. 2017; Qiao et al. 2017); conversely, the current study and our recent study used an intermittent cue paradigm in which the switch trial occurred after successive correct trials for the alternative task without presenting a cue (Fig. 1C,D; Tsumura et al. 2021). The variability in the cueing procedures among the studies may yield the variability of classification accuracy.

In the current task, correct responses could be made even if participants perform an incorrect alternative task, because the identical set of buttons was used in the place and face tasks. Due to this button press procedure, the classification performance of the whole-brain CNN and SVM classifiers could be affected

by the mislabeling of activation images. However, the greater switch effect on classification accuracy than behavioral accuracy suggests that the switch effect on classification accuracy cannot be fully explained by the image mislabeling. On the other hand, the absence of the coherence and interaction effects on the classification accuracy may be attributable to the image mislabeling. Nonetheless, in low-coherence trials, because of the increased perceptual uncertainty, participants were uncertain about the task to be performed, rather than certain to perform the incorrect task by incorrectly perceiving the motion stimulus. Thus, it seems less likely that the mislabeling due to the performance of the incorrect task with certainty produces critical bias for our results regarding the coherence effects on classification accuracy.

Classifier Training Using Independent Open Resource Data

One notable analysis procedure in the current machine learning-based functional brain mapping is that classifier was trained using an open resource dataset that was independently collected from the current experiment. This procedure ensured independence between the training and testing data.

Task switch and WM may involve distinct cognitive control demands, with the former related to behavioral flexibility (Allport et al. 1994; Rogers and Monsell 1995) and the latter related to active maintenance and updates of goal-relevant information (D'Esposito and Postle 2015). However, the 2 tasks used common visual stimulus categories (face and place); thus, perceptual demands may involve some degree of commonality. Recognition demands for the presented stimuli were also distinct: the current task-switching paradigm used male–female and indoor–outdoor discriminations during face and place tasks, respectively, but HCP WM tasks used discrimination of identicalness to past stimuli. Additionally, the current task used face–place superimposed stimuli, and thus, the identical stimulus set was used during face and place tasks, whereas HCP WM task used distinct visual stimulus sets during face and place blocks. Thus, task representation examined in the current study may reflect visual perception or attention rather than low level visual features.

For the CNN classifier, training involved 2 steps: retraining of HCP WM maps and additional retraining of target-only trials. CNN classifies maps based on whole cortical areas including frontoparietal regions in which differential subregions are recruited during task switching (Dove et al. 2000; Rushworth et al. 2002; Koechlin et al. 2003; Bunge et al. 2005; Derrfuss et al. 2005; Crone et al. 2006; Yeung et al. 2006; Jimura and Braver 2010; Kim et al. 2012; Nee and D'Esposito 2016; Bissonette and Roesch 2017; Malagon-Vina et al. 2018; Fouragnan et al. 2019; Tsumura et al. 2021) and WM (Courtney et al. 1997; Miller and Cohen 2001; D'Esposito and Postle 2015). Thus, additional retraining of the CNN model based on identical recognition demands (i.e., target-only trials) was effective in classifying tasks to optimize the CNN model for classification using whole cortical images. Distinct layer activation differences in the parietal cortex (Fig. 3*D*,*E*) may partially be attributable to higher performance with the additional retraining.

Because incremental training of HCP WM trials and the target-only trials in the current study is irrelevant to SVM, these 2 datasets were separately trained for whole-brain cortical SVM, and weight maps were consistent especially in occipitotemporal regions (Fig. 4*B* and Supplementary Fig. S8B). Importantly, classification accuracy for the cue conditions was consistent in SVMs with the 2 training datasets and also with the CNN classifier. The sample size was much smaller for the current target-only trials than HCP dataset, but classification performance was comparable between those 2 classifiers (Fig. 4*A* and Supplementary Fig. S8A). Thus, SVM may thus not require a larger sample size like the HCP data for training, while CNN training needed incremental training even with large sets of image data.

The searchlight SVM using HCP WM maps as training data identified occipitotemporal regions spatially closed to the FFA/PPA showing higher classification accuracy for target-only trials. Moreover, classification accuracy was higher in high-coherence switch trials. Interestingly, these classification results were absent in frontoparietal regions, well known to be involved in executive control (Supplementary Fig. S10F,G), even when the searchlight classifier was trained by the target-only trials of the current task (Supplementary Fig. S10H,I). One possibility for this discrepancy is that control and recognition demands are incompatible while perceptual modality is compatible in HCP WM trials, target-only trials, and switch/repeat trials. Then, the distinct control and recognition demands might be reflected in classification incompatibility in the frontoparietal regions.

A CNN model successfully learned and decoded task-related fMRI images of HCP without flattening the images, and retraining of the classifier was also effective for small subset of the HCP data (Wang et al. 2020). However, this technique is unavailable for the current experiment using event-related design (see Materials and Methods for more details), and retraining of the VGG16/ImageNet model based on flattened 2D activation maps was powerful for the current dataset.

Supplementary Material

Supplementary material is available at *Cerebral Cortex* online.

Authors' Contributions

K.T. and K.J. designed the experiment and study. K.T., R.A., K.N., and K.J. collected the data. K.T., K.K., and K.J. analyzed the data. M.T., J.C., and Y.H. contributed to the analysis design of imaging data and to the development of machine learning classifier. K.T., M.T., J.C., K.N., and K.J. wrote the manuscript.

Funding

Kakenhi (Japan Society for the Promotion of Science) (19H04914, 17K01989, 17H05957, 17H00891, 26350986, and 26120711 to K.J.; 20H00521, 18H04953, and 18H05140 to M.T.; 18H05017 to J.C.; 17H00891 to K.N.); Uehara Memorial Foundation (grant to K.J.); Takeda Science Foundation (grant to K.J. and M.T.); Japan Agency for Medical Research and Development (AMED) (JP20dm0207086 to J.C.).

Notes

We thank Drs Akira Funahashi, Shori Nishimoto, and Teppei Matsui for scientific comments on the study and manuscript. We thank Ms Maoko Yamanaka for administrative assistance. *Conflict of Interest*: The authors declare no competing interests.

References

Allport DA, Styles EA, Hsieh S. 1994. *Attention and performance XV: conscious and nonconscious information processing*. Cambridge (MA): The MIT Press.

Badre D. 2008. Cognitive control, hierarchy, and the rostro-caudal organization of the frontal lobes. *Trends Cogn Sci*. 12: 193–200.

Barch DM, Burgess GG, Harms MP, Petersen SE, Schlaggar BL, Corbetta M, Glasser MF, Curtiss S, Dixit S, Feldt C, et al. 2013. Function in the Human Connectome: task-fMRI and individual differences in behavior. *Neuroimage*. 80:169–189.

Beauchamp MS, Cox RW, Deyoe EA. 1997. Graded effects of spatial and featural attention on human area MT and associated motion processing areas. *J Neurophysiol*. 78:516–520.

Bissonette GB, Roesch MR. 2017. Neurophysiology of rule switching in the corticostriatal circuit. *Neuroscience*. 345:64–76.

Brainard DH. 1997. The psychophysics toolbox. *Spat Vis*. 10:433–436.

Britten KH, Shadlen MN, Newsome WT, Movshon JA. 1993. Responses of neurons in macaque MT to stochastic motion signals. *Vis Neurosci*. 10:1157–1169.

Bunge SA, Wallis JD, Parker A, Brass M, Crone EA, Hoshi E, Sakai K. 2005. Neural circuitry underlying rule use in human and nonnuman primates. *J Neurosci*. 25:10347–10350.

Chen MY, Jimura K, White CN, Maddox WT, Poldrack RA. 2015. Multiple brain networks contribute to the acquisition of bias in perceptual decision-making. *Front Neurosci*. 9:63.

Chikazoe J, Lee D, Kriegeskorte N, Anderson AK. 2019. Distinct representation of basic taste qualities in human gustatory cortex. *Nat Commun*. 10:1048.

Cole MW, Reynolds JR, Power JD, Repovs G, Anticevic A, Braver T. 2013. Multi-task connectivity reveals flexible hubs for adaptive task control. *Nat Neurosci*. 16:1348–1355.

Corbetta M, Miezin FM, Dobmeyer S, Shulman GL, Petersen SE. 1991. Selective and divided attention during visual discriminations of shape, color, and speed: functional-anatomy by positron emission tomography. *J Neurosci*. 11:2383–2402.

Corbetta M, Shulman GL. 2002. Control of goal-directed and stimulus-driven attention in the brain. *Nat Rev Neurosci*. 3:201–215.

Courtney SM, Ungerleider LG, Keil K, Haxby JV. 1997. Transient and sustained activity in a distributed neural system for human working memory. *Nature*. 386:608–611.

Crone EA, Wendelken C, Donohue SE, Bunge SA. 2006. Neural evidence for dissociable components of task-switching. *Cereb Cortex*. 16:475–486.

D'Esposito M, Postle BR. 2015. The cognitive neuroscience of working memory. *Annu Rev Psychol*. 66:115–142.

Deng J, Dong W, Socher R, Li LJ, Li K, Fei-Fei L. 2009. ImageNet: a large-scale hierarchical image database. 2009 IEEE Conference on Computer Vision and Pattern Recognition. Miami (FL): IEEE. p. 248–255.

Derrfuss J, Brass M, Neumann J, von Cramon DY. 2005. Involvement of the inferior frontal junction in cognitive control: meta-analyses of switching and Stroop studies. *Hum Brain Mapp*. 25:22–34.

Desimone R, Duncan J. 1995. Neural mechanisms of selective visual-attention. *Annu Rev Neurosci*. 18:193–222.

Donahue J, Jia Y, Vinyals O, Hoffman J, Zhang N, Tzeng E, Darrell T. 2014. Decaf: a deep convolutional activation feature for generic visual recognition. *Proc 31st Int Conf Mach Learn*. 32:647–655.

Dove A, Pollmann S, Schubert T, Wiggins CJ, von Cramon DY. 2000. Prefrontal cortex activation in task switching: an event-related fMRI study. *Cogn Brain Res*. 9:103–109.

Egner T, Hirsch J. 2005. Cognitive control mechanisms resolve conflict through cortical amplification of task-relevant information. *Nat Neurosci*. 8:1784–1790.

Eklund A, Nichols TE, Knutsson H. 2016. Cluster failure: inflated false positives for fMRI. *Proc Natl Acad Sci U S A*. 113:7900–7905.

Epstein R, Kanwisher N. 1998. A cortical representation of the local visual environment. *Nature*. 392:598–601.

Fouragnan EF, Chau BKH, Folloni D, Kolling N, Verhagen L, Klein-Flügge M, Tankelevitch L, Papageorgiou GK, Aubry JF, Sallet J, et al. 2019. The macaque anterior cingulate cortex translates counterfactual choice value into actual behavioral change. *Nat Neurosci*. 22:797–808.

Freiwald WA, Tsao DY. 2010. Functional compartmentalization and viewpoint generalization within the macaque face-processing system. *Science*. 330:845–851.

Friston KJ, Harrison L, Penny W. 2003. Dynamic causal modelling. *Neuroimage*. 19:1273–1302.

Friston KJ, Kitvak V, Oswal A, Razi A, Stephan KE, van Wijk BCM, Ziegler G, Zeidman P. 2016. Bayesian model reduction and empirical Bayes for group (DCM) studies. *Neuroimage*. 128:413–431.

Gazzaley A, Cooney JW, McEvoy K, Knight RT, D'Esposito M. 2005. Top-down enhancement and suppression of the magnitude and speed of neural activity. *J Cogn Neurosci*. 17:507–517.

Glasser MF, Smith SM, Marcus DS, Andersson JL, Auerbach EJ, Behrens TE, Coalson TE, Harms MP, Jenkinson M, Moeller S, et al. 2016a. The Human Connectome Project's neuroimaging approach. *Nat Neurosci*. 19:1175–1178.

Glasser MF, Coalson TS, Robinson EC, Hacker CD, Harwell J, Yacoub E, Ugurbil K, Andersson J, Beckmann CF, Jenkinson M, et al. 2016b. A multi-modal parcellation of human cerebral cortex. *Nature*. 536:171–178.

Gold JJ, Shadlen MN. 2007. The neural basis of decision making. *Annu Rev Neurosci*. 30:535–574.

Hanks TD, Summerfield C. 2017. Perceptual decision making in rodents, monkeys, and humans. *Neuron*. 93:15–31.

Haynes JD, Rees G. 2006. Decoding mental states from brain activity in humans. *Nat Rev Neurosci*. 7:523–534.

Huk AC, Dougherty RF, Heeger DJ. 2002. Retinotopy and functional subdivision of human areas MT and MST. *J Neurosci*. 22:7195–7205.

Ishai A, Ungerleider LG, Martin A, Schouten JL, Haxby JV. 1999. Distributed representation of objects in the human ventral visual pathway. *Proc Natl Acad Sci U S A*. 96:9379–9384.

Jimura K, Braver TS. 2010. Age-related shifts in brain activity dynamics during task switching. *Cereb Cortex*. 20:1420–1431.

Jimura K, Poldrack RA. 2012. Analyses of regional-average activation and multivoxel pattern information tell complementary stories. *Neuropsychologia*. 50:544–552.

Kamitani Y, Tong F. 2005. Decoding the visual and subjective contents of the human brain. *Nat Neurosci*. 8:679–685.

Kanwisher N, McDermott J, Chun MN. 1997. The fusiform face area: a module in human extrastriate cortex specialized for face perception. *J Neurosci*. 17:4302–4311.

Kastner S, Ungerleider LG. 2000. Mechanisms of visual attention in the human cortex. *Annu Rev Neurosci*. 23:315–341.

Kayser AS, Erickson DT, Bushsdaum BR, D'Esposito M. 2010. Neural representations of relevant and irrelevant features in perceptual decision making. *J Neurosci*. 30:15778–15789.

Kim C, Cilles SE, Johnson NF, Gold BT. 2012. Domain general and domain preferential brain regions associated with different types of task switching: a meta-analysis. *Hum Brain Mapp*. 33:130–142.

Koechlin E, Ody C, Kouneiher F. 2003. The architecture of cognitive control in the human prefrontal cortex. *Science*. 302:1181–1185.

Konishi S, Hayashi T, Uchida I, Kikyo H, Takahashi E, Miyashita Y. 2002. Hemispheric asymmetry in human lateral prefrontal cortex during cognitive set shifting. *Proc Natl Acad Sci U S A*. 99:7803–7808.

Kriegeskorte N, Goebel R, Bandettini P. 2006. Information-based functional brain mapping. *Proc Natl Acad Sci U S A*. 103:3863–3868.

Krizhevsky A, Sutskever I, Hinton G. 2012. Imagenet classification with deep convolutional neural network. *Adv Neural Inf Proc Sys*. 25:1097–1105.

Kumano H, Suda Y, Uka T. 2016. Context-dependent accumulation of sensory evidence in the parietal cortex underlies flexible task switching. *J Neurosci*. 36:12192–12202.

Kurikawa T, Haga T, Handa T, Harukuni R, Fuai T. 2018. Neuronal stability in medial frontal cortex set individual variability in decision-making. *Nat Neurosci*. 21:1764–1773.

LeCun Y, Bengio Y, Hinton G. 2015. Deep learning. *Nature*. 521:436–444.

Lee TG, D'Esposito M. 2012. The dynamic nature of top-down signals originating from prefrontal cortex: a combined fMRI-TMS study. *J Neurosci*. 32:15458–15466.

Lewis-Peacock JA, Postle BR. 2008. Temporary activation of long-term memory supports working memory. *J Neurosci*. 28:8765–8771.

Loose LS, Wisniewski D, Rusconi M, Goschke T, Haynes JD. 2017. Switch-independent task representations in frontal and parietal cortex. *J Neurosci*. 37:8033–8042.

Malagon-Vina H, Ciocchi S, Passecker J, Dorffner G, Klausberger T. 2018. Fluid network dynamics in the prefrontal cortex during multiple strategy switching. *Nat Commun*. 9:309.

Mante V, Sussillo D, Shenoy KV, Newsome WT. 2013. Context-dependent computation by recurrent dynamics in prefrontal cortex. *Nature*. 503:78–84.

McCarthy G, Puce A, Gore JC, Allison T. 1997. Face-specific processing in the human fusiform gyrus. *J Cogn Neurosci*. 9:605–610.

Miller EK, Cohen JD. 2001. An integrative theory of prefrontal cortex function. *Annu Rev Neurosci*. 24:167–202.

Misaki M, Kim Y, Bandettini PA, Kriegeskorte N. 2010. Comparison of multivariate classifiers and response normalization for pattern-information fMRI. *Neuroimage*. 53:103–118.

Moeller S, Yacoub E, Olman CA, Auerbach E, Strupp J, Harel N, Uğurbil K. 2010. Multiband multislice GE-EPI at 7 tesla, with 16-fold acceleration using partial parallel imaging with application to high spatial and temporal whole-brain fMRI. *Magn Reson Med*. 63:1144–1153.

Nakahara K, Adachi K, Kawasaki K, Matsuo T, Sawahata H, Majima K, Takeda M, Sugiyama S, Nakata R, Iijima A, et al. 2016. Associative-memory representations emerge as shared spatial patterns of theta activity spanning the primate temporal cortex. *Nat Commun*. 7:11827.

Nee DE, D'Esposito M. 2016. The hierarchical organization of the lateral prefrontal cortex. *Elife*. 5:e12112.

Newsome WT, Pare EBA. 1988. A selective impairment of motion perception following lesions of the middle temporal visual area (MT). *J Neurosci*. 8:2201–2211.

Nichols TE, Holmes AP. 2001. Nonparametric permutation tests for functional neuroimaging: a primer with examples. *Hum Brain Mapp*. 15:1–25.

Nishimoto S, Vu AT, Naselaris T, Benjamini Y, Yu B, Gallant JL. 2011. Reconstructing visual experiences from brain activity evoked by natural movies. *Curr Biol*. 21:1641–1646.

Norman KA, Polyn SM, Detre GJ, Haxby JV. 2006. Beyond mind-reading: multi-voxel pattern analysis of fMRI data. *Trends Cogn Sci*. 10:424–430.

Palmer J, Huk AC, Shadlen MN. 2005. The effect of stimulus strength on the speed and accuracy of a perceptual decision. *J Vis*. 5:376–404.

Pan SJ, Yang QA. 2010. A survey on transfer learning. *IEEE Trans Knowl Data Eng*. 22:1345–1359.

Penny WD, Stephan KE, Mechelli A, Friston KJ. 2004. Comparing dynamic causal models. *Neuroimage*. 22:1157–1172.

Penny WD, Stephan KE, Daunizzeau J, Rosa MJ, Friston KJ, Schofield TM, Leff AP. 2010. Comparing families of dynamic causal models. *PLoS Comput Biol*. 6:e1000709.

Penny WD. 2012. Comparing dynamics causal models using AIC, BIC, and free energy. *Neuroimage*. 59:319–330.

Qiao L, Zhang L, Chen A, Egner T. 2017. Dynamic trial-by-trial recoding of task-set representations in the frontoparietal cortex mediates behavioral flexibility. *J Neurosci*. 37:11037–11050.

Rogers RD, Monsell S. 1995. Cost of a predictable switch between simple cognitive tasks. *J Exp Psychol Gen*. 124:207–231.

Rushworth MF, Passingham RE, Nobre AC. 2002. Components of switching intentional set. *J Cogn Neurosci*. 14:1139–1150.

Sakai, K. 2008. Task set and prefrontal cortex. *Annu Rev Neurosci*. 31:129–245.

Selvaraju RR, Cogswel M, Das A, Vedantam R, Parikh D, Batra D. 2017. Grad-CAM: visual explanations from deep networks via gradient-based localization. 2017 IEEE Inter Conf Comp Vis. Venice, Italy: IEEE. p. 618–626.

Shadlen MN, Britten KH, Newsome WT, Movshon JA. 1996. A computational analysis of the relationship between neuronal and behavioral responses to visual motion. *J Neurosci*. 16:1486–1510.

Simonyan K, Zisserman A. 2015. Very deep convolutional networks for large-scale image recognition. arXiv.1409.1556.

Smith SM, Miller KL, Salimi-Khorshidi G, Webster M, Beckmann CF, Nichols TE, Ramsey JD, Woolrich MW. 2011. Network modelling methods for FMRI. *Neuroimage*. 54:875–891.

Stephan KE, Penny WD, Moran RJ, den Ouden HE, Daunizeau J, Friston KJ. 2010. Ten simple rules for dynamic causal modeling. *Neuroimage*. 49:3099–3109.

Stoet G, Snyder LH. 2009. Neural correlates of executive control functions in the monkey. *Trends Cogn Sci*. 13:228–234.

Tsumura K, Aoki R, Takeda M, Nakahara K, Jimura K. 2021. Cross-hemispheric complementary prefrontal mechanisms during task switching under perceptual uncertainty. *J Neurosci*. 41:2197–2213.

Vapnik VN. 1998. *Statistical learning theory*. New York: Wiley.

Wang X, Liang X, Jiang Z, Nguchu BA, Zhou Y, Wang Y, Wang H, Li Y, Zhu Y, Wu F, et al. 2020. Decoding and mapping task states of the human brain via deep learning. *Hum Brain Mapp*. 41:1505–1519.

Waskom ML, Kumaran D, Gordon AM, Rissman J, Wagner AD. 2014. Frontoparietal representations of task context support the flexible control of goal-directed cognition. *J Neurosci*. 34:10743–10755.

Worsley DB, Friston KJ. 1995. Analysis of fMRI time-series revisited again. *Neuroimage*. 2:173–181.

Yarkoni T, Poldrack RA, Nichols TE, van Essen DC, Wager TD. 2011. Large-scale automated synthesis of human functional neuroimaging data. *Nat Methods*. 8:665–670.

Yeung N, Nystrom LE, Aronson JA, Cohen JD. 2006. Between-task competition and cognitive control in task switching. *J Neurosci*. 26:1429–1438.

Zanto TP, Rubens MT, Thangavel A, Gazzaley A. 2011. Causal role of the prefrontal cortex in top-down modulation of visual processing and working memory. *Nat Neurosci*. 14: 656–661.

Zhang J, Kriegeskorte N, Carlin JD, Rowe JB. 2013. Choosing the rules: distinct and overlapping frontoparietal representations of task rules for perceptual decisions. *J Neurosci*. 33:11852–11862.