Behavioral/Cognitive

Dynamic Trial-by-Trial Recoding of Task-Set Representations in the Frontoparietal Cortex Mediates Behavioral Flexibility

Lei Qiao,[1] **Lijie Zhang,**[1] **Antao Chen,**[1] **and Tobias Egner**[2]

[1]Key Laboratory of Cognition and Personality of Ministry of Education, Faculty of Psychology, Chongqing Collaborative Innovation Center for Brain Science, Southwest University, Chongqing 400715, China, and [2]Center for Cognitive Neuroscience, Department of Psychology and Neuroscience, Duke University, Durham, North Carolina 27708

Cognitive flexibility forms the core of the extraordinary ability of humans to adapt, but the precise neural mechanisms underlying our ability to nimbly shift between task sets remain poorly understood. Recent functional magnetic resonance imaging (fMRI) studies employing multivoxel pattern analysis (MVPA) have shown that a currently relevant task set can be decoded from activity patterns in the frontoparietal cortex, but whether these regions support the dynamic transformation of task sets from trial to trial is not clear. Here, we combined a cued task-switching protocol with human (both sexes) fMRI, and harnessed representational similarity analysis (RSA) to facilitate a novel assessment of trial-by-trial changes in neural task-set representations. We first used MVPA to define task-sensitive frontoparietal and visual regions and found that neural task-set representations on switch trials are less stably encoded than on repeat trials. We then exploited RSA to show that the neural representational pattern dissimilarity across consecutive trials is greater for switch trials than for repeat trials, and that the degree of this pattern dissimilarity predicts behavior. Moreover, the overall neural pattern of representational dissimilarities followed from the assumption that repeating sets, compared with switching sets, results in stronger neural task representations. Finally, when moving from cue to target phase within a trial, pattern dissimilarities tracked the transformation from previous-trial task representations to the currently relevant set. These results provide neural evidence for the longstanding assumptions of an effortful task-set reconfiguration process hampered by task-set inertia, and they demonstrate that frontoparietal and stimulus processing regions support "dynamic adaptive coding," flexibly representing changing task sets in a trial-by-trial fashion.

Key words: cognitive control; cognitive flexibility; fMRI; frontoparietal cortex; representational similarity analysis; task switching

---

Significance Statement

Humans can fluently switch between different tasks, reflecting an ability to dynamically configure "task sets," rule representations that link stimuli to appropriate responses. Recent studies show that neural signals in frontal and parietal brain regions can tell us which of two tasks a person is currently performing. However, it is not known whether these regions are also involved in dynamically reconfiguring task-set representations when switching between tasks. Here we measured human brain activity during task switching and tracked the similarity of neural task-set representations from trial to trial. We show that frontal and parietal brain regions flexibly recode changing task sets in a trial-by-trial fashion, and that task-set similarity over consecutive trials predicts behavior.

Introduction

"Cognitive control" describes a set of processes that allows individuals to coordinate thoughts and actions in accordance with internal goals (Miller and Cohen, 2001). A key component of

cognitive control is the capacity to flexibly configure, and switch between, "task sets," rule representations that define currently relevant stimulus attributes and determine how they map onto responses (for review, see Monsell, 2003; Kiesel et al., 2010; Vandierendonck et al., 2010). Shifting task sets incurs a "switch cost," as reflected in slower and more error-prone responses when participants are cued to perform a different task (switch trials) from the previous trial, compared to when participants repeat the same task (repeat trials). A large behavioral literature suggests that this response-time (RT) cost is attributable to a combination of time required for resolving proactive interference from residual activation of the previous, no-longer relevant set ("task-set inertia"; Allport et al., 1994; Wylie and Allport, 2000) and of time required for constructing or retrieving the newly cued set ("task-set recon-

figuration"; Rogers and Monsell, 1995; Meiran, 1996, 2000). Moreover, higher error rates on switch trials indicate that a newly reconfigured set is more labile than a repeated set. However, the exact way in which the flexible transition between task sets across trials maps onto dynamic neural coding of task representations in the human brain is presently not well understood.

One major reason for this is that most functional magnetic resonance imaging (fMRI) studies delineating neural substrates of task switching have simply contrasted mean activity levels between switch and repeat cues or trials. This work has implicated the lateral inferior frontal cortex, the presupplementary motor area (pre-SMA), superior and inferior parietal cortices, and the striatum in switch processes (Dove et al., 2000; Sohn et al., 2000; Brass and von Cramon, 2002, 2004; Johnston et al., 2007; Braem et al., 2013; Ruge et al., 2013; Korb et al., 2017). However, mean activation contrasts cannot speak directly to these regions' potential roles in representing and modifying task sets. For that purpose, multivariate, information-based brain mapping techniques, like multivoxel pattern analysis (MVPA: Haynes and Rees, 2006; Kriegeskorte et al., 2006) are required. A handful of recent studies employing such multivariate methods have shown that the currently relevant task set can be decoded from activity patterns in frontoparietal and visual cortices (Woolgar et al., 2011; Waskom et al., 2014; Wisniewski et al., 2015), and that task distinctiveness in these areas can be enhanced by reward (Etzel et al., 2016) and training (Garner and Dux, 2015). These studies suggest a key role for frontoparietal and stimulus processing regions in representing task sets, but they do not characterize the role of these regions in the dynamic reconfiguration of task-set representations from trial to trial.

In the present study, we generated a set of fundamental predictions about neural task-set representations based on the cognitive literature, and tested those hypotheses by combining fMRI during a cued task-switching protocol with MVPA and representational similarity analyses (RSA; Kriegeskorte and Kievit, 2013). Specifically, we first replicated previous studies (Woolgar et al., 2011; Waskom et al., 2014) in using MVPA to decode the currently relevant task set from activation patterns in frontoparietal and visual regions, and we show that task sets are encoded with less fidelity on switch trials than on repeat trials. Subsequently, we used RSA to probe trial-to-trial dynamics of task-set representations. We first show that neural pattern similarity across trials is greater for repeat trials than for switch trials, and that the degree of this cross-trial task-set (dis)similarity predicts behavior. We then use RSA to track the within-trial transformation of task-set representations, showing how, on switch trials, multivoxel activity patterns gradually become more dissimilar to the no-longer relevant, previous-trial task set (and more similar to the newly relevant task set) when moving from processing of the switch cue to the onset of the imperative task stimulus. Together, these data show that behavioral rule switching is supported by dynamic trial-by-trial recoding of task-set representations in frontoparietal and stimulus processing regions.

Materials and Methods

Participants. Forty-six right-handed, healthy young adults participated in our study. All of them had normal or corrected-to-normal vision, and none of them had a history of neurological or psychiatric illness. Two volunteers were excluded because they did not complete the tests. The final sample (44 subjects; age range: 19–25 years) included 27 females with a mean age of 20.9 years (SD, 1.26) and 17 males with a mean age of 21.8 years (SD, 1.9). The study was approved by the Southwest University

Brain Imaging Center Institutional Review Board. We obtained written informed consent from all the volunteers.

Stimuli. Cue stimuli consisted of black circle and square outlines on a gray background, which were sized to be half the width of the target stimuli. Each target stimulus consisted of one superimposed semitransparent face and house picture (Fig. 1A). The face stimuli belonged to a set of eight emotionally neutral faces, four males and four females, selected from the Chinese Affective Picture System collection (Lu et al., 2005). A set of eight house stimuli, including four one-story and four two-story houses, was selected from the Internet. Each face was combined with each house to form the target stimuli, thus producing 64 unique compound stimuli. All target stimuli were gray-scaled, set to the same luminance, and presented with the same resolution (260 × 300 pixels; Fig. 1A). The contrast of the monitor was set to be constant across subjects.

Task and procedure. We employed a cued task-switching paradigm (Allport et al., 1994) where, on each trial, subjects were cued to perform one of two alternative tasks (face task vs house task). In the face task, subjects had to decide if the face feature of the compound stimulus was male or female. The right index and the left middle finger were used to indicate the gender of the face. Responses with the right and left hand were balanced such that half the participants responded to male with the right index finger and female with the left middle finger, and the other half had the opposite stimulus response mapping. In the house task, subjects had to decide whether the house feature of the compound stimulus was a one-story or a two-story building. The left index finger and the right middle finger were used to categorize the house, and right-hand and left-hand responses for the house task were also counter-balanced across subjects. Thus, the task stimuli were physically identical in both tasks as well as for repeat and switch trials.

Subjects underwent two scanning runs in the experiment, and each run consisted of 128 trials. Each target stimulus was used twice in each run in random sequence. During each run, participants were granted a 10 s rest after each block of 32 trials. Half the trials required participants to perform the face task while the other half required participants to perform the house task. Which task to perform next was presented in a pseudorandom sequence and set unpredictably from trial to trial. Specifically, there were 56 switch trials (where the cued task for the current trial differed from that on the previous trial) and 68 repeat trials (where the cued task for the current trial was the same as on the previous trial) in each run, consisting of 28 face task-set repetition trials, 35 face task-set switch trials, 28 house task-set repetition trials, and 33 house task-set switch trials. For each run, four trials (the first in each block) were neither repeat nor switch trials. As shown in Figure 1A, each trial started with a fixation cross (random jitter for 1000, 2000, or 3000 ms), followed by a cue stimulus (500 ms) indicating which task to perform. The circle cue cued the face task, and the square cue cued the house task. A blank screen (random jitter for 500, 1500, or 2500 ms) was then shown during the cue-target interval (CTI). The target stimuli were present for 2000 ms, and participants were asked to respond as soon as possible after target onset (response deadline, 2000 ms), while striving for accuracy. Before the main task, practice blocks comprising 20 trials were conducted outside the scanner. The formal task started if the accuracy rate in a practice block reached 90%.

Note that the present task protocol employed a 1:1 mapping of cue-to-task, which means that it was not designed to distinguish between costs due to cue switching versus task switching (Arrington et al., 2007). Rather, we here treat the processes of cue encoding/interpretation and the subsequent task-set reconfiguration as representing two sequential components of the broader process of interest, namely, the modification of mental rule representations that takes place when participants are prompted to change tasks.

Experimental design and statistical analysis. As detailed above, in terms of behavior, the experimental task design created two conditions of interest: task-repeat trials and task-switch trials. Behavioral effects of this factor were assessed with paired-sample *t* tests on mean response time and accuracy. Neuroimaging analyses examined both the task factor (face task vs house task) and the task transition (repeat vs switch trial) factors. As described in detail below, MVPA was used to search for brain regions whose activity could distinguish between face-task and house-task trials
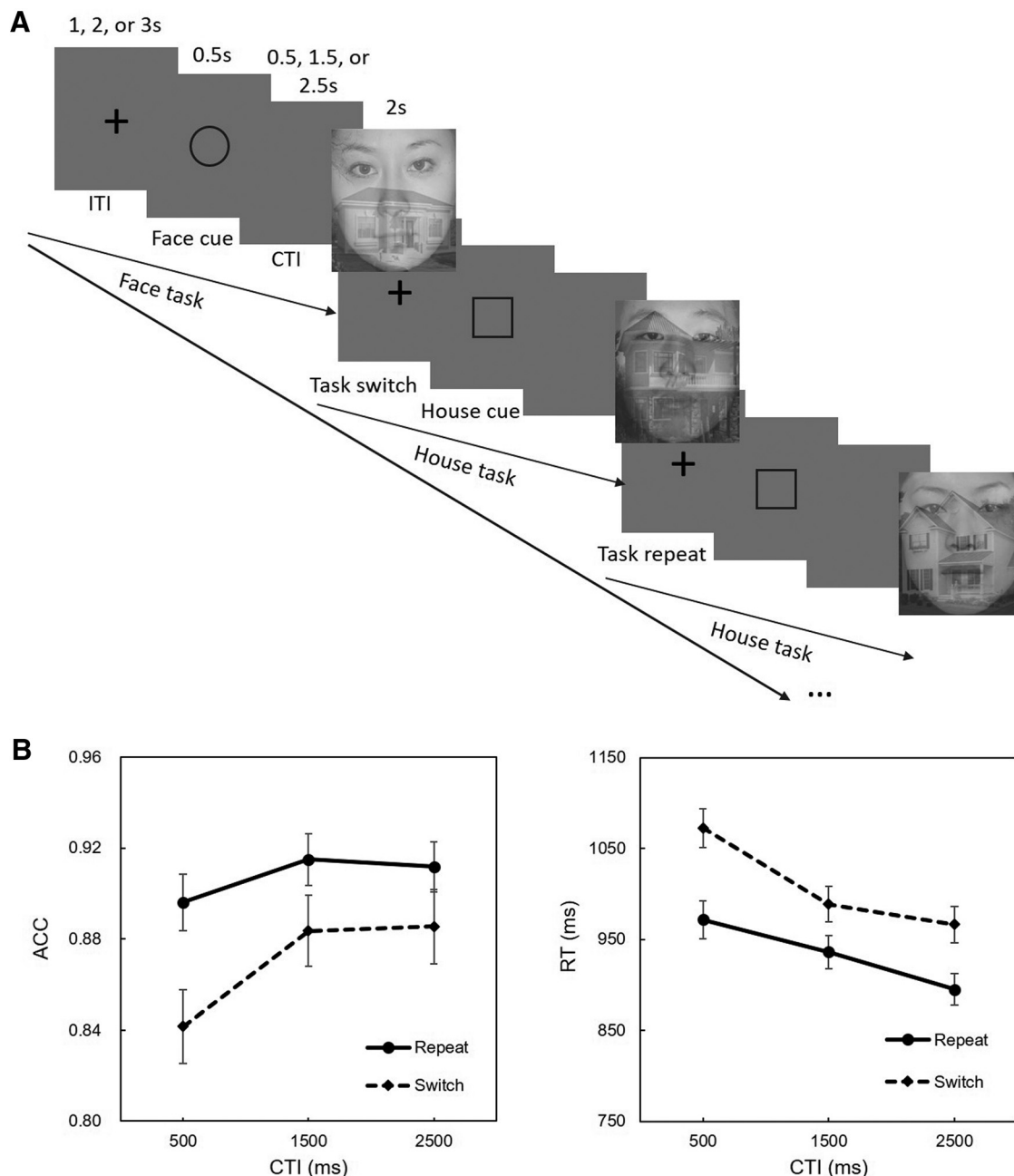
Figure 1. Experimental protocol and behavior. *A*, On each trial of a cued task-switching paradigm, subjects were cued to perform either a face task (decide whether the face shown is male or female) or a house task (decide whether the house shown is a one-story house or a two-story house) on stimuli consisting of semitransparent, overlaid face and house pictures. *B*, Mean reaction time and accuracy rates (± mean SE) are displayed as a function of trial types (switch, repeat) and CTI. ACC, Accuracy.

(at above-chance levels, corrected for multiple comparisons), and also to determine whether this classification success differed between repeat and switch trials. As also described in detail below, RSA was used to measure representational dissimilarity between different trial types from data in regions of interest (ROIs) resulting from the above MVPA. Dissimilarity was assessed between consecutive trials as a function of whether they were repeat or switch trials, and between the four possible trial categories of face–repeat, face–switch, house–repeat, and house–switch trials. All statistical tests were corrected for false-positive inflation due to multiple comparisons.

Generating predictions for neural representations. A number of fundamental, but largely untested, implications for the nature of neural task-set representations follow from the cognitive psychology literature, which we here make explicit in the form of hypotheses that can be tested with MVPA and RSA approaches. To probe the nature of neural task-set

representations, we begin by localizing the key brain regions supporting those representations. Following previous studies (Woolgar et al., 2011; Waskom et al., 2014; Wisniewski et al., 2015), we conducted an MVPA to delineate regions from which the nature of the currently active task set (face or house task) can be reliably decoded (see *Defining task-set-sensitive regions with searchlight MVPA*). Based on the common behavioral finding of better performance in repeat compared with switch trials, we derive Hypothesis 1 as follows: task-set representations on switch trials should be more labile/less stable than on repeat trials (see, *Classification accuracy and distance to the hyperplane analysis*).

Second, the basic assumption of task-set recoding within a common set of brain regions implies Hypothesis 2 as follows: task-set representations over consecutive trials should be more similar for task-repeat trials (since the same task set remains active) than for task-switch trials (where sets need to be changed; see *RSA of consecutive trials*). Moreover, based on

the fundamental assumption that the fidelity of neural task-set representations determines our ability to implement the task rules behaviorally, we derived Hypothesis 3 as follows: the relative (dis)similarity of neural task-set representation between consecutive trials should predict behavior; high similarity should benefit performance on repeat trials but impair it on switch trials.

Hypothesis 4 concerns the overall pattern of representational (dis)similarities implied by the findings of task-set inertia and reconfiguration costs. Figure 2A depicts the predicted relative pattern dissimilarity of the four possible trial types in our experiment: repeat trials with faces as the repeated relevant stimulus feature [FaceRepeat (FR)], repeat trials with houses as the repeated relevant stimulus feature [HouseRepeat (HR)], switch trials with faces as the newly relevant stimulus feature [FaceSwitch (FS)], and switch trials with houses as the newly relevant stimulus feature [HouseSwitch (HS)]. In Figure 2A, the point of origin reflects the absence of a face/house task set, and displacement to the left or right along the horizontal axis reflects the adoption of an increasingly stable/reliable representation of the face task set or house task set, respectively. The assumed dissimilarities of neural patterns are reflected in the length of horizontal arrows between conditions, where a longer line represents greater neural pattern dissimilarity.

Starting in an assumed neutral position (the point of origin), if a subject is cued to perform the face or house task, the respective task set is being configured and representations get displaced from the point of origin accordingly (Fig. 2A, First trial arrows). If on the subsequent trial, the same task is cued again (repeat trial), the task-set representation is reinforced, and thus moves farther away from the point of origin (Fig. 2A, FR and HR arrows). By contrast, if the alternative task is cued, the representation has to be reconfigured, meaning that it has to pass through the point of origin, and our model assumes (in line with the behavioral literature) that this traversing of representational space results in a less stable task-set representation (less distance from the point of origin) than on repeat trials (Fig. 2A, FS and HS arrows).

Based on these simple assumptions, the following predictions can be derived: unsurprisingly, distances between trial pairs that involve the same task are generally expected to be shorter than those between pairs involving different tasks, such that the dissimilarity between FR and HS (as well as FS and HR) trial types (Fig. 2A, iii, iv) is predicted to be greater than that between FR and FS (as well as HR and HS) conditions (Fig. 2A, i, iv). Note though that we cannot predict whether the dissimilarity between FR and FS (as well as between HR and HS) conditions (Fig. 2A, i, vi) would be smaller or greater than that between FS and HS conditions (Fig. 2A, v), because we do not know the absolute starting and end points of the FR/HR and HS/FS conditions. More interestingly though, in terms of contrasting conditions involving face versus house task sets, we can predict that FR and HR trials should have the most dissimilar representations (Fig. 2A, ii), and FS and HS trials the least dissimilar representations (Fig. 2A, v), with the distances between FS and HR conditions and between FR and HS conditions (Fig. 2A, iii, iv) falling between these extremes. In sum, if the pattern of neural representations conforms to the predictions of this model (Hypothesis 4), then the neural pattern dissimilarities should conform to the following
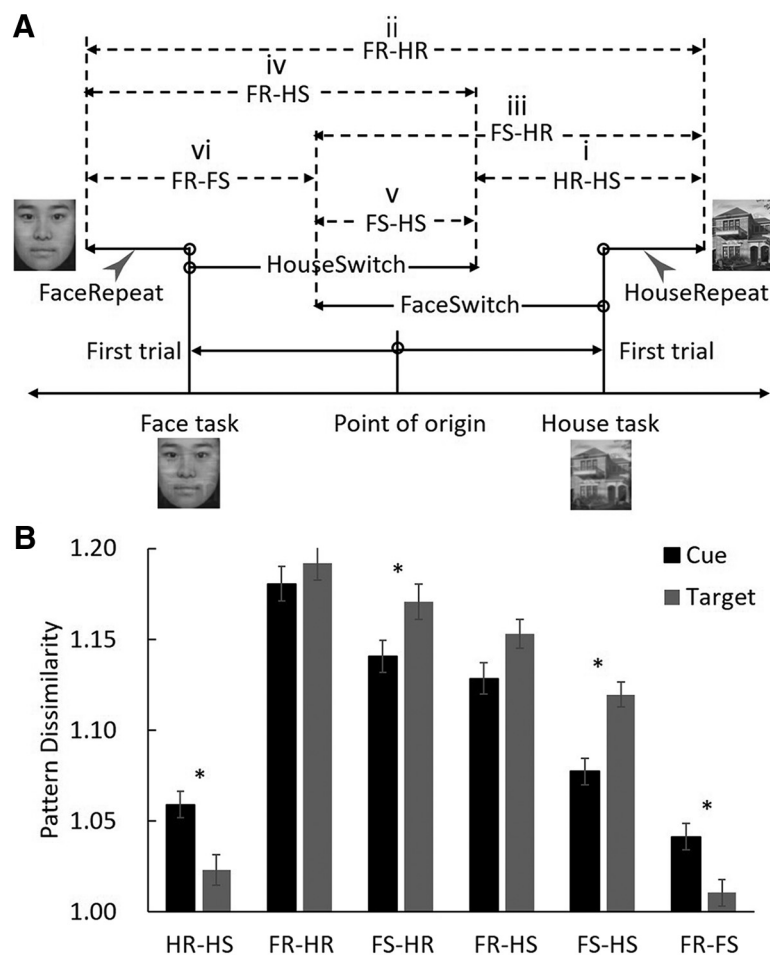

Figure 2. Predicted and observed patterns of ROI-based representational dissimilarities. *A*, A graphical depiction of the predicted pattern of neural representational dissimilarities (or distances) between the four possible trial types: FR, FS, HR, and HS. Displacement to the left or right along the horizontal axis from the point of origin reflects the adoption of an increasingly stable/reliable representation of the face task set or house task set, respectively. The assumed dissimilarities of neural patterns are reflected in the length of horizontal arrows between conditions, where a longer line represents greater neural pattern dissimilarity. Roman numerals are used to denote distances between different pairs of conditions (see *Generating predictions for neural representations* and *Using RSA to track task-set transformation*). *B*, The observed mean neural pattern dissimilarities (± mean SE), averaged across ROIs, are shown for the cue and target phase, respectively, for each condition pairing. HR–HS, house repeat– house switch distance (i); FR–HR, face repeat– house repeat distance (ii); FS–HR, face switch– house repeat distance (iii); FR–HS, face repeat– house switch distance (iv); FS–HS, face switch– house switch distance (v); FR–FS, face repeat–face switch distance (vi). *p < 0.05.

pattern: (FS–HS, HR–HS, FR–FS) < (FS–HR, FR–HS) < (FR–HR) (see, *Representational RSA of consecutive trials*).

Finally, this model sketch also allows us to make predictions about how representational distance (dissimilarity) should dynamically change in the course of cue and target processing. Specifically, based on the fundamental assumption that during switching, the neural representation moves from representing the previous trial set (task-set inertia) to being reconfigured to represent the newly relevant set, we derive Hypothesis 5 as follows: as the participant moves from encoding the task cue to the onset of the target stimulus, the dissimilarity between FS and HS should increase, while the dissimilarity between FR and FS (as well as between HR and HS) should decrease; moreover, the dissimilarity between FR and HS (as well as between FS and HR) should increase when moving from the cue to the target period (see *Using RSA to track task-set transformation*).

fMRI data collection. All fMRIs were collected on a Siemens 3T Trio scanner (Siemens Medical Systems). A foam pad was used to minimize subjects' head motion. fMRI images were acquired using gradient-echo echo planar imaging sequence (TR/TE = 1500/30 ms; flip angle, 90°; resolution matrix, 64 × 64; FOV, 192 × 192 mm^2; slice thickness, 5 mm;

acquisition voxel size, $3.0 \times 3.0 \times 5$ mm^3). A total of 24 slices were used to cover the whole brain.

Data preprocessing. fMRI data were preprocessed using Statistical Parametric Mapping 8 (Welcome Trust Centre for Neuroimaging, London, UK; http://www.fil.ion.ucl.ac.uk/spm; RRID:SCR_007037) implemented in Matlab 2009b (Mathworks; RRID:SCR_001622). For each run, the first five functional volumes of each participant served for the scanner to reach steady state and were discarded before analyses. The remaining scans were slice-time corrected and subsequently realigned to the first image to correct for head motion. The MVPA and pattern-similarity analyses were performed within native space. The time series for each voxel within a run was detrended to correct for drift, and z-score normalized to account for voxelwise signal change between runs (Pereira et al., 2009; Guidotti et al., 2015). For classification analyses, the data were not smoothed (Waskom et al., 2014; Vickery et al., 2015).

Behavioral data analysis. The mean RT and mean accuracy rate for trial types × CTIs were calculated for each subject. Extreme RTs (3 SDs above or below the mean) and incorrect responses were excluded from the RT analyses. To assess switch costs as a function of CTI, a 2 (trial type: switch vs repeat) × 3 (CTI: 500, 1500, 2500 ms) repeat-measures ANOVA was performed for the mean RTs and accuracy rates, respectively.

MVPA. MVPAs were implemented using the Princeton MVPA toolbox (http://code.google.com/p/princeton-mvpa-toolbox) in combination with a linear support vector machine (SVM) classifier from the Library for Support Vector Machines (RRID:SCR_010243; Chang and Lin, 2001). Whole-brain β maps were estimated from the first-level subject-wise GLM analyses, and a searchlight approach was then applied to these β maps (Kriegeskorte et al., 2006; Haynes et al., 2007). To obtain these β maps, we adopted a "least-squares all" approach (Mumford et al., 2012). The decoding analysis focused on the cue stage of each trial rather than on the target stage, because in this analysis we were interested in brain activity associated with cue-related task-set reconfiguration and interference resolution processes rather than task implementation and response execution. Specifically, a separate task-cue event regressor (stick function with 0 duration, convolved with SPM's hemodynamic response function) for each trial was defined in the design matrix and regressed against the neural data. Additionally, six head-motion parameters were modeled as nuisance variables. With this procedure, the neural cue response for each trial resulted in a single β map per trial. Because each trial was defined by a separate regressor in the same design matrix, the β's reflected the unique variance accounted for by each cue response. The patterns of activation (one β map for each trial) were then concatenated to construct a subject-wise activation magnitude (Rissman et al., 2004).

Defining task-set-sensitive regions with searchlight MVPA. In our initial analyses, we employed searchlight MVPA to identify brain regions that encoded task-set representations, asking from which voxel clusters in the brain we could decode the currently relevant task set (cf. Woolgar et al., 2011; Waskom et al., 2014). This analysis served to replicate previous studies and to define ROIs for our main analyses, where we used RSA to track changes in neural task-set representations across and within trials. We employed two-voxel radius searchlights (each searchlight contains ≤33 voxels) in the searchlight analysis, and trained a classifier for each subject separately. For each participant, the searchlight sphere extracted data from the non-normalized and unsmoothed β images, and the sphere moved around to include each voxel in the brain as its center voxel once. A linear SVM, with the parameter set at $c = 1$, then classified each sphere's data into face task set or house task set. The classification accuracy value of each sphere was assigned to the center voxel in the sphere. We calculated mean classification accuracies with a split-halves cross-validation procedure. Specifically, for each cross-validation, one of the two fMRI runs served as a training set, and the other run was employed as the test set. This step was repeated twice, and the two accuracies for each voxel were averaged to get the final mean accuracy map. A whole-brain gray matter mask was employed to restrict the search regions for the searchlights. Incorrect trials and trials with outlier RTs (>3 SDs or <3 SDs compared with the mean RT) were excluded from the classification analysis. Because of this procedure, the number of face-task samples and house-task samples was often unequal, which can bias the classifier. To

avoid this bias, we randomly sampled the valid face trials if the number of valid face trials was larger than that of the house trials, and vice versa. In this way, we conserved the most trials while submitting an equal number of face and house samples to the classifier. The same was done before the RSA analysis, including the RSA analysis among FR, FS, HR, and HS trials. The accuracy images for individuals were normalized by applying the warping parameters acquired at the preprocessing stage, resampled to a resolution of $3 \times 3 \times 3$ mm^3, and smoothed using an isotropic Gaussian kernel with 6 mm^3 full-width at half-maximum (FWHM). These normalized and smoothed images were then submitted to a second-level (random effects) one-sample t test to locate voxels that showed significantly higher classification accuracy than chance (0.5) at the group level. The group statistical t map was then corrected for multiple comparisons by using a cluster-level false discovery rate (FDR) at a corrected threshold of $p < 0.05$. The FDR correction was done using SPM with a cluster-forming threshold of $p < 0.001$, which is considered relatively conservative (Eklund et al. (2016). To create individual ROIs, 6-mm-radius spheres were centered on the peak voxels of clusters identified in the above analysis, resulting in 12 ROIs. These local peak-focused ROIs were employed instead of the mean activity across entire clusters because some of the clusters we observed stretched across different brain regions, which we aimed to keep separate in our analyses. We used these ROIs for all subsequent analyses.

Classification accuracy and distance to the hyperplane analysis. To test Hypothesis 1, that task sets would be more stable and distinguishable for repeat trials compared with those for switch trials, we compared the respective distances of switch-trial versus repeat-trial exemplars from the hyperplane in the above MVPA. It was expected that the distance from the hyperplane would be shorter for switch trials than for the repeat trials, and that there would be fewer correct classifications in switch trials compared with repeat trials. Specifically, in the MVPA described above, the fMRI data of the face and house task sets were projected onto a one-dimensional "hyperplane," defined by the decision boundary of the linear SVM classifier, which divides the feature space (the voxel-response space in our study) into two areas, one corresponding to the response pattern of the face task and the other one to the response pattern of the house task. The multivoxel β value of each test exemplar is located at some direction and distance from the hyperplane in one of the two areas of the space. The ability of a voxel to classify a given exemplar is reflected by the distance between the test exemplar and the SVM hyperplane: exemplars farther from the hyperplane can be more easily classified and more confidently assigned to that class than exemplars closer to the hyperplane (Frankland and Greene, 2015; Manelis and Reder, 2015; Etzel et al., 2016).

We calculated the distance between each test exemplar and the hyperplane for each subject. The distances of the misclassified exemplars were defined as negative. The classification function for each test exemplar is given by the following equation (Eq. 1):

$$y_i = sng(w^T x_i + b),$$

$$i = 1, \ldots, N, \text{ (N: number of samples)}$$

Which class the exemplar belongs to is determined by the class label y_i. If the sample comes from a class of interest, then $y_i = 1$. If the sample comes from the other class (for binary classification problem), then $y_i = -1$. $sgn(.)$ is a sign function, $sgn(x) = +1, 0, -1$ if $x > 0, x = 0$, and $x < 0$, respectively. x_i denotes the voxel response vector for the present test trial, w indicates the weight vector for each cross-validation iteration, ^{T} represents vector transposition, and b is the constant term. w and b are estimated by a linear SVM from the training dataset. The weight vector w is obtained by minimizing the following objective function (Eqs. 2–4):

$$\frac{1}{2} w^T w + C \sum_{i=1}^{N} \xi_i$$

$$\text{subject to } y_i(w^T x_i + b) \geq 1 - \xi_i$$

$$\text{and } \xi_i \geq 0$$

The slack variable ξ_i is introduced to account for nonseparable cases (i.e., the two classes cannot be separated without classification error). The slack variable is weighted by C (i.e., to the extent misclassification is allowed).

We can combine the objective function Equation 2 and the constraint Equations 3 and 4 into one nonconstraint function by introducing a Lagrange multiplier. The weighted vector w can be solved from the combined formula as follows (Eq. 5):

$$w = \sum_{i=1}^{N} \alpha_i y_i x_i$$

Where α_i is the Lagrange multiplier, which determines whether the input vector x_i is a support vector. x_i is the support vector if the corresponding α_i is nonzero.

The distance between a test exemplar and the hyperplane can then be computed as follows (Eq. 6):

$$d_i = \frac{|w^T x_i + b|}{\|w\|}$$

We calculated the mean distance and classification accuracy for switch and repeat trials, respectively (per voxel, at the individual subject level, using the searchlight method). Subsequently, these distance and accuracy rate maps were normalized, resampled to $3 \times 3 \times 3$ mm^3, and smoothed with a 6 mm FWHM kernel. The distance and accuracy rate data were then extracted for the ROIs from the maps, and several second-level paired-sample t tests were employed to explore the difference between the switch and repeat trials. To control the risk of false-positive results, multiple-comparison correction was applied to keep the FDR at <0.05. The same processing procedure pipeline for producing the statistical maps (normalization, resampling, smoothing, and ROI data extraction) was also applied for all the following analyses.

RSA of consecutive trials. To probe Hypothesis 2, the fundamental assumption that neural task-set representations remain stable (similar) in repeat trials but are altered in switch trials, we performed RSA to compare the neural pattern similarities between consecutive trials for switch versus repeat trials. RSA is used to describe representational structure and validate competing models of that structure (Kriegeskorte et al., 2008; Kriegeskorte and Kievit, 2013). RSA is complementary to classification analyses and tests whether the similarity of different stimuli or conditions can be matched with the similarity of their corresponding neural patterns. This approach provides a representational similarity matrix (e.g., a pairwise correlation matrix for brain activity from different conditions) of the neural data to calculate the similarity of different conditions (Khaligh-Razavi and Kriegeskorte, 2014; Skerry and Saxe, 2015). Here, we first computed the correlation of the β maps between consecutive trials for each voxel at the individual subject level. These correlation maps were then r-to-z transformed, and averaged across trials for task switch and repeat trials, respectively. To verify whether consecutive trial similarities would be larger in task-repeat trials compared with task-switch trials, group-level paired-sample t tests were performed for consecutive trial similarities between switch and repeat trials for each ROI. Note that high temporal autocorrelation in the BOLD signal is likely to promote a relatively high similarity of neural patterns across consecutive trials in a given brain region. However, this does not represent a confound in the current analysis, as we make no claims about absolute levels of pattern similarity but are only interested in comparing the relative dissimilarities of consecutive switch trials versus those of consecutive repeat trials. Moreover, as each trial was defined by a separate regressor in the same design matrix, the unique variance accounted for by each trial is isolated by the multiple-regression analysis.

In addition to assessing representational similarities between task-set representations of consecutive trials between repeat versus switch trials (i.e., first-order trial sequences), we also conducted an RSA on second-order task repetition (e.g., three face-set trials in a row) and switch sequences (e.g., a face-set, house-set, face-set trial sequence). This provided another test of our model's predictions. Here, our model predicts a greater task-set similarity for consecutive (second-order) repeat trials compared with a repeat trial that was preceded by a task switch (first-

order), as the successively repeated task set will become more ingrained (moving farther away from the point of origin in Fig. 2A). For consecutive (second-order) switch trials, the prediction is the same, but for a different reason: namely, that a first-order switch (a switch trial following a repeat trial) represents a transition from a well-established set (a long distance from the point of origin in Fig. 2A) to a new set, which involves a greater representational distance compared with a second-order switch (two switch trials in a row), which represents a transition between two weakly established sets (each of which lies closer to the point of origin). Thus, we predicted that representational similarities would be greater for consecutive repeat trials than repeat trials preceded by a switch trial (FR–FR/HR–HR > FS–FR/HS–HR), and greater for consecutive switch trials than for switch trials preceded by a repetition trial (FS–HS/HS–FS > FR–HS/HR–FS).

Hypothesis 3 was that the cross-trial (dis)similarity of neural patterns should be predictive of behavior, based on the assumptions that high neural pattern similarity over consecutive trials should be beneficial for repeat-trial performance but detrimental for switch-trial performance. Specifically, greater representational similarity between consecutive trials in repeat trials would indicate a better retention of the previous-trial task set, which should be associated with faster behavioral responses than in repeat trials where the cross-trial pattern similarity was low. Conversely, greater neural pattern similarities between consecutive trials in switch trials would indicate stronger interference from the previous trial or a failure to reconfigure the new task set, which should be associated with slower behavioral responses than in switch trials where the cross-trial pattern similarity was low. Therefore, we computed the correlation between the consecutive trial similarities (r-to-z-transformed correlation coefficients) and RTs of the corresponding trials. We predicted that the consecutive trial similarities and trial-by-trial RTs would be positively correlated in switch trials and negatively correlated in repeat trials. Only correct trials were employed in these analyses.

Using RSA to track task-set transformation. Hypothesis 4 concerned the predicted overall pattern of neural representational similarities over all trial types (Fig. 2A), and Hypothesis 5 relates to how this pattern should be dynamically transformed when moving from the time of cue presentation to the time of task-set implementation in the target period. To test these predictions, we conducted an additional set of RSAs assessing the relative pattern similarity of the four possible trial types (FR, FS, HR, and HS) for cue and target events, respectively. We computed the voxel response pattern in the target phase in the same way as it was done for the cue phase (see *MVPA*). Then, the β values for the cue and target phases were subjected to pattern-similarity analysis. We first labeled trials as belonging to the four trial types at the individual-subject level, and then the neural activation patterns for the cue stimuli of each voxel (using searchlights) were averaged across trials for each type of trial, respectively. For the four averaged activation patterns, pairwise correlations were calculated, resulting in six correlation coefficients for each voxel. The six correlation coefficients were submitted for r-to-z transformation. We then took the mean of the ROIs to get an "averaged ROI." Thus, there are six pattern dissimilarities (1 − pattern similarity) for each subject. Every two of the six pattern dissimilarities were paired and subjected to second-level paired-sample t tests (C_6^2), resulting in 15 paired-sample t tests for the averaged ROI. The above procedures were also done for the voxel response of the target phase, also producing 15 paired-sample t tests for the target phase. To avoid possible false-positive results, multiple-comparison correction was applied by maintaining FDR at <0.05 (Benjamini and Hochberg, 1995; Benjamini and Yekutieli, 2001).

To validate whether the pattern dissimilarities change with time, these similarities were averaged across ROIs for the cue and target phase, resulting in six dissimilarities for the cue and target phase respectively. Then, these mean pattern dissimilarities were submitted to paired-sample t tests (cue vs target phase).

Results

Behavioral results

The repeated-measures ANOVAs for RTs and accuracy rates revealed that switch trials were slower ($F_{(1,43)} = 84.8$, $p = 1 \times 10^{-11}$, $\eta_p^2 = 0.76$) and less accurate ($F_{(1,43)} = 19.7$, $p = 1 \times 10^{-5}$,
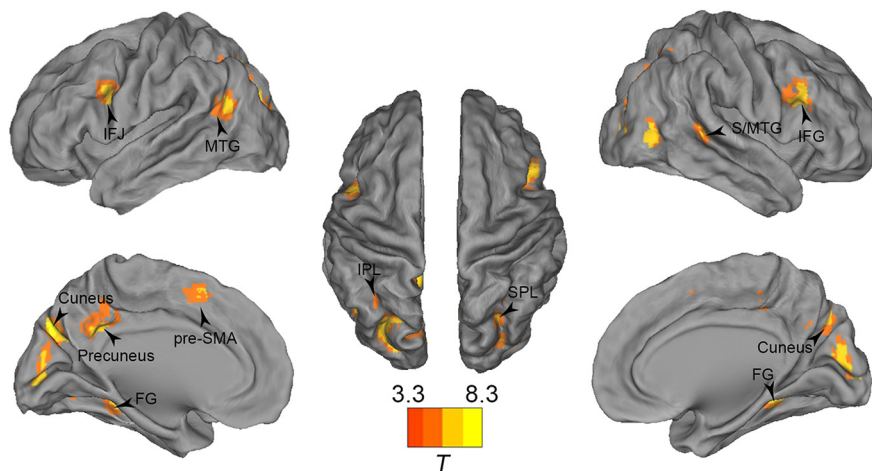
Figure 3. Brain regions encoding task-set representations. Three-dimensional surface rendering of brain regions from which whole-brain searchlight MVPA could decode the currently relevant task set at above-chance accuracy (corrected for multiple comparisons). The *t* scores, voxel numbers, and MNI coordinates are listed in Table 1. FG, Fusiform gyrus; IFJ, inferior frontal junction; IFG, right inferior frontal gyrus; IPL, inferior parietal lobule.

$\eta_p^2 = 0.35$) than repeat trials, which confirmed the classical switch-cost effect. The main effects of CTI for RTs ($F_{(1,43)} = 64.9$, $p = 3.9 \times 10^{-10}$, $\eta_p^2 = 0.53$) and accuracy rates ($F_{(1,43)} = 12.7$, $p = 0.001$, $\eta_p^2 = 0.22$) were also significant. *Post hoc* comparisons revealed that RTs decreased with increasing CTI, with slower RT at a 500 ms CTI than at 1500 ms ($t_{(43)} = 6.6$, $p = 6.2 \times 10^{-8}$, $d = 0.46$, corrected) and at 2500 ms ($t_{(43)} = 8.1$, $p = 1.2 \times 10^{-9}$, $d = 0.7$, corrected). The same was true for the comparison between the 1500 and 2500 ms CTIs ($t_{(43)} = 4.4$, $p = 7.8 \times 10^{-5}$, $d = 0.26$, corrected). Likewise, the CTIs of 1500 ms ($t_{(43)} = 4.7$, $p = 6.9 \times 10^{-5}$, $d = 0.35$, corrected) and of 2500 ms ($t_{(43)} = 3.6$, $p = 0.001$, $d = 0.34$, corrected) produced more accurate responses than that of 500 ms. The difference of accuracy rates between 1500 and 2500 ms CTIs was not significant ($t_{(43)} = 0.1$, $p = 0.9$, $d = 0.008$, corrected). Moreover, the interaction between trial type and CTI was significant for both RTs ($F_{(1,43)} = 4.7$, $p = 0.036$, $\eta_p^2 = 0.14$) and accuracy rates ($F_{(1,43)} = 4.3$, $p = 0.04$, $\eta_p^2 = 0.06$), as switch costs decreased with increasing CTI (Fig. 1B), thus replicating a classic finding of the task-switching literature. Finally, to determine to what degree task switching was affected by practice, we contrasted performance between the first and the second half of trials. There was no effect of practice on accuracy ($t_{(43)} = -0.8$, $p = 0.4$). Moreover, while mean RTs got faster over time ($t_{(43)} = 3.5$, $p = 0.001$, $d = 0.31$), this was equally the case for repeat ($t_{(43)} = 3$, $p = 0.004$, $d = 0.28$) and switch trials ($t_{(43)} = 3.3$, $p = 0.002$, $d = 0.32$), such that practice had no influence on switch costs ($t_{(43)} = 0.6$, $p = 0.5$). In sum, task practice did not appear to significantly alter the factors underlying switch costs.

Decoding neural task-set representations with MVPA
We used searchlight MVPA to identify brain regions that encoded the face and house task-set representations. We found that task set could be decoded at above-chance levels from the bilateral frontal (left inferior frontal junction, right inferior frontal gyrus) parietal [left precuneus and inferior parietal lobule, right superior parietal lobule (SPL)], and occipitotemporal cortex [bilateral middle temporal gyrus (MTG), bilateral fusiform gyrus, and bilateral cuneus], as well as the pre-SMA (Fig. 3). The T scores, cluster sizes, and MNI coordinates are listed in Table 1. These data broadly replicate data from previous studies (Woolgar et al., 2011; Waskom et al., 2014; Wisniewski et al., 2015) and

served as task-sensitive ROIs for testing our main hypotheses on dynamic neural representations of task sets.

Neural task-set representations are more distinct in repeat trials than in switch trials
To test Hypothesis 1, namely, that task-set representations on switch trials should be more labile/less stable than on repeat trials, we contrasted the hyperplane distance and classification accuracy between switch and repeat trials (see Materials and Methods). We predicted that the hyperplane distance for switch trials would be shorter than that for the repeat trials, and that there would be more incorrect classifications in switch trials compared with repeat trials. Paired-sample *t* tests revealed that the distance to the hyperplane was longer in the repeat trials compared with switch trials for all 12 ROIs (Fig. 4A; all *t*'s > 0, *p*'s < 0.05, *d*'s > 0.48; Table 2). Moreover, the same was true for task classification accuracy, which was universally greater for repeat trials than for switch trials (Fig. 4B; all *t*'s > 0, *p*'s < 0.01, *d*'s > 0.48; Table 3). Decoding of task sets was significantly above chance (*p*'s < 0.05, one-sample *t* tests) for all 12 ROIs for repeat trials, and for 8 of 12 ROIs for switch trials (excluding the left MTG, right SPL, left SMA, and right MTG). Thus, our first hypothesis was confirmed: in line with the idea that the higher error rate on switch trials compared with repeat trials reflects a less well-established task set, multivoxel patterns in task-sensitive brain regions were more predictive of the currently relevant task set in repeat trials than in switch trials.

Representational similarity over consecutive trials predicts behavior
To test Hypothesis 2, that task-set representations over consecutive trials should be more similar for repeat trials than for switch trials, we computed the mean neural pattern similarity of consecutive trials for switch and repeat trials, respectively. These correlation values were then submitted to group-level paired-sample *t* test for each ROI. We found that compared with switch trials, the pattern similarity of consecutive trials was greater in repeat trials for all 12 task-sensitive ROIs (all *t*'s > 0, *p*'s < 0.005, *d*'s > 0.26; Fig. 5; Table 4). These results confirm the fundamental assumption that neural task-set representations remain stable (similar) in repeat trials but are altered in switch trials.

To supply an additional test of our model, we also derived predictions for second-order trial sequences, namely that representational similarity of task sets for consecutive (second-order) repeat trials should be greater than for first-order repeat trials that were preceded by a task switch (FR–FR/HR–HR > FS–FR/HS–HR), as the active task set should become more ingrained over successive repetitions; and that representational similarity of the task set for consecutive (second-order) switch trials should be greater than that for first-order switch trials preceded by a task-repeat trial (FS–HS/HS–FS > FR–HS/HR–FS), as the former involve a distance from a well-established set to a new set, whereas the latter involves the relatively shorter distance between two weakly established sets. Contrasting representational similarities (averaged across the ROIs) between these trial types confirmed these predictions: representational similarity was greater for

second-order repeat trials than for first-order repeat trials ($t = 9.3$, $p = 8 \times 10^{-12}$, $d = 1.7$, corrected), and for second-order compared with first-order switch trials ($t = 20$, $p < 1 \times 10^{-16}$, $d = 4.7$, corrected).

To test Hypothesis 3, that the (dis)similarity of task-set representation over consecutive trials should be predictive of behavior, the associations between the neural pattern similarities of consecutive trials and trial-by-trial RTs were calculated for switch and repeat trials, respectively (see Materials and Methods). We predicted that greater cross-trial representational similarity in repeat trials should indicate a better retention of the previous-trial task set, and thus be associated with faster responses, whereas, conversely, greater similarity in switch trials would indicate stronger interference from the previous set or a failure to reconfigure the new task set, and thus be associated with slower responses. Results indicate that the pattern similarities of consecutive trials and trial-by-trial RTs were indeed positively correlated in switch trials for all 12 ROIs (all t's > 0, p's < 0.005, d's > 0.44, corrected for multiple comparisons; Table 5), while they were negatively correlated in repeat trials for three of the ROIs (left cuneus, $t = -2.48$, $p = 0.037$, $d = -0.37$; right cuneus, $t = -2.45$, $p = 0.037$, $d = -0.37$; right fusiform gyrus, $t = -2.56$, $p = 0.037$, $d = -0.39$, corrected), and displayed a marginally significant negative correlation in another two ROIs (left precuneus, $t = -2.18$, $p = 0.057$, $d = -0.32$; left inferior frontal junction, $t = -2.03$, $p = 0.058$, $d = -0.31$; other t's < 0, p's > 0.05, $0 < d$'s < -0.27, corrected; Table 6). Together, Hypothesis 3 was supported by the data, thus confirming the fundamental assumption that trial-by-trial behavioral flexibility is directly related to the fidelity of concomitant neural coding of task-set representations.

Finally, as with the behavioral data, we probed whether the similarities between neural task-set representations were affected by task practice. In line with the behavioral findings of unchanging switch costs, comparing the neural pattern similarities between the first and second half of the experiment did not yield any differences (p's > 0.1).

Table 1. Brain regions showing above-chance accuracy for classifying the currently active task set

Regions	Voxels (n)	T	Effect size (d)	MNI coordinates x	y	z
Left precuneus	116	4.79	0.72	−6	−51	33
Left MTG	143	3.33	0.50	−47	−48	15
Right cuneus	46	4.94	0.74	18	−72	39
Left fusiform gyrus	73	4.53	0.68	−35	−42	−13
Right SPL	65	4.86	0.73	22	−55	57
Left inferior parietal lobule	38	4.14	0.62	−30	−57	42
Right inferior frontal gyrus	107	5.37	0.81	48	21	27
Left inferior frontal junction	127	5.54	0.83	−45	8	27
Left pre-SMA	54	3.91	0.59	−3	15	57
Right fusiform gyrus	59	4.23	0.64	43	−47	−11
Right superior/middle temporal gyrus	60	4.67	0.70	48	−39	6
Left cuneus	114	7.97	1.20	−15	−78	36

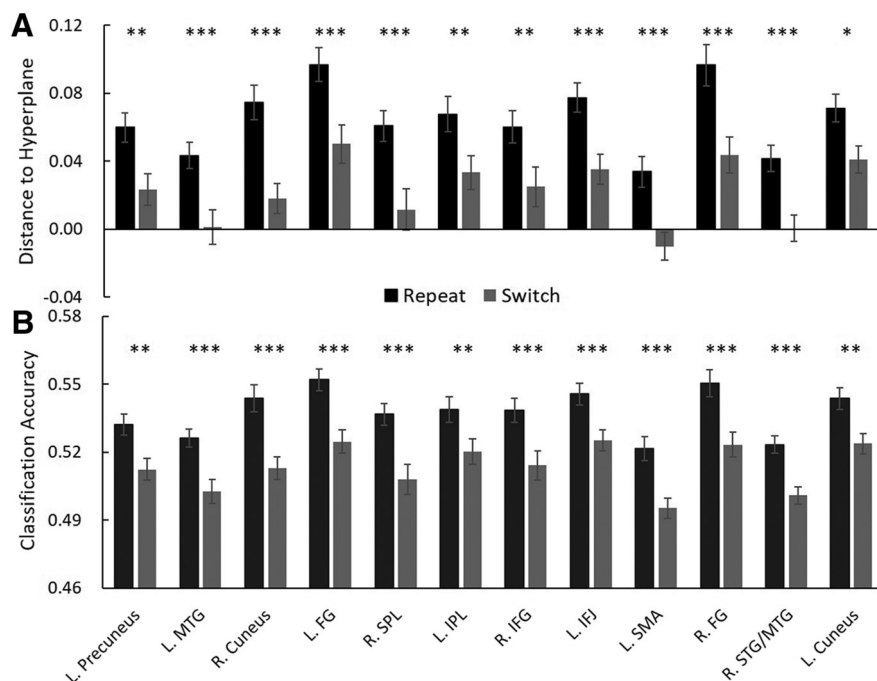Corrected at $p < 0.05$.

Figure 4. Task-set classification as a function of task-repeat versus task-switch trials. ***A***, Mean distance to the hyperplane ($\pm$ mean SE) as a function of trial type (task repeat vs switch) is displayed for each ROI. See Table 2 for specific t and p values. ***B***, Mean task-set classification accuracy as a function of trial type (task repeat vs switch) is displayed for each ROI. See Table 3 for specific t and p values. FG, Fusiform gyrus; IFG, right inferior frontal gyrus; IFJ, inferior frontal junction; L, left; R, right. *$p < .05$, **$p < 0.01$, ***$p < 0.001$.

Neural pattern similarity tracks between-trial and within-trial dynamics of task switching

In Figure 2A, we translated the basic notion that task-set inertia and reconfiguration processes lead to more distinct/stable task-set representations under repeat than under switch conditions into representational (dis)similarities, or distances between the four possible first-order task transitions (FR, FS, HR, HS) in representational space. Specifically (see *Generating predictions for neural representations*), Hypothesis 4 stated that during the cue phase, representational dissimilarities should follow the following pattern: (HS–FS, HR–HS, FR–FS) < (FS–HR, FR–HS) < (FR–HR). However, as the switch or repeat process progresses when moving from the cue to the target processing phase, we expect the pattern dissimilarities of FS–HS, FR–HS, and FS–HR

to increase and pattern dissimilarities of FR–FS and HR–HS to decrease (Hypothesis 5).

Pattern dissimilarities among the four types of trials were analyzed and compared with each other to test our model predictions. We analyze and display these similarities as averaged across ROIs for the sake of conciseness, but the patterns held for individual ROIs, too. In line with the predictions, results for the cue phase showed that the pattern dissimilarities of FS–HS, HR–HS, and FR–FS were lowest, those between FR and HR were the highest, and pattern dissimilarities between FR and HS (as well as between FS and HR) fell between these extremes (Fig. 2B). This pattern of results was supported by inferential statistics: the pattern dissimilarities of FS–HS, HR–HS, and FR–FS were significantly lower than that of FR–HS and FS–HR, and FR–HR (p's < 0.005, d's < −0.9, corrected), and the pattern dissimilarities of

Table 2. Paired-sample *t* test comparing ROI-based distance to the hyperplane between repeat and switch trials (repeat > switch)

ROI	T	P-corrected	Effect size (d)
Left precuneus	3.54	1.30×10^{-3}	0.61
Left MTG	4.34	1.84×10^{-4}	0.70
Right cuneus	4.80	1.84×10^{-4}	0.88
Left fusiform gyrus	4.33	1.84×10^{-4}	0.66
Right SPL	4.08	2.87×10^{-4}	0.68
Left inferior parietal lobule	3.04	4.44×10^{-3}	0.50
Right inferior frontal gyrus	3.31	2.25×10^{-3}	0.49
Left inferior frontal junction	4.27	1.84×10^{-4}	0.72
Left pre-SMA	4.35	1.84×10^{-4}	0.77
Right fusiform gyrus	4.31	1.84×10^{-4}	0.69
Right superior/middle temporal gyrus	4.61	1.84×10^{-4}	0.81
Left cuneus	2.66	1.10×10^{-2}	0.56

Corrected for multiple comparisons.

Table 3. Paired-sample *t* test comparing ROI-based classification accuracy between repeat and switch trials (repeat > switch)

ROI	T	P-corrected	Effect size (d)
Left precuneus	3.56	1.11×10^{-3}	0.62
Left MTG	4.59	1.09×10^{-4}	0.76
Right cuneus	4.57	1.09×10^{-4}	0.84
Left fusiform gyrus	5.25	5.35×10^{-5}	0.83
Right SPL	4.70	1.09×10^{-4}	0.74
Left inferior parietal lobule	2.94	5.71×10^{-3}	0.49
Right inferior frontal gyrus	3.94	4.39×10^{-4}	0.62
Left inferior frontal junction	3.85	5.19×10^{-4}	0.65
Left pre-SMA	4.54	1.09×10^{-4}	0.80
Right fusiform gyrus	4.06	3.52×10^{-4}	0.71
Right superior/middle temporal gyrus	4.48	1.10×10^{-4}	0.87
Left cuneus	2.79	7.78×10^{-3}	0.63

Corrected for multiple comparisons.

FR–HS and FS–HR were significantly lower than that of FR–HR (both *p*'s < 0.01, *d*'s < −0.5, corrected). Thus, the data provided strong support for the set of representational similarity predictions grouped under Hypothesis 4.

To test Hypothesis 5, that these representational dissimilarities should dynamically change when moving from the time of cue presentation to the time of task-set implementation in the target period, we compared cue-related versus target-related dissimilarities for the different trial pairs. Paired-sample *t* tests revealed that pattern dissimilarity of FR–FS ($t_{(43)} = -2.53, p = 0.031, d = -0.44$, corrected) and HR–HS ($t_{(43)} = -2.96, p = 0.015, d = -0.44$, corrected) conditions decreased, whereas pattern dissimilarity of FS–HS ($t_{(43)} = 3.22, p = 0.015, d = 0.54$, corrected) as well as FS–HR ($t_{(43)} = 2.32, p = 0.038, d = 0.36$, corrected) conditions increased from cue to target phase. The pattern dissimilarity of FR–HS also displayed a trend-level increase ($t_{(43)} = 1.93, p = 0.073, d = 0.33$, corrected). These results confirm Hypothesis 5, showing that during switching, the neural task-set representation moves in representational space, away from representing the previous-trial set (due to task-set inertia) and toward representing the reconfigured, newly relevant set.

Control analyses
Representational similarity
It is noteworthy that the pattern dissimilarities (1 − pattern similarity) we observed were all >1 (Fig. 2B), indicating that the neural patterns were negatively correlated. While this is intuitively plausible for the condition pairs where distinct task sets are being correlated, it is surprising that representational patterns for

the same task sets that only differ by their task-repeat versus task-switch trial status (i.e., HR–HS and FR–FS) should also be slightly negatively correlated. To ensure that these results are truly attributable to representational differences between repeat and switch trials, we calculated the pattern similarities between pairs of trials within each trial type (i.e., FR, FS, HR, and HS) and then averaged these pattern-similarity values across trial pairs. Here, pattern correlations should be expected to be positive. In line with expectations, the within-trial-type similarities were in fact all highly positive (FR, $t = 38.8, p < 1 \times 10^{-16}, d = 5.9$; FS, $t = 37.9, p < 1 \times 10^{-16}, d = 5.7$; HR, $t = 49.3, p < 1 \times 10^{-16}, d = 7.4$; HS, $t = 41.4, p < 1 \times 10^{-16}, d = 6.2$). By contrast, when we applied the same analysis to trial types that either involved a different task set or a different trial transition (e.g., face repeat vs face switch), the correlations were all negative (HR–HS: $t = -6.9, p = 2.2 \times 10^{-8}, d = -1$; FR–HR: $t = -19.1, p < 1 \times 10^{-16}, d = -2.9$; FS–HR: $t = -14.3, p < 1 \times 10^{-16}, d = -2.2$; FR–HS: $t = -12.1, p = 3.6 \times 10^{-15}, d = -1.8$; FS–HS: $t = -9.2, p = 1.6 \times 10^{-11}, d = -1.4$; FR–FS: $t = -2.5, p = 0.02, d = -3.7$), though—importantly—the pattern of these dissimilarities again conformed to the predictions of our model: (HS–FS, HR–HS, FR–FS) < (FS–HR, FR–HS) < (FR–HR).

Given the results of these control analyses, we conclude that the counterintuitive findings of negative correlations between FR and FS trial types and between HR and HS trial types are genuine. The fact that we observe negative correlation values for identical task sets as a function of whether these sets were repeated or just switched to likely reflects a substantial influence of task-set inertia: it appears that on switch trials, the prior set still impacts the task representation to such a degree that it renders the switch-trial task-set representation quite unlike the repeat-trial task-set representation. This idea is supported by the fact that the degree of negative correlation becomes smaller for these trial types when moving from the cue to the target phase (Fig. 2B, HR–HS, FR–FS). Moreover, it is possible that dissimilarity between these trial types is further enhanced by differences in the cognitive processes active during repeat trials (where an ongoing set needs to be maintained) and switch trials (where a prior set needs to be suppressed and a new set established). Future studies addressing this issue in designs involving >2 task types would be highly desirable.

Temporal proximity
Due to the sluggish nature of the BOLD response, consecutive task events closely spaced in time will produce BOLD responses with temporal overlap, which can lead to misattribution of activation between the events due to a "bleed-over" of the response evoked by the first event into the epoch of the second event. Given that the main analyses in the present paper focus on successive trial events, we took the following measures to rule out the possibility that our results could be driven by this kind of bleed-over. First, the design presented the four different trial types in random order, and included randomly jittered intertrial and cue-to-target intervals, both of which facilitate the statistical separation of activation evoked by the different events (Wager and Nichols, 2003). Second, each trial was modeled with a separate regressor, thus imposing orthogonalization of parameter estimates between trials (including consecutive ones). However, the results reported above were all derived from design matrices that either exclusively modeled the cue events (testing Hypotheses 1–4) or the target events (in testing Hypothesis 5). To ensure that these results are not based on misattributing cue-related variance to target-related variance, and vice versa, we reran all of
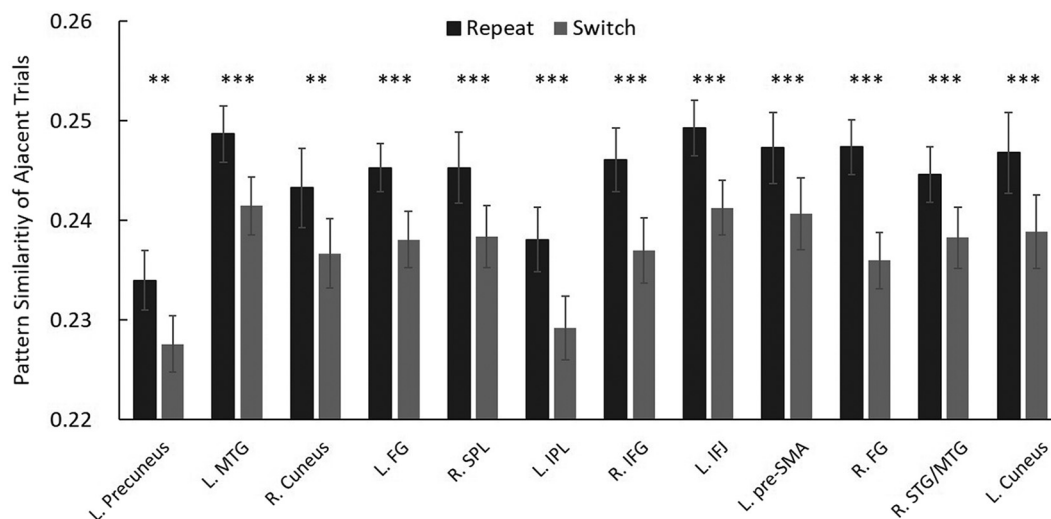
Figure 5. Mean neural pattern similarity across consecutive trials. Bars display the neural pattern correlations ($\pm$ mean SE) between consecutive trials as a function of whether they were repeat or switch trials for each ROI. See Table 4 for specific t and p values quantifying these pattern dissimilarities with group-level paired-sample t tests. **$p < 0.01$, ***$p < 0.001$. FG, Fusiform gyrus; IPL, inferior parietal lobule; IFG, right inferior frontal gyrus; IFJ, inferior frontal junction; L, left; R, right.

Table 4. Paired-sample t test comparing ROI-based adjacent-trial pattern similarities between repeat and switch trials (repeat > switch)

ROI	T	P-corrected	Effect size (d)
Left precuneus	3.10	1.71×10^{-3}	0.32
Left MTG	4.19	1.63×10^{-4}	0.37
Right cuneus	3.16	1.56×10^{-3}	0.27
Left fusiform gyrus	3.82	3.67×10^{-4}	0.40
Right SPL	3.47	7.90×10^{-4}	0.31
Left inferior parietal lobule	4.66	4.52×10^{-5}	0.41
Right inferior frontal gyrus	5.01	2.98×10^{-5}	0.42
Left inferior frontal junction	4.79	4.04×10^{-5}	0.43
Left pre-SMA	3.43	7.98×10^{-4}	0.27
Right fusiform gyrus	7.18	4.19×10^{-8}	0.61
Right superior/middle temporal gyrus	3.96	2.77×10^{-4}	0.32
Left cuneus	3.76	3.82×10^{-4}	0.30

Corrected for multiple comparisons.

Table 6. Group level one-sample t test on the ROI-based correlation between consecutive trial similarities of the repeat trials and trial-by-trial RTs

ROI	T	P-corrected	Effect size (d)
Left precuneus	-2.181	0.057	-0.322
Left MTG	-1.731	0.088	-0.261
Right cuneus	-2.452	0.037	-0.370
Left fusiform gyrus	-1.426	0.097	-0.215
Right SPL	-1.568	0.093	-0.236
Left inferior parietal lobule	-1.667	0.088	-0.251
Right inferior frontal gyrus	-1.503	0.093	-0.227
Left inferior frontal junction	-2.031	0.058	-0.306
Left pre-SMA	-0.955	0.172	-0.144
Right fusiform gyrus	-2.556	0.037	-0.385
Right superior/middle temporal gyrus	-1.137	0.143	-0.171
Left cuneus	-2.480	0.037	-0.374

Corrected for multiple comparisons.

Table 5. Group level one-sample t test on the ROI-based correlation between adjacent trial similarities of the switch trials and trial-by-trial RTs

ROI	T	P-corrected	Effect size (d)
Left precuneus	4.56	1.05×10^{-4}	0.32
Left MTG	3.55	5.76×10^{-4}	0.37
Right cuneus	4.01	2.04×10^{-4}	0.27
Left fusiform gyrus	4.12	1.72×10^{-4}	0.40
Right SPL	3.81	2.94×10^{-4}	0.31
Left inferior parietal lobule	3.17	1.52×10^{-3}	0.41
Right inferior frontal gyrus	4.34	1.09×10^{-4}	0.42
Left inferior frontal junction	5.00	6.14×10^{-5}	0.43
Left pre-SMA	2.98	2.39×10^{-3}	0.27
Right fusiform gyrus	4.32	1.09×10^{-4}	0.61
Right superior/middle temporal gyrus	4.49	1.05×10^{-4}	0.32
Left cuneus	3.90	2.51×10^{-4}	0.30

Corrected for multiple comparisons.

the analyses, but now defining the target onsets as additional regressors when computing the β images of the cue phase, and adding the cue onsets as regressors when computing the β images of the target phase.

We found that all the main results reported above were preserved under these new models. First, the 12 ROIs displayed above-chance classification accuracy in distinguishing the two

task sets (all t's > 0, p's < 0.05, d's > 0.34, with FDR correction). Second, the distance to hyperplane and the classification accuracy were greater for repeat trials than switch trials (all t's > 0, p's < 0.05, d's > 0.34, corrected). Third, the pattern similarities of consecutive trials were also greater in repeat trials than in switch trials (all t's > 0, p's < 1×10^{-7}, d's > 0.6, corrected). Moreover, representational similarity was greater for second-order repeat trials than for first-order repeat trials ($t_{(43)} = 9.1$, $p = 1.6 \times 10^{-11}$, $d = 1.6$, corrected), and also greater for second-order switch trials than for first-order switch trials ($t_{(43)} = 20$, $p < 1 \times 10^{-16}$, $d = 4.4$, corrected). Fourth, greater consecutive trial similarity was beneficial for repeat-trial performance (t's < 0, p's < 0.05 in several ROIs, corrected), but detrimental for switch-trial performance (all t's > 0, p's < 0.05, d's > 0.53, corrected). Finally, the RSA pattern corresponding to our model predictions for the cue and target phases was also preserved (data not shown). Thus, the current results are unlikely to be related to misattribution of BOLD responses between trials or between cue and target events. In addition to these control analysis results, it should also be noted that the pattern-similarity results themselves speak against the possibility that cue-driven and target-driven BOLD responses are confounded, as some pattern similarities increased while others decreased from cue to target phases.
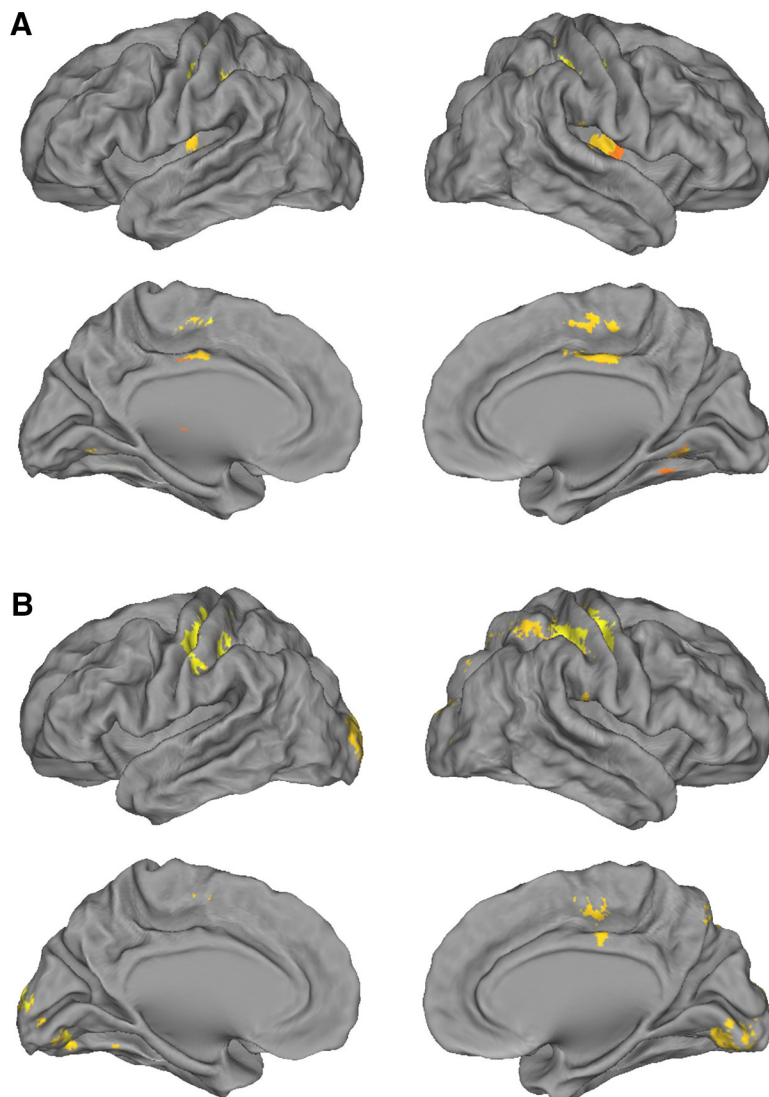
Figure 6. Brain regions encoding stimulus feature representations. **A**, 3D surface rendering of brain regions from which whole-brain searchlight MVPA could decode whether the face feature of a current stimulus was male or female at above-chance accuracy (corrected for multiple comparisons). **B**, 3D surface rendering of brain regions from which whole-brain searchlight MVPA could decode whether the house feature of a current stimulus was a one-story or a two-story house at above-chance accuracy (corrected for multiple comparisons).

Finally, another factor that might affect processing during switch trials is that a repeated target stimulus could reactivate the previous-trial task set that had just been applied to that stimulus. In the present study, however, this would be an extremely rare occurrence, as each target was shown only twice within each run of 128 randomly ordered trials, and task-set reactivation through target repetition is therefore highly unlikely to be a significant contributor to the present findings.

Task versus stimulus representations
In another set of control analyses, we sought to confirm that our results do in fact selectively reflect the neural representation of task sets, independent of neural sensitivities to specific physical stimulus features (male or female face; one-story or two-story house). While the current task design renders the task and switch factors independent of particular stimulus features, it is nevertheless possible that the brain regions we identified as supporting task-set coding may also have preferential responses to different stimulus features. Moreover, demonstrating that the results we

report above are completely independent of repetitions or changes of stimulus features over consecutive trials would provide strong additional evidence against the possibility that these results are driven by sluggish BOLD responses being misattributed across trials.

To ensure that our results reflect task-set rather than stimulus representations, we first submitted the stimulus features to an MVPA analysis. Here, male versus female faces and one-story versus two-story buildings were used as training and testing labels. Since the stimulus was not determined until the presentation of the target, we employed the β values of the target phase in this MVPA analysis. The other procedures were identical to the task-set analysis. Results showed that the stimulus features could be decoded from clusters that peaked in visual and motor regions (Fig. 6), but had partial overlap with some of task-sensitive ROIs (including in cuneus, right SPL, and pre-SMA). Next, we tested whether any of the RSA results we observed for task-set processing in the 12 task-sensitive ROIs could be replicated at the level of stimulus features. To ensure the orthogonality between stimulus features and task sets in these analyses, we excluded trials where the repetition or switching of task sets and stimulus types were confounded. The other procedures were identical to the pattern-similarity analyses at the task level. The pattern-similarity effects over consecutive trials was not found in the stimulus repeat/switch analysis; that is, pattern similarities of consecutive trials for repeated stimulus features were no greater than for switched stimulus features in the 12 ROIs (all p's > 0.1). Moreover, the classification accuracy and distance to the hyperplane were no greater for stimulus feature repeat trials compared with switch trials in the 12 ROIs

(all p's > 0.1). Finally, the trials were also grouped into four types according to the stimulus feature repetitions (male–repeat, male–switch, female–repeat, female–switch, as well as one-story–repeat, one-story–switch, two-story–repeat, two-story–switch). The pattern similarities between these trial types were calculated and contrasted in reference to the predictions and findings reported in Figure 2. The results did not replicate the pattern of representational dissimilarities found in the task-set analysis (all p's > 0.1).

Finally, it should be noted that by using a 1:1 cue-to-task mapping, the current task design is in principle open to the argument that the multivoxel patterns that allowed us to distinguish face-task trials from house-task trials may not actually reflect the currently active task set (face task/house task) or the meaning of the cues (face task/house task), but rather that they reflect the mere physical features of the cues (circle/square). While this is a logical possibility, this scenario appears to be unlikely for most of our ROIs, in particular the frontoparietal regions of prime inter-

est. First, the frontoparietal cortex is not known to be inherently sensitive to low-level visual shapes like squares and circles, so it is unlikely that we would be able to decode these perceptual characteristics from these regions unless the cues were processed as signifying their respective associated task sets. Second, much prior work has shown instead that the frontoparietal cortex is involved in representing task sets (Woolgar et al., 2011, Waskom et al., 2014), even when those are explicitly dissociated from perceptual cue feature encoding (Etzel et al., 2016). Third, much prior literature has shown that low-level shape features of the kind we employed as cues are instead encoded in the visual cortex, most prominently the lateral occipital complex (for review, see Grill-Spector and Malach, 2004). It therefore seems unlikely that our results, especially those regarding neural representations in the frontoparietal cortex, would reflect the decoding of low-level visual features rather than of task-set information. However, the possibility that distinct visual cue features may contribute to our results cannot be ruled out.

In sum, our control analyses strongly suggest that the results we obtained in testing Hypotheses 1–5 all pertained selectively to the encoding of task-set representations, which were unlikely to be confounded by sensitivities to specific stimulus categories or physical cue features (though the latter cannot be logically excluded). Moreover, the results are unlikely to be contaminated by bleed-over of BOLD responses across trial phases and trials.

Discussion

Cognitive flexibility lies at the core of the extraordinary human ability to adapt. Many univariate neuroimaging studies of task switching have previously attributed the process of updating task-set representations to frontal and parietal cortices, and a small number of recent fMRI studies employing MVPA have shown that the currently relevant task set can in fact be decoded from activity patterns in these regions (Woolgar et al., 2011; Waskom et al., 2014; Wisniewski et al., 2015). Prior studies have not probed whether these (or other) brain regions can be shown to support the dynamic transformation of task sets from trial to trial, however. In the present study, we therefore used RSA to examine a set of fundamental assumptions about neural task-set representations in relation to the switching process. We found that frontoparietal and visual stimulus processing regions that harbored information about the currently relevant set (Fig. 3) displayed the following characteristics: (1) the representation of task sets on switch trials was weaker than on repeat trials (Fig. 5); (2) the neural pattern dissimilarity across consecutive trials was greater for switch trials than for repeat trials (Fig. 4); (3) the degree of this pattern dissimilarity predicted behavior (Tables 3, 4); (4) the overall neural pattern of representational dissimilarities between different trial types (FR, FS, HR, HS) followed predictions derived from the simple assumption that repeating sets results in stronger neural task representations than switching sets (Fig. 2); and (5) when moving from cue to target phase within a trial, the pattern dissimilarities tracked the transformation from representations corresponding to the previously relevant set to the currently relevant set. Overall, these results provide neural support for the longstanding assumptions of an effortful (and imperfect) task-set reconfiguration process (Rogers and Monsell, 1995) that is hampered by task-set inertia (Allport et al., 1994), and they demonstrate that frontoparietal and stimulus processing regions provide the neural substrates of "dynamic adaptive coding" (Duncan, 2013; Stokes et al., 2013), flexibly representing changing task sets in a trial-by-trial fashion.

The results of our MVPA identification of brain regions involved in encoding the currently relevant task set closely replicate previous findings (Woolgar et al., 2011; Waskom et al., 2014; Garner and Dux, 2015; Wisniewski et al., 2015; Etzel et al., 2016) and thus confirm that a frontoparietal network, in collaboration with (here: visual) stimulus processing regions, supports task-set representations (Fig. 3). Moreover, in a novel test of the stability of these representations, we also documented that neural task-set encoding is more labile/less stable on switch trials than on repeat trials (Fig. 5). This finding is perfectly congruent with much behavioral literature that has documented time and again that performance is more error prone on switch trials than on repeat trials (for review, see Monsell, 2003; Kiesel et al., 2010; Vandierendonck et al., 2010). However, another recent study has come to the opposite conclusion, namely, that frontoparietal task-set representations are unaffected by switch processes (Loose et al., 2017). A probable reason for these divergent findings is that, unlike the current design, the task of Loose and colleagues did not require subjects to switch between different cognitive tasks but only between varying stimulus–response mappings. While there is some evidence to suggest that at least the frontal component of the frontoparietal network encodes task goals rather than stimulus–response mappings (Muhle-Karbe et al., 2014), a future MVPA study that independently varies switches in task goal versus response mapping should test directly whether switching differentially affects these two types of representations in the frontoparietal cortex.

At first glance, the present results may also seem at odds with previous MVPA studies reporting that increased perceptual difficulty (Woolgar et al., 2011) and increased demand for cognitive control (Waskom et al., 2014) are associated with higher decoding accuracy compared with less demanding conditions. However, these results are only superficially incongruous, and can easily be accounted for by key differences in the nature of the task protocols and analyses. First, Woolgar and colleagues (2011) manipulated the perceptual difficulty in identifying a task-relevant stimulus feature (stimulus position) in a block-wise fashion and found that representation of that stimulus feature in the frontoparietal cortex was enhanced under greater perceptual difficulty. By contrast, the strength of task-rule representations was not actually affected by this manipulation, and the authors did not contrast task-set decoding success as a function of repeat versus switch trials (Woolgar et al., 2011). Thus, the present results are not in conflict with these prior findings. Second, Waskom and colleagues (2014) employed a protocol where a task cue, followed by a long (6.5 s) cue-target interval, indicated which task to perform in a subsequent miniblock of three consecutive trials. Within these miniblocks, the authors found that task-rule classification accuracy decreased over repeated trials (Waskom et al., 2014). While this finding differs from the present demonstration of stronger task-set encoding on task-repetition trials, the particulars of the Waskom et al. (2014) design provide a straightforward explanation for this divergence: unlike in our task, task switches in that study could occur only on a predictable subset of trials. Given that subjects knew that only the end of a miniblock could be followed by a task switch, and were then granted a 6.5 s cued preparation period for the subsequent task, it is plausible that task-set implementation would be stronger at the beginning of the new miniblock and then move to a less controlled, "autopilot" setting for the predictable subsequent task-repeat trials (Waskom et al., 2014). By contrast, the current results show that when task switches are unpredictable and preparation time is brief, frontoparietal task-set encoding is weaker on switch trials

than on repeat trials, in line with classic behavioral findings of poorer accuracy on switch trials.

To test our main hypotheses, we pursued a novel application of RSA to assessing the neural pattern dissimilarities between consecutive trials as a function of whether the task set was repeated or switched. Our findings support the fundamental assumption that task-coding regions track the currently relevant set from trial to trial, such that task-set repetitions are associated with a shorter distance (or dissimilarity) in neural representational space than task-set switches. Moreover, we showed that this neural measure of (dis)similarity of cognitive representations from one trial to the next predicts behavior: in line with the assumption that greater similarity should be beneficial for repeat trials but detrimental for switch trials, RTs were positively correlated with the degree of cross-trial representational similarity for task-switch trials and negatively for task-repeat trials. To our knowledge this is the first demonstration of a direct link between dynamic neural task-set recoding across trials and behavior, a finding that clearly supports the notion that the fidelity of neural task-set representations determines one's ability to implement task rules to guide response selection. Moreover, this set of results also offers a novel neurobehavioral demonstration of the oft-hypothesized stability–flexibility trade-off (Goschke, 2000; Cohen et al., 2007), the notion that a strong encoding of an ongoing task set (i.e., cognitive stability) to ensure accurate performance comes at the cost of a diminished ease at changing sets, and vice versa, that a more labile task set is more error-prone but also more easily updated (i.e., promoting cognitive flexibility).

We next tested a set of RSA predictions on the relative distances between the different trial types produced by our experimental design (Fig. 2). The simple assumption that repeating sets results in stronger neural task representations than switching sets predicted that FR and HR trials would exhibit the most dissimilar representations and FS and HS trials the least dissimilar ones, with the distances between FS and HR conditions and between FR and HS conditions falling between these extremes. This similarity pattern in representational space was clearly confirmed by our results (Fig. 2). Moreover, we also employed RSA to test predictions about how representational distances should dynamically change when moving from cue to target processing. We found that compared with cue distances, the dissimilarity between FS and HS increased, while the dissimilarity between FR and FS (as well as between HR and HS) decreased during target processing. Similarly, the dissimilarity between FR and HS (as well as between FS and HR) increased for the target relative to the cue period. All of these results are perfectly in line with the fundamental assumption that during switching, the frontoparietal cortex moves from representing the previous-trial set (due to task-set inertia) to being reconfigured to represent the newly relevant set.

In conclusion, we combined a cued task-switching protocol with fMRI and harnessed RSA to facilitate a novel assessment of trial-by-trial changes in neural task-set representations. Task-sensitive frontoparietal and visual stimulus processing regions were found to display weaker task-set representation on switch trials than on repeat trials, and greater neural pattern dissimilarity across consecutive trials for switch trials than for repeat trials, which was predictive of behavior. The overall pattern of representational dissimilarities followed the assumption that repeating sets results in stronger neural task representations than switching sets, and when moving from cue to target phase within a trial, the pattern dissimilarities tracked task-set transformation. These results provide strong support for the involvement of frontoparie-

tal and stimulus processing regions in dynamic adaptive coding of task sets from trial to trial.

References

Allport DA, Styles EA, Hsieh S (1994) Shifting intentional set: exploring the dynamic control of tasks. Attention Perform 15:421–452.

Arrington CM, Logan GD, Schneider DW (2007) Separating cue encoding from target processing in the explicit task-cuing procedure: are there "true" task switch effects? J Exp Psychol Learn Mem Cogn 33:484–502. CrossRef Medline

Benjamini Y, Hochberg Y (1995) Controlling the false discovery rate: a practical and powerful approach to multiple testing. J R Statist Soc B (Methodological) 57:289–300.

Benjamini Y, Yekutieli D (2001) The control of the false discovery rate in multiple testing under dependency. Ann Stat 29:1165–1188. CrossRef

Braem S, King JA, Korb FM, Krebs RM, Notebaert W, Egner T (2013) Affective modulation of cognitive control is determined by performance-contingency and mediated by ventromedial prefrontal and cingulate cortex. J Neurosci 33:16961–16970. CrossRef Medline

Brass M, von Cramon DY (2002) The role of the frontal cortex in task preparation. Cereb Cortex 12:908–914. CrossRef Medline

Brass M, von Cramon DY (2004) Decomposing components of task preparation with functional magnetic resonance imaging. J Cogn Neurosci 16:609–620. CrossRef Medline

Chang CC, Lin CJ (2001) Training ν-support vector classifiers: theory and algorithms. Neural Comput 13:2119–2147. CrossRef Medline

Cohen JD, McClure SM, Yu AJ (2007) Should I stay or should I go? How the human brain manages the trade-off between exploitation and exploration. Philo Trans R Soc Lond B Biol Sci 362:933–942. CrossRef Medline

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. CrossRef Medline

Duncan J (2013) The structure of cognition: attentional episodes in mind and brain. Neuron 80:35–50. CrossRef Medline

Eklund A, Nichols TE, Knutsson H (2016) Cluster failure: why fMRI inferences for spatial extent have inflated false-positive rates. Proc Natl Acad Sci U S A 113:7900–7905. CrossRef Medline

Etzel JA, Cole MW, Zacks JM, Kay KN, Braver TS (2016) Reward motivation enhances task coding in frontoparietal cortex. Cereb Cortex 26:1647–1659. CrossRef Medline

Frankland SM, Greene JD (2015) An architecture for encoding sentence meaning in left mid-superior temporal cortex. Proc Natl Acad Sci U S A 112:11732–11737. CrossRef Medline

Garner KG, Dux PE (2015) Training conquers multitasking costs by dividing task representations in the frontoparietal-subcortical system. Proc Natl Acad Sci U S A 112:14372–14377. CrossRef Medline

Goschke T (2000) Intentional reconfiguration and involuntary persistence in task-set switching. In: Attention and performance, Volume 18 (Monsell S, Driver TS, eds), pp 331–355. Cambridge, MA: MIT.

Grill-Spector K, Malach R (2004) The human visual cortex. Annu Rev Neurosci 27:649–677. CrossRef Medline

Guidotti R, Del Gratta C, Baldassarre A, Romani GL, Corbetta M (2015) Visual learning induces changes in resting-state fMRI multivariate pattern of information. J Neurosci 35:9786–9798. CrossRef Medline

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

Haynes JD, Sakai K, Rees G, Gilbert S, Frith C, Passingham RE (2007) Reading hidden intentions in the human brain. Curr Biol 17:323–328. CrossRef Medline

Johnston K, Levin HM, Koval MJ, Everling S (2007) Top-down control-signal dynamics in anterior cingulate and prefrontal cortex neurons following task switching. Neuron 53:453–462. CrossRef Medline

Khaligh-Razavi SM, Kriegeskorte N (2014) Deep supervised, but not unsupervised, models may explain IT cortical representation. PLoS Comput Biol 10:e1003915. CrossRef Medline

Kiesel A, Steinhauser M, Wendt M, Falkenstein M, Jost K, Philipp AM, Koch I (2010) Control and interference in task switching—a review. Psychol Bull 136:849–874. CrossRef Medline

Korb FM, Jiang J, King JA, Egner T (2017) Hierarchically organized medial frontal cortex-basal ganglia loops selectively control task- and response-selection. J Neurosci 37:7893–7905. CrossRef Medline

Kriegeskorte N, Kievit RA (2013) Representational geometry: integrating

cognition, computation, and the brain. Trends Cogn Sci 17:401–412. CrossRef Medline

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

Kriegeskorte N, Mur M, Bandettini P (2008) Representational similarity analysis-connecting the branches of systems neuroscience. Front Syst Neurosci 2:4. CrossRef Medline

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. CrossRef Medline

Lu B, Hui M, Yuxia H (2005) The development of native Chinese affective picture system—a pretest in 46 college students. Chinese Mental Health J 19:719–722.

Manelis A, Reder LM (2015) He who is well prepared has half won the battle: an fMRI Study of task preparation. Cereb Cortex 25:726–735. CrossRef Medline

Meiran N (1996) Reconfiguration of processing mode prior to task performance. J Exp Psychol Learn Mem Cogn 22:1423. CrossRef

Meiran N (2000) Modeling cognitive control in task-switching. Psychol Res 63:234–249. CrossRef Medline

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

Monsell S (2003) Task switching. Trends Cogn Sci 7:134–140. CrossRef Medline

Muhle-Karbe PS, Andres M, Brass M (2014) Transcranial magnetic stimulation dissociates prefrontal and parietal contributions to task preparation. J Neurosci 34:12481–12489. CrossRef Medline

Mumford JA, Turner BO, Ashby FG, Poldrack RA (2012) Deconvolving BOLD activation in event-related designs for multivoxel pattern classification analyses. Neuroimage 59:2636–2643. CrossRef Medline

Pereira F, Mitchell T, Botvinick M (2009) Machine learning classifiers and fMRI: a tutorial overview. Neuroimage 45:S199–S209. CrossRef Medline

Rissman J, Gazzaley A, D'Esposito M (2004) Measuring functional connectivity during distinct stages of a cognitive task. Neuroimage 23:752–763. CrossRef Medline

Rogers RD, Monsell S (1995) Costs of a predictable switch between simple cognitive tasks. J Exp Psychol 124:207. CrossRef

Ruge H, Jamadar S, Zimmermann U, Karayanidis F (2013) The many faces of preparatory control in task switching: reviewing a decade of fMRI research. Hum Brain Mapp 34:12–35. CrossRef Medline

Skerry AE, Saxe R (2015) Neural representations of emotion are organized around abstract event features. Curr Biol 25:1945–1954. CrossRef Medline

Sohn MH, Ursu S, Anderson JR, Stenger VA, Carter CS (2000) The role of prefrontal cortex and posterior parietal cortex in task switching. Proc Natl Acad Sci U S A 97:13448–13453. CrossRef Medline

Stokes MG, Kusunoki M, Sigala N, Nili H, Gaffan D, Duncan J (2013) Dynamic coding for cognitive control in prefrontal cortex. Neuron 78:364–375. CrossRef Medline

Vandierendonck A, Liefooghe B, Verbruggen F (2010) Task switching: interplay of reconfiguration and interference control. Psychol Bull 136:601–626. CrossRef Medline

Vickery TJ, Kleinman MR, Chun MM, Lee D (2015) Opponent identity influences value learning in simple games. J Neurosci 35:11133–11143. CrossRef Medline

Wager TD, Nichols TE (2003) Optimization of experimental design in fMRI: a general framework using a genetic algorithm. Neuroimage 18:293–309. CrossRef Medline

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. CrossRef Medline

Wisniewski D, Reverberi C, Momennejad I, Kahnt T, Haynes JD (2015) The role of the parietal cortex in the representation of task–reward associations. J Neurosci 35:12355–12365. CrossRef Medline

Woolgar A, Hampshire A, Thompson R, Duncan J (2011) Adaptive coding of task-relevant information in human frontoparietal cortex. J Neurosci 31:14592–14599. CrossRef Medline

Wylie G, Allport A (2000) Task switching and the measurement of "switch costs". Psychol Res 63:212–233. CrossRef Medline