

Choosing the Rules: Distinct and Overlapping Frontoparietal Representations of Task Rules for Perceptual Decisions

Jiaxiang Zhang,¹ Nikolaus Kriegeskorte,¹ Johan D. Carlin,¹ and James B. Rowe^{1,2,3}

¹Cognition and Brain Sciences Unit, Medical Research Council, Cambridge CB2 7EF, United Kingdom, ²Department of Clinical Neurosciences, University of Cambridge, Cambridge CB2 2QQ, United Kingdom, and ³Behavioural and Clinical Neuroscience Institute, Cambridge CB2 3EB, United Kingdom

Behavior is governed by rules that associate stimuli with responses and outcomes. Human and monkey studies have shown that rule-specific information is widely represented in the frontoparietal cortex. However, it is not known how establishing a rule under different contexts affects its neural representation. Here, we use event-related functional MRI (fMRI) and multivoxel pattern classification methods to investigate the human brain's mechanisms of establishing and maintaining rules for multiple perceptual decision tasks. Rules were either chosen by participants or specifically instructed to them, and the fMRI activation patterns representing rule-specific information were compared between these contexts. We show that frontoparietal regions differ in the properties of their rule representations during active maintenance before execution. First, rule-specific information maintained in the dorsolateral and medial frontal cortex depends on the context in which it was established (chosen vs specified). Second, rule representations maintained in the ventrolateral frontal and parietal cortex are independent of the context in which they were established. Furthermore, we found that the rule-specific coding maintained in anticipation of stimuli may change with execution of the rule: representations in context-independent regions remain invariant from maintenance to execution stages, whereas rule representations in context-dependent regions do not generalize to execution stage. The identification of distinct frontoparietal systems with context-independent and context-dependent task rule representations, and the distinction between anticipatory and executive rule representations, provide new insights into the functional architecture of goal-directed behavior.

Introduction

A hallmark of cognitive control is the ability to regulate goal-directed behavior according to task rules, which coordinate cognitive and motor processes with knowledge of the associations among stimuli, responses, and outcomes (Sakai, 2008). Single-unit recordings have identified neuronal encoding of specific rules in a frontoparietal network (Quintana and Fuster, 1999; Asaad et al., 2000; Wallis et al., 2001). Human imaging studies have shown consistent rule-specific activity in homologous regions (Bunge et al., 2003; Sakai and Passingham, 2003; Haynes et al., 2007; Woolgar et al., 2011a). These representations of rules can be sustained between instruction and execution phases of a trial, and influence the subsequent behavioral performance (Sakai and Passingham, 2006) or drive selective attention to task-relevant information (Duncan, 2001; Miller and Cohen, 2001).

Task rules can be established under different contexts. For example, a rule may be formally identical, and executed equally

effectively, when it has been specifically instructed and when it has been chosen from several alternatives. Previous studies have examined the neural correlates of choosing a task rule, compared with specific instruction, identifying activation in the same frontoparietal network that also represents rule-relevant information (Forstmann et al., 2006; Rowe et al., 2008; Bengtsson et al., 2009). The functional overlap between contextual modulation and rule representation in these cortical regions raises the following three unresolved issues: (1) how rules are established under different contexts; (2) whether the context influences the neural representation of a given rule; and (3) whether the rule representation that is maintained in working memory matches the representation used when executing the same rule. There may exist robust and invariant coding of task rules under different contexts and at different stages of rule implementation, allowing consistent behavior in a changing environment. Alternatively, rule representations may be flexible and dependent on the context in which they were established.

To test these hypotheses, we examined the impact of context on rule representations. Three highly practiced rules that regulated perceptual decisions on distinct visual features (motion, color, and size) were established under two contexts: either chosen by the participant (chosen context), or instructed to them (specified context). We first identified regions with differential functional MRI (fMRI) responses when establishing a rule under the two contexts. Both in the regions exhibiting contextual differences and across the whole brain, we used multivariate pattern analysis (MVPA) (Norman et al., 2006) to investigate whether the

Received Nov. 7, 2012; revised April 19, 2013; accepted April 21, 2013.

Author contributions: J.Z., N.K., and J.B.R. designed research; J.Z. performed research; J.D.C. contributed unpublished reagents/analytic tools; J.Z. analyzed data; J.Z., N.K., J.D.C., and J.B.R. wrote the paper.

This work was supported by Medical Research Council Grant MC-A060-5PQ30 and Wellcome Trust Grant 088324. We thank John Duncan for very helpful comments.

The authors declare no competing financial interests.

This article is freely available online through the *JNeurosci* Author Open Choice option.

Correspondence should be addressed to Jiaxiang Zhang, MRC Cognition and Brain Sciences Unit, Cambridge CB2 7EF, United Kingdom. E-mail: jiaxiang.zhang@mrc-cbu.cam.ac.uk.

DOI:10.1523/JNEUROSCI.5193-12.2013

Copyright © 2013 the authors 0270-6474/13/3311852-11\$15.00/0

neural representation of an actively maintained rule is context dependent or context independent: a context-independent account predicts that the rule representation is independent of the context from which the rule is established, whereas a context-dependent account predicts distinct rule representations when the context changes. We further investigated whether the rule representation sustains at different stages of a trial, from maintaining to executing.

Our results demonstrate that the context under which rules are established can influence the representation of the rule in different frontoparietal regions, even when it is the same rule that guides subsequent behavior. However, the influence of context differs between cortical regions, indicating functionally distinct frontoparietal systems for adaptive cognitive control.

Materials and Methods

Participants. Twenty adults participated in the study (16 females; age range, 20–31 years; mean age, 24.65 years) following written informed consent. Participants were right handed with normal or corrected-to-normal vision, and none had a history of significant neurological or psychiatric illness. None had previous experience with the task. The data from one participant was excluded before analysis because of severe bias in selecting task rules. The study was approved by the local research ethics committee.

Stimuli. For all psychophysical training and fMRI scan sessions, stimuli for perceptual decisions were random dot kinematograms (RDKs) displayed within a central circular aperture (7.5° diameter) on a black background (100% contrast) with 50 ms frame duration. Eighty dots with three visual features (dot motion, dot color, and dot size) were plotted during each stimulus frame. To introduce apparent motion direction (upward or downward), a certain proportion (i.e., motion coherence) of the dots was replotted in the next frame at an appropriate spatial displacement (5°/s velocity), and the rest of the dots were replotted at random locations. Dots that moved outside the aperture were repositioned inside the aperture to avoid attention cues along the aperture edges. To ensure that the dots were evenly distributed across the entire aperture, the minimum distance between any two dots in each frame was 0.4°. To introduce apparent color information (blue or red), a certain proportion (i.e., color coherence) of the dots was rendered in one color while other dots were randomly rendered in either blue or red. The RGB values of the dot colors were equiluminant blue (5, 137, 255) and red (255, 65, 2) (Kayser et al., 2010). To introduce apparent size information (large or small), a certain proportion (i.e., size coherence) of the dots was presented in one size while other dots were randomly presented in either large (0.2° diameter) or small size (0.12° diameter). In particular, the coherence levels were held constant on each trial and the three visual features were independently applied to each dot in each frame, resulting in RDK stimuli with three uncorrelated visual features. Visual stimuli were presented by using Matlab 7.8 (MathWorks) and Psychophysics Toolbox 3 (www.psychtoolbox.org), and were displayed onto a screen with a resolution of 1024 × 768 pixels.

Task and procedures. All participants finished five psychophysical training sessions conducted on different days, followed by an MRI scan session. On each trial of the training and scan sessions, one of the three visual features was relevant to a pending task rule and needed to be attended. For the motion rule, participants were instructed to discriminate the coherent direction of motion (upward or downward). For the color rule, participants were instructed to discriminate the predominant color (blue or red). For the size rule, participants were instructed to discriminate the predominant size (large or small). Participants were instructed to respond as quickly and as accurately as possible by pressing one of two buttons on a button response box. The stimulus–response mappings for different rules were counterbalanced across participants.

Psychophysical training sessions. Participants were first familiarized with the three task rules before their first training session during a short practice session comprising 16 trials for each rule. To ensure that the participants were well trained on all rules, each training session com-

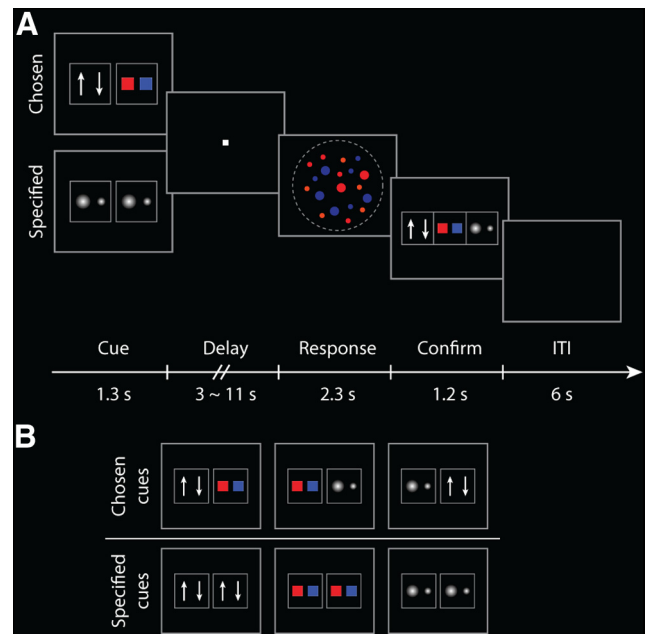


Figure 1. Behavioral paradigm for scan session. **A**, A rule cue was presented for 1300 ms at the beginning of each trial. In a chosen trial, participants selected one of two possible rules by themselves. In a specified trial, participants were instructed to a particular rule. Task cues were followed by a variable delay period (3000–11,000 ms), during which a fixation point was presented. After the onset of the RDK stimulus, participants made binary decision using the established rule. The RDK stimulus appeared for 2000 ms, after which there was a 300 ms gap and then the participants made a second response reporting which rule they had performed. **B**, Possible rule cues for the chosen and specified conditions.

prised three blocks, and within each block the participants were trained on one of the three task rules. The order of the three blocks was randomized across training sessions and participants. Each block comprised 216 trials, and participants were informed about which rule to implement at the beginning of each block. The rule-relevant feature (e.g., motion coherence in motion trials) was randomly presented with nine coherence levels at 2%, 3.84%, 5.76%, 8.64%, 12.96%, 19.44%, 29.16%, 43.75%, and 65.61% (24 trials per level). The two rule-irrelevant features (e.g., color and size coherences in motion trials) were randomly presented with two coherence levels at 2% and 65.61%. On each trial, the RDK stimulus was presented for 2000 ms, followed by a 2000 ms interstimulus interval during which a white fixation square (0.1°) was presented on the screen. Auditory feedback was given on incorrect responses (tone, 600 Hz; duration, 150 ms). To determine individual sensitivity, the discrimination accuracy of each rule in each training session was fit with a Weibull function using maximum likelihood methods. For each participant, the 82% correct thresholds of the three rules in the last training session were then estimated from individual psychometric functions and were used in the fMRI scan session.

fMRI scan session. Three cues (2.5° × 2.5° visual angle) were associated with the three task rules (motion discrimination, color discrimination, and size discrimination). At the beginning of each trial, two rule cues were presented horizontally (separated by 0.05°) for 1300 ms (Fig. 1A). On chosen trials, the two cues were associated with different rules, and the participants chose any one of the two available rules to perform. On specified trials, the two rule cues were the same and participants were instructed to implement the corresponding rule (Fig. 1B). Cue instruction was followed by a variable and unpredictable delay of 3000, 5000, 7000, 9000, or 11000 ms, during which a white fixation square (0.1°) was presented on the screen. After the delay, the RDK stimulus was presented for 2000 ms and the participants were required to make a two-alternative forced-choice response based on the selected or specified rule. To match the task difficulty across rules and individuals, the coherence levels of each visual feature were set to individual 82% correct thresholds estimated from the last training session. To register the chosen rule and

validate response accuracy, cues of all three possible rules were presented horizontally 300 ms after the RDK stimulus offset (1200 ms duration) and the participants were instructed to confirm which rule they performed in the current trial by a button press. The assignment of rules to buttons for the confirmation stage was fixed between trials, to avoid an additional visual search process during rule confirmation and to minimize potential errors regarding which task participants had performed. The intertrial interval was 6000 ms. The scan session comprised eight experimental runs, each of which lasted 534 s. Each run comprised 30 experimental trials, with each of the three chosen (motion/color, color/size, motion/size) or the three specified (motion, color, size) cue combinations presented in 5 trials. For each cue combination, each delay length was presented once. The orders of cue combinations and delay durations were randomized across runs and participants.

Data acquisition. A Siemens Tim Trio 3 T scanner (Siemens Medical Systems) was used to acquire BOLD-sensitive T2*-weighted EPI images [repetition time (TR), 2000 ms; echo time (TE), 30 ms, flip angle (FA), 78°; 32 slices; 3.0 mm thick; in-plane resolution 3×3 mm with slice separation 0.75 mm] in sequential descending order. Two hundred seventy volumes were acquired in each run, and the first three were discarded in further analysis to allow for steady-state magnetization. Participants also underwent high-resolution structural T1-weighted magnetization-prepared rapid acquisition gradient echo scans (MP-RAGE; TR, 2250 ms; TE, 2.99 ms; FA, 9°; inversion time, 900 ms; $256 \times 256 \times 192$ isotropic 1 mm voxels).

fMRI data preprocessing. MRI data were processed using SPM 8 (www.fil.ion.ucl.ac.uk/spm) in Matlab 7.8 (MathWorks). Functional MRI data were converted from DICOM to NIFTII format, spatially realigned to the first image in the first run, and interpolated in time to the middle slice to correct acquisition delay. The mean fMRI and MP-RAGE volumes were coregistered using mutual information. For univariate analysis, the MP-RAGE volumes were segmented and normalized to the Montreal Neurological Institute T1 template in SPM by linear and nonlinear deformations. The normalization parameters were applied to all spatiotemporally realigned functional images, the mean image, and structural images, obtaining normalized volumes with a voxel size of $3 \times 3 \times 3$ mm. Normalized fMRI data were smoothed with an isotropic Gaussian kernel with full-width at half-maximum of 8 mm. No normalization or spatial smoothing was performed on the fMRI data used for the multivariate analysis. Freesurfer (<http://surfer.nmr.mgh.harvard.edu>) was used for individual cortical surface reconstruction. The T1-weighted structural images from individual participants were used for intensity normalization, skull stripping, segmentation, 3D cortex reconstruction, and cortical surface inflation.

Univariate analysis. To evaluate the effects of rule selection and maintenance, normalized and smoothed functional images were analyzed using a conventional univariate approach in SPM 8. First-level general linear model (GLM) included six epoch regressors representing the sustained activations during delay periods associated with the three task rules under chosen and specified conditions, with onset times locked to the onsets of rule cues and with durations matched to the length of the delays. The model also included (1) six event regressors representing the onsets of the rule cues for the three task rules under the two contexts; (2) three event regressors for activations in response to the RDK stimuli under the three task rules; and (3) one event regressor for activations in response to the second button response in all conditions. Six rigid-body motion correction parameters and run means were included as nuisance covariates. All epoch and event regressors were convolved with a canonical hemodynamic response function, and the data were high-pass filtered with a frequency cutoff at 128 s. Two second-level ANOVAs were conducted, one on the transient activations at rule cue onsets and the other on the sustained activations during delay periods. Each analysis included six first-level contrast images (three task rules in chosen and specified context conditions) from each participant, adjusted for nonsphericity with dependence between measures and unequal variance. Statistical parametric maps were then generated for effects of interest and corrected for multiple comparison at $p < 0.05$ (FWE) across the whole brain.

Regions of interest. We identified regions of interest (ROIs) in the frontoparietal cortical surface that showed significant difference between chosen and specified cues (voxelwise $p < 0.001$ uncorrected, and a cluster

extent threshold of 35 voxels). We also identified one additional ROI [Brodmann's area (BA) 17] in the visual cortex from a Brodmann's area map (Fischl et al., 2008). All ROIs were defined on a cortical surface template and mapped onto an individual's cortical surface using FreeSurfer. The individual surface ROIs were then transformed into individual native space and resampled at the raw resolution with a voxel size of $3 \times 3 \times 3.75$ mm.

Multivoxel pattern analysis. We used MVPA (Haynes and Rees, 2005; Norman et al., 2006) to decode rule-specific information during active rule maintenance. MVPA takes advantage of aggregating information across multiple voxels with different response selectivity (Kriegeskorte et al., 2010) and has been shown to be more sensitive than conventional univariate approaches (Kamitani and Tong, 2005; Zhang et al., 2010).

To obtain fMRI activation patterns for multivariate analysis, a first-level GLM was estimated for each participant using the realigned and slice-timed (but un-normalized and unsmoothed) images. The model used one epoch for each delay period, with durations matched to the length of the delay, resulting in 240 parameter estimates of sustained activations for each participant. Similar to the univariate analysis described above, the model also included covariates for transient effects of rule cue onsets, RDK stimulus onsets, and the second button responses. Parameter estimates of the voxels in each ROI were extracted to construct the multivoxel patterns.

We trained linear support vector machine (SVM) classifiers using the multivoxel patterns per ROI and calculated mean classification accuracies following a leave-one-run-out cross-validation procedure. That is, for each cross-validation seven of eight fMRI scan runs were used as a training set, and the one independent remaining run was used as a test set. The classifiers were trained on the training set to identify patterns corresponding to each rule, and were tested for classification accuracy on the test set. The procedure was repeated eight times, and the accuracies were averaged from eightfold cross-validation. We included all trials in the main analysis, and also conducted an additional analysis only using the subset of trials with correct decision responses. Analysis was performed using MATLAB routines provided in the Princeton MVPA Toolbox (www.csmb.princeton.edu/mvpa) and LIBSVM implementation (<http://www.csie.ntu.edu.tw/~cjlin/libsvm>). The tradeoff between errors of the SVM on training data and margin maximization was set to 1.

To enable comparisons across ROIs and participants, for each cross-validation we selected the same number of voxels across ROIs and participants with the highest difference between conditions (one-way ANOVA across conditions). We selected 120 voxels per ROI, comparable with the dimensionality used in previous studies (Haynes and Rees, 2005; Kamitani and Tong, 2005; Zhang and Kourtzi, 2010). The feature selection procedure was used only on the data from the training dataset, independently during each fold of cross-validation. Pattern classification and feature selection were, thus, performed on different datasets, and no risk of circular inference was introduced (Kriegeskorte et al., 2009). Also, to ensure that the classifier output did not simply result from univariate differences across conditions, for each ROI we subtracted the mean value across voxels from the response of each voxel and divided by the SD across voxels (Misaki et al., 2010).

To determine whether we could predict the patterns from the three task rules, we used multiple pairwise (one-against-one) binary classifiers (Kamitani and Tong, 2005; Preston et al., 2008; Serences et al., 2009; Zhang et al., 2010). In particular, we trained and tested three pairwise classifiers and collated their results for each test pattern. The predicted rule corresponded to the condition that received the fewest "votes against" when collating the results across all pairwise classifications. In the event of a tie, the prediction was randomly assigned to one of the conditions. We expressed the accuracy of the three-way classifier as the proportion of test patterns for which it correctly predicted the true conditions.

To estimate the significance of the classification accuracies, we used nonparametric permutation tests, without making assumptions about the probability distribution of the data (Edgington and Onghena, 2007; Pereira et al., 2009). In particular, for each ROI in each participant, we ran the classification analysis in 5000 permutations, with the data labels

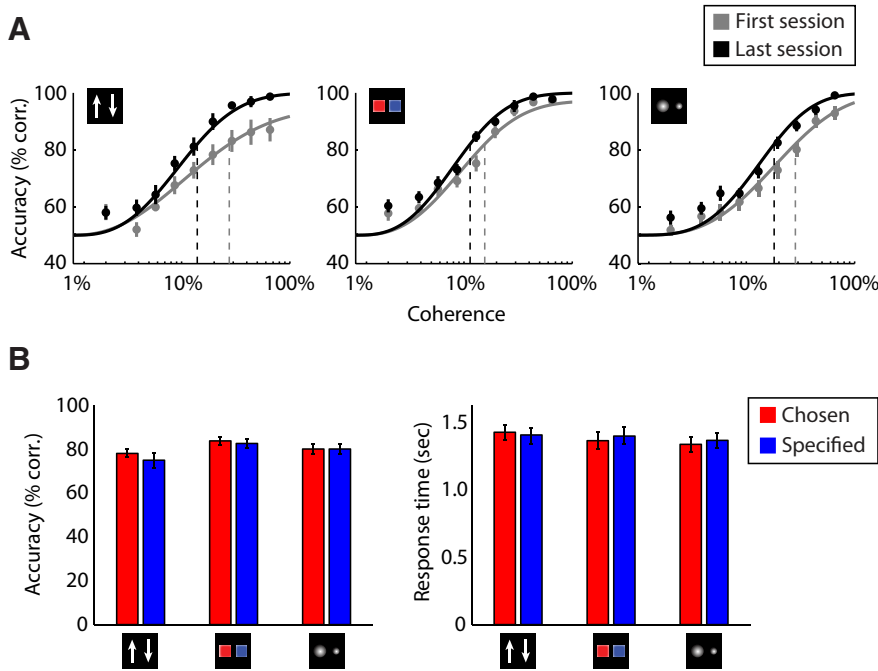


Figure 2. Behavioral results. **A**, Decision accuracy (average data across participants) of the three task rules across coherence levels in the first and the fifth training sessions. The curves indicate the best fit of the Weibull function. The dashed lines indicate the 82% correct threshold. **B**, Average decision accuracy and response time across participants in the scan session. Error bars denote SEs.

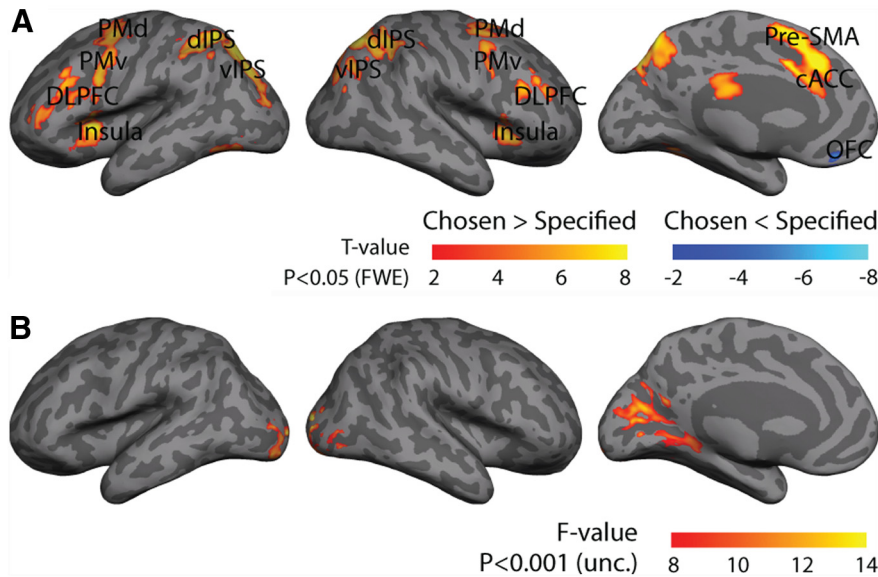


Figure 3. fMRI responses of establishing a rule at the initial cue stage. **A**, Brain regions that showed increased and decreased activity in voluntary rule selection compared with specified rule instruction ($p < 0.05$, FWE corrected). **B**, Brain regions that showed a main effect of task rules when establishing the rule (voxelwise $p < 0.001$ uncorrected, with a cluster extent threshold of 35 voxels).

shuffled separately in the training and test sets for each cross-validation in each permutation. This gave us a sampling distribution of the mean classification accuracy under the null hypothesis that there is no information of rule representations in the multivoxel patterns. The level of significance (p value) was estimated by the fraction of the permutation samples that were greater than or equal to the classification accuracy from the data without label shuffling. To account for the multiple statistical tests that were performed for all ROIs, we evaluated the results using a Bonferroni-corrected threshold ($p < 0.05$) for significance.

Whole-brain searchlight analysis. A secondary whole-brain searchlight analysis was performed to identify all regions coding task-relevant infor-

mation (Kriegeskorte et al., 2006). For each participant, multivoxel patterns (un-normalized and unsmoothed data) were extracted from a sphere with a radius of 2 voxels, and the sphere moved through each voxel in the brain. A linear SVM was trained and tested as in the ROI analysis, using the data from each sphere. The classification accuracy of each sphere was assigned to the central voxel in the sphere, yielding an image of whole-brain classification accuracy. The classification accuracy images for individual participants were normalized by applying the deformation parameters obtained at the preprocessing stage, and smoothed with an isotropic Gaussian kernel with full-width at half-maximum of 8 mm. These images were then entered into a second-level one-sample t test to identify voxels that showed significantly higher than chance level classification accuracy.

Eye-tracking data analysis. We recorded eye-movement data during fMRI scan sessions from 11 participants using a 50 Hz SMI eye-tracker (SensoMotoric Instruments). Eye-movement data were preprocessed using the BeGaze software (SensoMotoric Instruments) and analyzed using custom codes in Matlab (MathWorks). For each condition, we computed the horizontal and vertical eye position and the number of saccades per condition, separately for rule cue stage (1300 ms intervals) and rule execution stage (2000 ms intervals during which RDK stimuli were presented).

Results

Behavioral results

Participants were trained on three task rules for making two-alternative perceptual decisions: motion, color, and size discrimination. To quantify individual participants' performance, we measured decision accuracy (percentage of correct responses) and 82% discrimination threshold (estimated from Weibull fits on individual data) for each task rule from five consecutive 1 h training sessions. A repeated-measures ANOVA on behavioral performance showed significant main effects of task rules (accuracy: $F_{(2,36)} = 10.71$, $p < 0.001$; threshold: $F_{(2,36)} = 5.15$, $p < 0.05$) and training sessions (accuracy: $F_{(4,72)} = 13.32$, $p < 0.0000001$; threshold: $F_{(4,72)} = 8.67$, $p < 0.00001$). Participants' sensitivity in executing the task rules improved over the course of training (Fig. 2A), as shown by a significant linear trend over training sessions (accuracy: $F_{(1,18)} = 39.55$, $p < 0.00001$; threshold: $F_{(1,18)} = 19.67$, $p < 0.001$). In particular, for each task rule we observed significantly increased decision accuracy (motion rule: $t_{(18)} = 4.54$, $p < 0.001$; color rule: $t_{(18)} = 2.81$, $p < 0.05$; size rule: $t_{(18)} = 4.09$, $p < 0.001$) and decreased 82% threshold (motion rule: $t_{(18)} = -3.33$, $p < 0.01$; color rule: $t_{(18)} = -2.29$, $p < 0.05$; size rule: $t_{(18)} = -2.71$, $p < 0.05$) by the last training session. Behavioral performance across the three rules did not significantly differ in the last two training sessions (accuracy: $F_{(1,18)} = 1.62$, $p = 0.22$; threshold: $F_{(1,18)} = 2.23$, $p = 0.15$), indicating that a steady level of

performance was attained after training. No significant differences on response times (RTs) were found between the first and the last training sessions (motion rule: $F_{(1,18)} = 1.44, p = 0.26$; color rule: $F_{(1,18)} = 1.10, p = 0.30$; size rule: $F_{(1,18)} = 0.03, p = 0.87$).

To test whether participants' responses could be affected by task-irrelevant information, we segregated behavior data according to coherence levels of the irrelevant features (e.g., color coherence for trials with the motion or size rules). No significant effects on the coherence levels of the irrelevant features were observed in the last training session (accuracy: $F_{(1,18)} = 0.20, p = 0.66$; response time: $F_{(1,18)} = 0.29, p = 0.60$). These results suggest that after training participants' responses were not likely to be influenced by irrelevant features in the rule-based perceptual decision tasks.

To match the difficulty of the three task rules in the fMRI session, the coherence level for each visual feature was set to the 82% accuracy threshold for each participant, estimated from the last training session. During scanning, participants performed the decision tasks with high accuracies (motion rule, $76.30 \pm 9.63\%$; color rule, $82.95 \pm 7.85\%$; size rule, $79.84 \pm 8.11\%$). Repeated-measures ANOVAs showed a marginal effect in performance between task rules (accuracy: $F_{(2,36)} = 2.94, p = 0.07$; response time: $F_{(2,36)} = 3.03, p = 0.06$) (Fig. 2B,C), but no significant difference in performance between chosen and specified conditions (accuracy: $F_{(1,18)} = 1.38, p = 0.26$; response time: $F_{(1,18)} = 1.367, p = 0.21$), between trials in which participants changed rules and stayed on the same rule as in the last trial (accuracy: $F_{(1,18)} = 0.35, p = 0.56$; response time: $F_{(1,18)} = 0.04, p = 0.84$), or between trials with different delay durations in the chosen condition (accuracy: $F_{(4,72)} = 0.43, p = 0.79$; response time: $F_{(4,72)} = 1.21, p = 0.32$) and specified condition (accuracy: $F_{(4,72)} = 0.50, p = 0.74$; response time: $F_{(4,72)} = 0.30, p = 0.88$). Further, for the chosen condition participants were not biased to choose any particular rule (Friedman test, $\chi^2 = 0.76, p = 0.69$) and were not biased to select rules presented left or right to the central fixation (Wilcoxon signed-rank test, $Z = -0.18, p = 0.86$). These results suggest that task difficulties were similar between rules and between contexts by our task manipulation, and that participants made rapid decisions on the rule to perform after cue onset in both chosen and specified conditions.

Establishing and maintaining the neural representations of rules under different contexts: univariate analysis

To compare BOLD responses in establishing task rules under chosen and specified conditions, we performed a random-effects ANOVA with one context factor (chosen, specified) and one rule factor (motion, color, size) on transient activations to rule cues presented at the beginning of each trial. A whole-brain univariate analysis yielded significantly higher activations ($p < 0.05$, FWE corrected) for voluntary rule selection, compared with specified rule instruction, in medial-frontal, dorsolateral prefrontal, and parietal cortices. We also observed a significantly greater fMRI response for establishing specified rules versus chosen rules in the medial orbitofrontal cortex (OFC) (Fig. 3A; Table 1).

Based on the univariate analysis and our hypothesis that a frontoparietal network establishes rule representations (Sakai and Passingham, 2003, 2006; Rushworth et al., 2004; Forstmann et al., 2006; Rowe et al., 2007, 2008; Woolgar et al., 2011a), we defined nine bilateral cortical ROIs for further multivariate analysis of between-rule effects. These ROIs showed different responses between establishing chosen and specified rules, including premotor dorsal (PMd), premotor ventral (PMv), dorsolateral prefrontal cortex (DLPFC), insula, pre-supplementary

Table 1. Regions associated with stronger transient activation for chosen rule cues or specified rule cues (Fig. 3A)

Region	t value	MNI coordinates		
		x	y	z
Chosen > specified				
PMd	7.65	-27	-4	52
	7.82	30	-1	49
PMv	7.30	-45	2	34
	5.31	45	2	37
DLPFC	6.79	-45	29	28
	6.85	45	32	28
Insula	9.34	-33	17	4
	7.97	30	23	4
Pre-SMA	10.81	-3	14	49
	7.23	-6	23	34
cACC	8.67	-33	-46	46
	7.47	36	-49	46
vIPS	8.72	-21	-67	40
	6.57	33	-73	31
Precuneus	8.60	9	-64	49
	7.94	-9	-64	49
Middle cingulate gyrus	6.67	-3	-25	28
	6.13	-30	17	4
Fusiform gyrus	6.13	-30	17	4
Talamus	5.18	12	-16	10
Chosen < specified				
Medial OFC	-5.66	-6	38	-17
Angular gyrus	-4.71	-51	-70	28

Values are statistics ($p < 0.05$, FWE corrected) and peak coordinates (separated by ≥ 8 mm) reported in MNI space (mm).

Table 2. Regions associated with stronger sustained activation for maintaining specified rules than voluntary rules during delay periods

Region	t value	MNI coordinates		
		x	y	z
Middle orbitofrontal cortex	5.97	0	47	-11
Middle frontal gyrus	5.36	-24	35	43
Ventral intraparietal sulcus	6.05	-39	-79	34
Precuneus	4.98	-6	-55	28
Middle cingulate gyrus	4.78	-3	-43	43
Calcarine sulcus	4.86	-6	-55	7

Values are statistics ($p < 0.05$, FWE corrected) and peak coordinates (separated by ≥ 8 mm) reported in MNI space (mm). The reverse contrast showed no significant activations.

motor area (SMA), caudal anterior cingulate cortex (cACC), medial OFC, dorsal intraparietal sulcus (dIPS), and ventral IPS (vIPS). Because the visual cortex may respond to the presentations of rule cues, we also defined one additional ROI in BA 17 from an independent template (Fischl et al., 2008), although no significant differences between chosen and specified conditions were observed in this region.

Our task design incorporated a variable delay period (3–11 s) between rule cue offset and RDK stimulus onset (Fig. 1A), during which participants maintained an established task rule (Bengtsson et al., 2009). BOLD response associated with rule maintenance was modeled as epochs with variable lengths, matched to delay durations. In this way, we can identify sustained activations during the delay period, separately from transient activations associated with establishing rules at the beginning of the trial or executing rules at the RDK stimulus onset (Sakai and Passingham, 2003, 2006). A random-effects ANOVA (context \times rule) on sustained activations during delay periods showed higher activations ($p < 0.05$, FWE corrected) for maintaining specified rules than maintaining chosen rules in the medial orbitofrontal and inferior parietal cortex (Table 2). The reverse contrast revealed no significant activation, indicating that a uni-

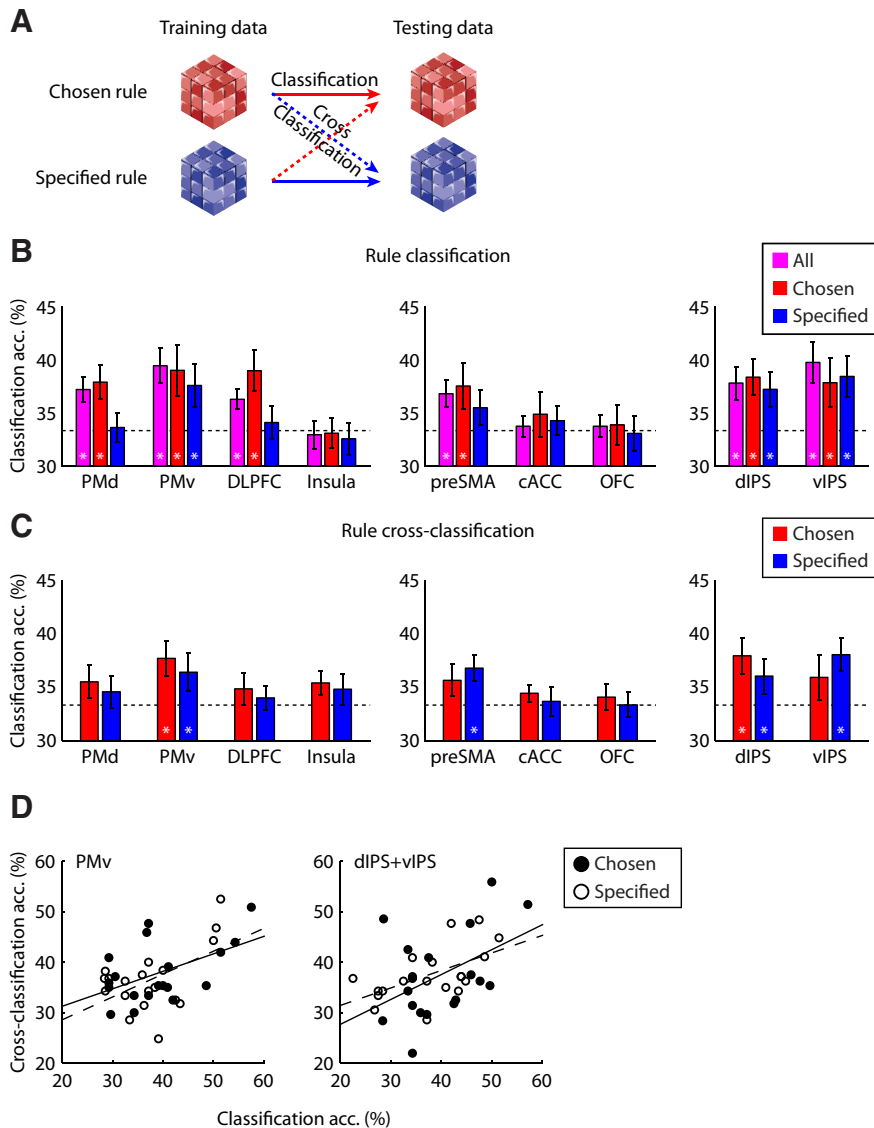


Figure 4. MVPA of rule representation for chosen and specified contexts. **A**, Schematic diagram of the MVPA procedure (classification and cross-classification). **B**, MVPA accuracy (percentage of correct classification) per ROI for the classification of the three rules. Data are shown for all conditions as well as the chosen and specified conditions separately. **C**, Cross-classification accuracy per ROI for discriminating the three rules. The classifier was trained on data from the specified condition and used to decode rules in the chosen condition (red), and vice versa (blue). Error bars denote SEM. Asterisks indicate regions where decoding accuracy was significantly ($p < 0.05$, nonparametric permutation tests) different from chance level (33.33%, dashed line). **D**, Correlation between classification and cross-classification. Each data point represents one participant. The lines indicate the linear regression for chosen (solid line) and specified (dashed line) conditions.

variate analysis failed to identify sustained activation as greater for chosen rules than specified rules, in contrast with the transient activation with selection at the start of each trial.

We then examined whether there was a main effect of task rules, regardless of whether the rules were voluntarily chosen or specifically instructed. Transient activation associated with rule cue presentation was modulated by task rules in the occipital lobe, including the inferior occipital gyrus, lingual gyrus, and cuneus (voxelwise $p < 0.001$ uncorrected and a cluster extent threshold of 35 voxels) (Fig. 3B). No significant effect of task rules during delay periods was observed. In other words, there was no univariate effect within the frontoparietal network that distinguished the three task rules.

Representation of chosen and specified rules: multivariate analysis

Using MVPA (Haynes and Rees, 2005; Norman et al., 2006), we examined whether distributed fMRI patterns in the frontoparietal ROIs contain rule-specific information during active rule maintenance. In particular, we constructed multivoxel activation patterns of rule maintenance from response-amplitude estimates of sustained activations during delay periods on a trial-by-trial basis (using unsmoothed and un-normalized data), and used linear support vector machines to discriminate between fMRI patterns associated with the three task rules (Fig. 4A).

We first considered which frontoparietal ROIs maintained rule-specific information, regardless how the rule was established. By pooling data from the chosen and specified conditions, we observed significant coding of which rule the participants were intending to perform during delay periods (Fig. 4B) in PMd, PMv, DLPFC, pre-SMA, dIPS, vIPS, and BA 17 ($p < 0.01$), but not in insula, cACC, and OFC ($p > 0.99$). We also conducted whole-brain searchlight analysis to identify any additional regions that encode task-relevant information. The searchlight analysis largely confirmed the ROI results (Fig. 5A), with significant rule coding in the PMv, DLPFC, dIPS, and the visual cortex (Table 3).

We then tested rule coding for the two context conditions separately. A repeated-measures ANOVA on classification accuracies showed a higher classification accuracy in the chosen condition compared with the specified condition in the lateral frontal cortex ($F_{(1,18)} = 7.60$, $p < 0.01$), but not in the medial frontal ($F_{(1,18)} = 2.16$, $p = 0.16$) or parietal cortex ($F_{(1,18)} = 0.05$, $p = 0.82$). Significant rule coding was observed in both chosen and specified conditions in PMv (chosen: $p < 0.01$; specified: $p < 0.01$), dIPS (chosen: $p < 0.01$; specified: $p < 0.01$), vIPS (chosen: $p < 0.01$; specified: $p < 0.01$), and BA

17 (chosen: $p < 0.01$; specified: $p < 0.01$). Rule-specific representation was significant only in the chosen condition but not the specified condition in PMd (chosen: $p < 0.01$; specified: $p > 0.99$), DLPFC (chosen: $p < 0.01$; specified: $p > 0.99$), and pre-SMA (chosen: $p < 0.01$; specified: $p = 0.20$). No significant encoding of rule information was observed in insula (chosen: $p > 0.99$; specified: $p > 0.99$), cACC (chosen: $p = 0.63$; specified: $p > 0.99$), or OFC (chosen: $p > 0.99$; specified: $p > 0.99$). When comparing mean classification accuracies in ROIs with significant coding in both chosen and specified conditions (PMv, dIPS, vIPS, and BA 17) versus those with significant coding only in the chosen condition (PMd, DLPFC, and pre-SMA), there was a significant interaction between rule-cue context (chosen vs specified) and ROI groups ($F_{(1,18)} = 5.01$, $p < 0.05$). These results suggest a broad representation of rule-specific information in a

frontoparietal network during maintenance of the three perceptual decision rules.

Generalization and consistency of rule representation across contexts

Maintaining a voluntarily chosen task rule might elicit the same pattern information as maintaining the same rule established by specific instructions. Alternatively, the representation may be distinct, according to the context under which the rules were established. To distinguish these alternatives, we examined the relationship between the fMRI pattern information of chosen and specified rules within ROIs. Specifically, we trained the pattern classifier on specified rules and tested its generalized performance on discriminating chosen rules, and vice versa (i.e., “cross-classification”; Fig. 4A).

The pattern classifier trained on specified rules could accurately discriminate between chosen rules in PMv ($p < 0.01$) and dIPS ($p < 0.01$), and marginally in vIPS ($p = 0.06$). Multivoxel patterns in the same regions can also discriminate between specified rules when the classifier was trained on chosen rules [PMv ($p < 0.05$), dIPS ($p < 0.05$), and vIPS ($p < 0.01$)], suggesting that rule representations in these brain regions are independent of the context under which the rule was established (Fig. 4C). We also observed significant classification accuracy in BA 17 for training on chosen rules and testing on specified rules ($p < 0.01$), and for the reverse direction of cross-classification ($p < 0.01$).

Interestingly, for pre-SMA the classifier trained on chosen rules could accurately discriminate between specified rules ($p < 0.05$), but not for the opposite direction ($p = 0.13$), suggesting that rule representation in the pre-SMA is not fully context independent, but that some voxels in this region may retain their response preferences independently under different contexts.

A whole-brain searchlight analysis for cross-classification between chosen and specified conditions was consistent with the ROI results (Fig. 5B; Table 3). Cross-classification accuracies were significant above chance in the PMv, and also in the IPS and the visual cortex at a lower threshold (voxelwise $p < 0.001$ uncorrected and a cluster extent threshold of 35 voxels).

We further examined whether there is a consistency of classification accuracies between different analyses in the context-independent ROIs. Correlation analysis performed on classification accuracies across all participants confirmed a significant relationship between classification and cross-classification results in PMv (chosen: $r = 0.52$, $p < 0.05$; specified: $r = 0.53$, $p < 0.05$). When dIPS and vIPS voxels were pooled together, we also observed consistency between classification and cross-classification accuracies (chosen: $r = 0.45$, $p = 0.05$; specified: $r = 0.56$, $p < 0.05$) (Fig. 4D). No significant correlation was observed in BA 17 (chosen: $r = 0.36$, $p = 0.13$; specified: $r = 0.34$, $p = 0.15$).

Invariant rule representation during maintenance and execution stages

Because participants executed rules that had been maintained during delay periods, it is possible that rule-specific information during maintenance relates to its representation during the later

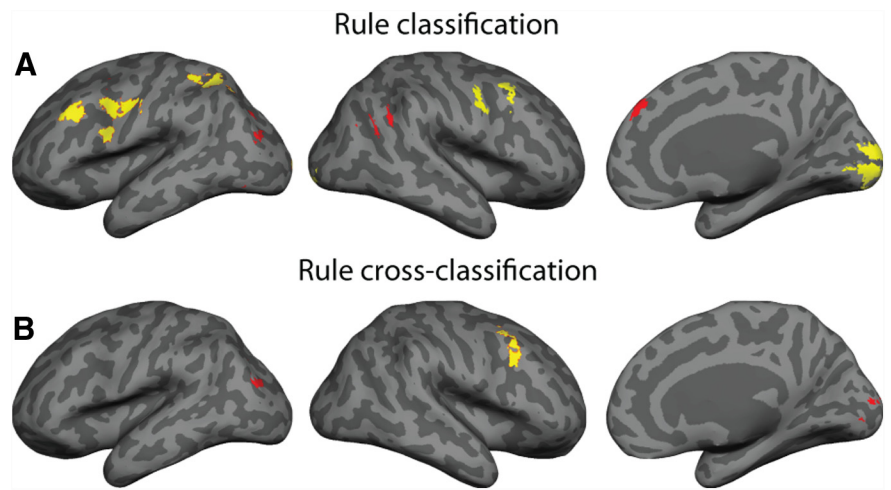


Figure 5. Whole-brain searchlight analysis results. Regions with significant classification accuracies above the chance level are displayed in yellow ($p < 0.05$ corrected for cluster-size threshold) or red (voxelwise $p < 0.001$ uncorrected, with a cluster extent threshold of 35 voxels). **A**, Multivariate searchlight analysis of rule classification during delay periods. Data from chosen and specified trials are pooled together. **B**, Multivariate searchlight analysis of rule cross-classification between chosen and specified conditions.

Table 3. Regions associated with significant coding of task rules in whole-brain searchlight analysis

Analysis	Region	t value	MNI coordinates		
			x	y	z
Rule classification	PMv*	4.51	-51	9	36
		4.10	57	9	42
	DLPFC*	5.96	-48	27	36
	dIPS*	4.99	-27	-54	51
	Calcarine sulcus*	5.13	-6	-96	-3
		5.31	15	-90	-3
	Medial superior frontal gyrus	5.21	9	45	42
	vIPS	3.78	-27	-72	30
	Supramarginal gyrus	5.08	45	-48	27
	Rule cross-classification	PMv*	4.84	45	21
vIPS		5.15	-42	-78	21
Calcarine sulcus		4.17	6	-93	6

Values are statistics ($*p < 0.05$ cluster corrected, or $p < 0.001$ uncorrected with a cluster extent threshold of 35 voxels) and peak coordinates reported in MNI space (mm).

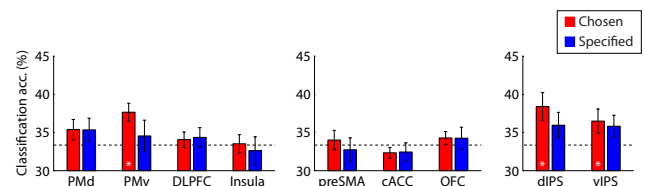


Figure 6. Generalization of rule representation between maintenance and execution. MVPA accuracies per ROI for the classification of three rules across maintenance and execution stages are shown for chosen and specified conditions. Error bars denote SEM.

execution stage. Therefore, we constructed fMRI patterns of rule execution from activations at the onset of the RDK stimuli, and examined whether classifiers trained on patterns from maintenance stage could discriminate patterns from execution stage (Fig. 6). This cross-stage classification analysis revealed invariant rule representations across maintenance and execution stages for the chosen condition in dIPS ($p < 0.01$), vIPS ($p < 0.05$), and PMv ($p < 0.01$), and marginally for the specified condition in dIPS ($p = 0.09$) and vIPS ($p = 0.09$), but not in PMv ($p > 0.99$). Other ROIs with significant rule-specific information during

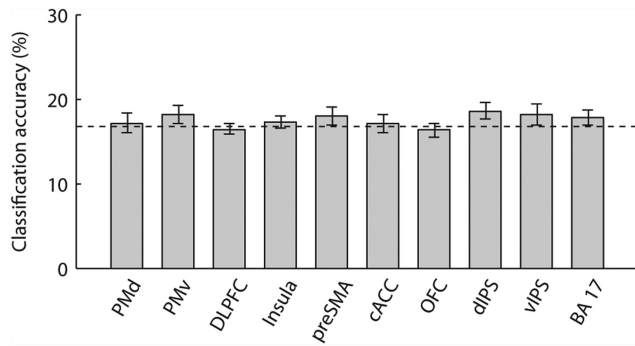


Figure 7. MVPA on task cues in the chosen condition. Classification accuracies to discriminate the six task cues in the chosen condition (motion/color, color/motion, size/color, color/size, motion/size, and size/motion) were estimated from a six-way classification for each ROI. Error bars denote SEMs across participants. Mean classification accuracies are not significantly different from chance level (16.67%, dashed line) in any ROI ($p > 0.10$, nonparametric permutation tests).

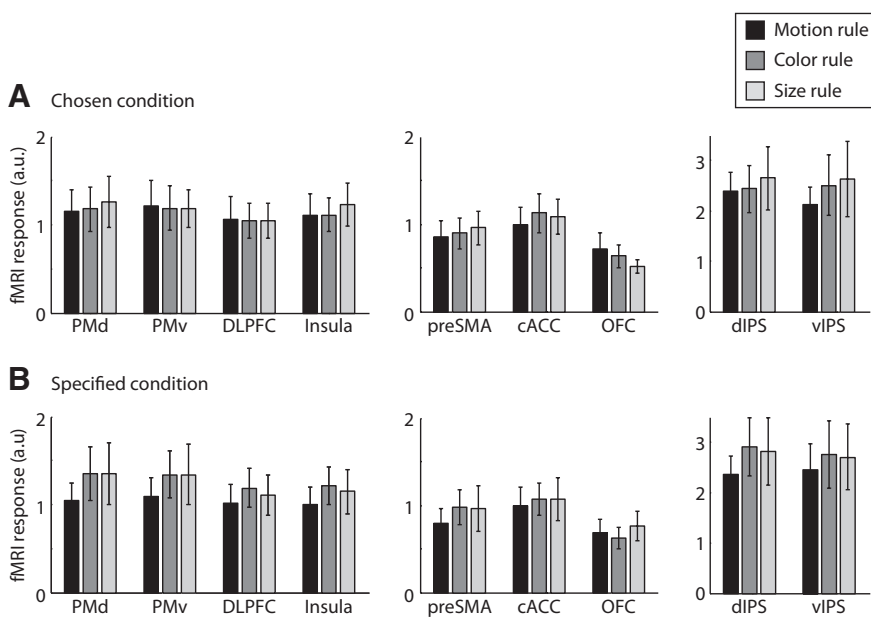


Figure 8. Regional fMRI responses during delay periods. fMRI responses during delay periods were estimated from unnormalized and unsmoothed data in each ROI. **A, B**, Data are shown for each task rule in chosen (**A**) and specified (**B**) conditions. Error bars denote SEMs across participants.

maintenance stage did not have invariant representations during execution stage: PMd (chosen: $p = 0.21$; specified: $p = 0.24$), DLPFC (chosen: $p > 0.99$; specified: $p > 0.99$), pre-SMA (chosen: $p > 0.99$; specified: $p > 0.99$), and BA 17 (chosen: $p > 0.99$; specified: $p > 0.99$).

Control analyses: low-level visual features, motor responses, rule difficulties, and eye movements

Is it possible that the significant rule-specific information we observed could be caused by differences in low-level visual features in the rule cues presented at the beginning of each trial? To rule out this possibility, we conducted a control analysis by using a six-way classifier to discriminate the six possible rule cues for the chosen condition (Fig. 1B). This analysis showed that no ROI encodes information of rule cues during delay periods (Fig. 7) and found no significant difference in the classification accuracies between the ROIs ($F_{(9,162)} = 0.73$, $p = 0.68$), suggesting that our results of rule representation during

maintenance cannot be attributed to differences in low-level visual features in rule cues. However, it is recognized that this six-way classification analysis has reduced power because of fewer patterns per class than the main classification analysis. Further ROI analysis of the univariate response on unsmoothed and unnormalized data showed no significant effect of rules in either chosen ($F_{(2,36)} = 0.25$, $p = 0.78$) or specified conditions ($F_{(2,36)} = 1.28$, $p = 0.29$) (Fig. 8), suggesting that our results could not be caused by differences in the overall BOLD response between rules.

We investigated whether planning, encoding, or executing motor responses for rule execution may have an effect on rule encoding during delay periods. Note that the correct responses were randomized across trials and hence were dissociated from task rules, allowing us to decode participant's binary motor responses. This analysis did not show any ROI that encodes information about future button press during delay periods ($p > 0.19$, permutation tests). A second analysis tested whether the observed rule-specific information could be explained by the

marginal differences in behavioral performance across rules. We repeated the MVPA procedure only using trials with correct responses and RT no longer than the 90% percentile of individual RT distributions (75.39% of total trials). The results were similar to the main analysis with all trials included ($F_{(1,18)} = 0.27$, $p = 0.61$). For the chosen condition, rule decoding was significant in the PMd, PMv, DLPFC, pre-SMA, dIps, and vIPS ($p < 0.01$, permutation tests). For the specified condition, rule decoding was significant in the PMv ($p < 0.01$), and marginally significant in the dIps ($p = 0.08$) and vIPS ($p = 0.05$). Therefore, the significant rule coding cannot be adequately explained by the marginal performance differences between rules.

Finally, analysis of eye movement data did not show any significant difference in the eye position or number of saccades between context conditions or task rules (Fig. 9), suggesting that our results could not readily be attributed to eye movements during rule cue or execution stages.

Discussion

Our findings provide novel insights into the functional organization and properties of the neural systems underlying goal-directed behaviors. Using MVPA, we found that a frontoparietal network prominently encodes rule-specific information when a task rule is actively maintained before its actual performance, consistent with previous studies (Li et al., 2007; Rowe et al., 2007; Woolgar et al., 2011a).

However, our results suggest that regions within this network differentially contribute to the representations of rules, according to whether the rules were chosen by participants themselves or instructed to the participants. In particular, rule representations in the dorsolateral and medial frontal cortex depend on rule-cue context: the rules could be decoded only when the rules were voluntarily chosen by participants, but not when the rules were instructed. In contrast, rule representations maintained in the ven-

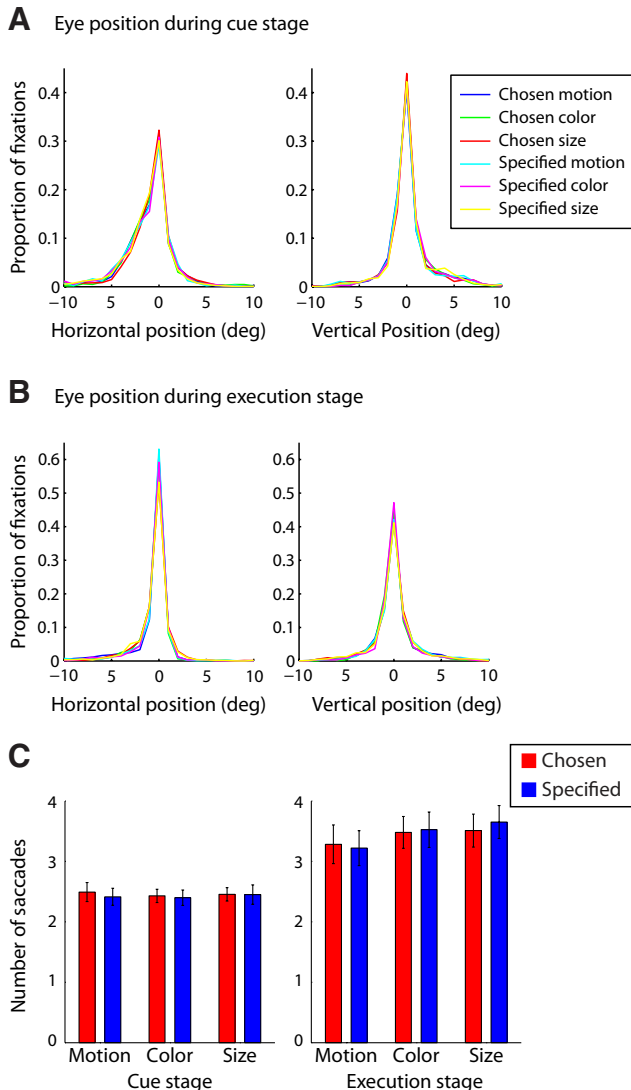


Figure 9. *A–C*, Eye-tracking data of the horizontal (*A*) and vertical (*B*) eye position and the number of saccades per condition (*C*), separately for rule cue and execution stages. A repeated-measures ANOVA indicated that there was no significant difference between task rules on mean horizontal eye position (cue stage: $F_{(2,20)} = 0.56, p = 0.58$; execution: $F_{(2,20)} = 0.56, p = 0.58$), mean vertical eye position (cue stage: $F_{(2,20)} = 0.60, p = 0.56$; execution: $F_{(2,20)} = 0.43, p = 0.66$), or the number of saccades per trial (cue stage: $F_{(2,20)} = 0.22, p = 0.80$; execution: $F_{(2,20)} = 1.41, p = 0.27$). In addition, no significant differences were observed between rule context (chosen vs specified) for mean horizontal eye position (cue stage: $F_{(1,10)} = 2.12, p = 0.17$; execution: $F_{(1,10)} = 1.09, p = 0.32$), mean vertical eye position (cue stage: $F_{(1,10)} = 2.68, p = 0.13$; execution: $F_{(1,10)} = 2.71, p = 0.31$), or number of saccades (cue stage: $F_{(1,10)} = 0.19, p = 0.67$; execution: $F_{(1,10)} = 0.56, p = 0.47$). These analyses suggest that our results could not readily be attributed to eye-movement differences between conditions.

trolateral frontal and parietal cortex were significant in both chosen and specified contexts and were context independent, as indicated by cross-classifications between different rule-cue contexts. Furthermore, our study shows how maintaining rule-specific information in anticipation of stimuli can influence the later execution of the rule: rule representations in context-independent regions are invariant from maintenance to execution stages, whereas rule representations in the context-dependent regions do not generalize to execution stage.

Previous work has linked dedicated frontal regions to particular task rules (Sakagami and Tsutsui, 1999; Sakai and Passingham, 2003). However, neurophysiological studies in nonhuman

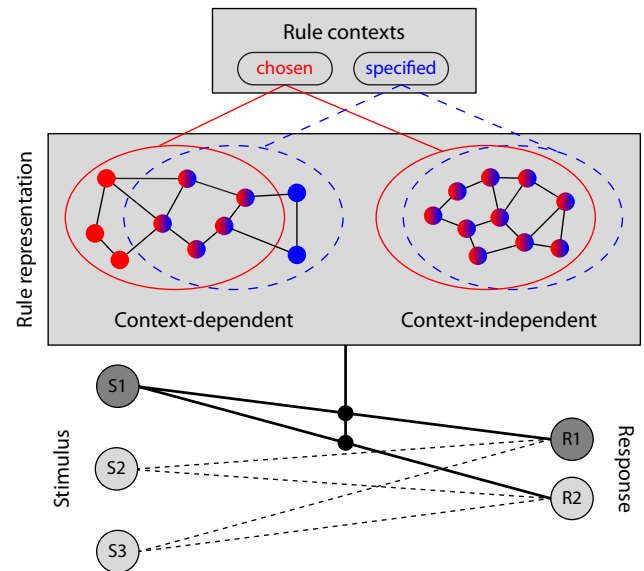


Figure 10. Schematic diagram illustrating context dependency of rule representations, extending the model of cognitive control of Miller and Cohen (2001). Neurons in the frontoparietal cortex may encode rule-specific information when the rule is chosen (red circles), instructed (blue circles), or both (mixed-color circles). Context-independent regions are dominated by neurons that are activated under different contexts, while context-dependent regions also contain neurons that only respond under a particular context. Rule representations provide a regulatory bias of the learned associations between stimuli (S1, S2, and S3) and responses (R1 and R2), necessary to perform a given task. In the diagram, a task rule pertaining to a stimulus dimension of S1 is chosen, and the appropriate response (R1) is evoked under this rule.

primates and human neuroimaging studies suggested that different rules can be differentially and dynamically represented by single neurons or single regions, within a dynamic and adaptive coding scheme (Asaad et al., 2000; Wallis et al., 2001; Wallis and Miller, 2003; Haynes et al., 2007; Soon et al., 2008; Bode and Haynes, 2009; Woolgar et al., 2011a,b).

The current study goes further in demonstrating the contextual influences on neural representations of task rules (Fig. 10). In context-dependent regions, a lack of common rule representations between chosen and specified conditions suggests that the specified and voluntarily selected rules are possibly encoded by separate neural populations even when the rule formally represents the same cognitive process. These subpopulations may be intermixed and thus detectable with multivariate but not univariate methods. Neurons in these regions may be turned to code rule-specific information (i.e., the rule to perform) and/or context-specific information (i.e., more than one rule available to choose from), and hence multivoxel fMRI patterns for task rules measured at a large population level are distinct between different contexts. Context-independent regions, on the other hand, only code information that is relevant to the pending task rules and do so with a similar neuronal ensemble (at least at the limits of MVPA resolution).

We used three task rules that shared the same procedural steps (i.e., two-alternative visual discriminations), but differed in (1) the visual feature to attend, and (2) the corresponding responses to the attended feature. Participants achieved high and similar accuracy in all the three rules and did not have significant switch cost when changing from one rule to another. This suggests that during delay periods, the participants not only maintained information for the visual feature to attend, but also information about associations between corresponding rule-specific categories and responses. Our study is therefore distinct from previous work

that manipulated top-down control of attention but did not change the operation of the tasks (Corbetta et al., 1990; Shulman et al., 1999; Meinhardt and Grabbe, 2002) or manipulated stimulus–response mapping while maintaining top-down control (Bode and Haynes, 2009; Reverberi et al., 2012). The three rules used here could be characterized as attentional sets that define the task-relevant features and associated responses (Corbetta and Shulman, 2002), and fall within the concept of task set (Sakai, 2008). The rule-specific information during the delay periods is more likely to be associated with a reconfiguration process that links visual features and responses, rather than selective attention to visual features alone.

Our study does not in itself rule out the possibility that, after extensive training, some regions might in principle maintain a “pointer” to long-term memory representation of the rules. However, this account is not sufficient to explain all our results. Consistent with our interpretation, the parietal cortex encodes task rules even after the participants only have a few minutes of practice (Woolgar et al., 2011a). In addition, verbal cuing of different tasks sets leads to differential activations in regions that are relevant to the respective spatial, verbal, or semantic nature of the task (Sakai and Passingham, 2003, 2006; Rowe et al., 2007). This cannot be explained if the pointer reflects the instruction rather than the rule itself.

Choosing to perform a rule, compared with specifically instructing the same rule, is associated with transiently increased activations in the frontoparietal cortex, similar to previously reported activations for voluntary selection of task rules (Forstmann et al., 2006; Haynes et al., 2007; Arrington, 2008; Rowe et al., 2008; Bengtsson et al., 2009), voluntary selection of actions (Walton et al., 2004; Karch et al., 2010; Rowe et al., 2010; Thimm et al., 2012; Zhang et al., 2012), and voluntary decisions of whether or not to implement a specific rule (Kühn and Brass, 2009; Kühn et al., 2009). This cortical network is commonly activated with multiple kinds of cognitive demands, including perception, selection, memory, and problem solving (Duncan and Owen, 2000; Duncan, 2001, 2010). In our paradigm, these regions may execute a competition process during voluntary selections (Rowe et al., 2010; Zhang et al., 2012) or monitor potential conflicts when more than one rule is available to choose (Botvinick et al., 2001). While participants voluntarily choose between two rules in the chosen condition, the design of our study discourages several potential selection strategies. First, only two of three possible rules are available in chosen trials, and the chosen trials are further intermixed with specified trials. Recent studies suggest that such a design encourages participants to make a choice in each trial and discourages simply repeating or switching from the last performed rules, because these strategies are not always possible (Lau et al., 2004; van Eimeren et al., 2006; Mueller et al., 2007; Zhang et al., 2012). Second, difficulties of each task rule were presented at individually calibrated level and the rules were not associated with differences in explicit or expected rewards. Therefore, a participant’s choice was not determined by reward-based mechanisms or implicit differences in reward that may arise from different levels of performance of each rule. Third, our data suggested that a participant’s choice was not determined by the cue locations. Although it remains conceivable that some participants use particular strategies in some trials, this would not affect our main research question regarding the contextual modulation of rule representations during maintenance periods.

Two results require further consideration. First, the role of the visual cortex in rule-based visual perceptual decisions remains unclear. We observed significant rule representations in the visual cortex under both chosen and specific contexts. However, the classification accuracies in the visual cortex did not correlate with that from cross-classification analysis, suggesting the rule-specific information in the visual cortex may not be fully context independent. It is possible this rule-specific information is modulated by top-down control signals from frontoparietal cortex (Yantis et al., 2002). Other imaging methods with superior temporal resolution like magneto- or electro-encephalography would be required to explore this hypothesis. Second is the lack of rule encoding in the cACC in both ROI and whole-brain searchlight analysis, contradictory to a previous study on rule representation with arithmetic operations (Haynes et al., 2007). Interestingly, we observed that the rule information can be decoded from the medial superior frontal gyrus, close to the areas identified by other studies of task rules (Haynes et al., 2007; Momennejad and Haynes, 2012). It is possible that representation of rule-specific information is domain dependent. We investigated three stimulus–response mapping rules for perceptual decisions, which may preferentially engage more posterior regions (Koechlin et al., 2003). A recent study on abstract stimulus–response mapping also demonstrated a lack of rule representation in the ACC (Woolgar et al., 2011a). Another possibility is that this difference arose because our participants underwent extensive prescan training, during which the cognitive load during early learning is later reduced (Chein and Schneider, 2005). Future studies using abstract task rules and multiple scan sessions over the course of training (Zhang and Kourtzi, 2010; Zhang et al., 2010) may test these possibilities.

In summary, these findings provide novel insights into the effects of acquisition context on the distributed neuronal representations of task rules. Rule representations in the ventrolateral and parietal cortex are independent of how the rules are established and are invariant from maintenance to execution. In contrast, rule representations in the dorsolateral and medial frontal cortex depend how the rules are first established: by choice or instruction. These results raise an intriguing possibility that context-independent representations enable robust selective attention to task-relevant information, while context-dependent representations enable efficient flexibility to external environment changes. Balancing these mechanisms to achieve one’s goals in a changing environment may be ecologically important, emphasizing the need for adaptive cognitive control.

References

- Arrington CM (2008) The effect of stimulus availability on task choice in voluntary task switching. *Mem Cognit* 36:991–997. [CrossRef Medline](#)
- Asaad WF, Rainer G, Miller EK (2000) Task-specific neural activity in the primate prefrontal cortex. *J Neurophysiol* 84:451–459. [Medline](#)
- Bengtsson SL, Haynes JD, Sakai K, Buckley MJ, Passingham RE (2009) The representation of abstract task rules in the human prefrontal cortex. *Cereb Cortex* 19:1929–1936. [CrossRef Medline](#)
- Bode S, Haynes JD (2009) Decoding sequential stages of task preparation in the human brain. *Neuroimage* 45:606–613. [CrossRef Medline](#)
- Botvinick MM, Braver TS, Barch DM, Carter CS, Cohen JD (2001) Conflict monitoring and cognitive control. *Psychol Rev* 108:624–652. [CrossRef Medline](#)
- Bunge SA, Kahn I, Wallis JD, Miller EK, Wagner AD (2003) Neural circuits subserving the retrieval and maintenance of abstract rules. *J Neurophysiol* 90:3419–3428. [CrossRef Medline](#)
- Chein JM, Schneider W (2005) Neuroimaging studies of practice-related change: fMRI and meta-analytic evidence of a domain-general control

- network for learning. *Brain Res Cogn Brain Res* 25:607–623. [CrossRef Medline](#)
- Corbetta M, Shulman GL (2002) Control of goal-directed and stimulus-driven attention in the brain. *Nat Rev Neurosci* 3:201–215. [CrossRef Medline](#)
- Corbetta M, Miezin FM, Dobmeyer S, Shulman GL, Petersen SE (1990) Attentional modulation of neural processing of shape, color, and velocity in humans. *Science* 248:1556–1559. [CrossRef Medline](#)
- Duncan J (2001) An adaptive coding model of neural function in prefrontal cortex. *Nat Rev Neurosci* 2:820–829. [CrossRef Medline](#)
- Duncan J (2010) The multiple-demand (MD) system of the primate brain: mental programs for intelligent behaviour. *Trends Cogn Sci* 14:172–179. [CrossRef Medline](#)
- Duncan J, Owen AM (2000) Common regions of the human frontal lobe recruited by diverse cognitive demands. *Trends Neurosci* 23:475–483. [CrossRef Medline](#)
- Edgington ES, Onghena P (2007) *Randomization tests*, Ed 4. Boca Raton, FL: Chapman and Hall/CRC.
- Fischl B, Rajendran N, Busa E, Augustinack J, Hinds O, Yeo BT, Mohlberg H, Amunts K, Zilles K (2008) Cortical folding patterns and predicting cytoarchitecture. *Cereb Cortex* 18:1973–1980. [CrossRef Medline](#)
- Forstmann BU, Brass M, Koch I, von Cramon DY (2006) Voluntary selection of task sets revealed by functional magnetic resonance imaging. *J Cogn Neurosci* 18:388–398. [CrossRef Medline](#)
- Haynes JD, Rees G (2005) Predicting the orientation of invisible stimuli from activity in human primary visual cortex. *Nat Neurosci* 8:686–691. [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](#)
- Kamitani Y, Tong F (2005) Decoding the visual and subjective contents of the human brain. *Nat Neurosci* 8:679–685. [CrossRef Medline](#)
- Karch S, Feueracker R, Leicht G, Meindl T, Hantschk I, Kirsch V, Ertl M, Lutz J, Pogarell O, Mulert C (2010) Separating distinct aspects of the voluntary selection between response alternatives: N2- and P3-related BOLD responses. *Neuroimage* 51:356–364. [CrossRef Medline](#)
- Kayser AS, Erickson DT, Buchsbaum BR, D'Esposito M (2010) Neural representations of relevant and irrelevant features in perceptual decision making. *J Neurosci* 30:15778–15789. [CrossRef Medline](#)
- Koechlin E, Ody C, Kouneiher F (2003) The architecture of cognitive control in the human prefrontal cortex. *Science* 302:1181–1185. [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, Simmons WK, Bellgowan PS, Baker CI (2009) Circular analysis in systems neuroscience: the dangers of double dipping. *Nat Neurosci* 12:535–540. [CrossRef Medline](#)
- Kriegeskorte N, Cusack R, Bandettini P (2010) How does an fMRI voxel sample the neuronal activity pattern: compact-kernel or complex spatio-temporal filter? *Neuroimage* 49:1965–1976. [CrossRef Medline](#)
- Kühn S, Brass M (2009) When doing nothing is an option: the neural correlates of deciding whether to act or not. *Neuroimage* 46:1187–1193. [CrossRef Medline](#)
- Kühn S, Haggard P, Brass M (2009) Intentional inhibition: how the “veto-area” exerts control. *Hum Brain Mapp* 30:2834–2843. [CrossRef Medline](#)
- Lau HC, Rogers RD, Haggard P, Passingham RE (2004) Attention to intention. *Science* 303:1208–1210. [CrossRef Medline](#)
- Li S, Ostwald D, Giese M, Kourtzi Z (2007) Flexible coding for categorical decisions in the human brain. *J Neurosci* 27:12321–12330. [CrossRef Medline](#)
- Meinhardt G, Grabbe Y (2002) Attentional control in learning to discriminate bars and gratings. *Exp Brain Res* 142:539–550. [CrossRef Medline](#)
- Miller EK, Cohen JD (2001) An integrative theory of prefrontal cortex function. *Annu Rev Neurosci* 24:167–202. [CrossRef Medline](#)
- Misaki M, Kim Y, Bandettini PA, Kriegeskorte N (2010) Comparison of multivariate classifiers and response normalizations for pattern-information fMRI. *Neuroimage* 53:103–118. [CrossRef Medline](#)
- Momennejad I, Haynes JD (2012) Human anterior prefrontal cortex encodes the “what” and “when” of future intentions. *Neuroimage* 61:139–148. [CrossRef Medline](#)
- Mueller VA, Brass M, Waszak F, Prinz W (2007) The role of the preSMA and the rostral cingulate zone in internally selected actions. *Neuroimage* 37:1354–1361. [CrossRef Medline](#)
- 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. [CrossRef Medline](#)
- Pereira F, Mitchell T, Botvinick M (2009) Machine learning classifiers and fMRI: a tutorial overview. *Neuroimage* 45:S199–S209. [CrossRef Medline](#)
- Preston TJ, Li S, Kourtzi Z, Welchman AE (2008) Multivoxel pattern selectivity for perceptually relevant binocular disparities in the human brain. *J Neurosci* 28:11315–11327. [CrossRef Medline](#)
- Quintana J, Fuster JM (1999) From perception to action: temporal integrative functions of prefrontal and parietal neurons. *Cereb Cortex* 9:213–221. [CrossRef Medline](#)
- Reverberi C, Görgen K, Haynes JD (2012) Compositionality of rule representations in human prefrontal cortex. *Cereb Cortex* 22:1237–1246. [CrossRef Medline](#)
- Rowe JB, Sakai K, Lund TE, Ramsøy T, Christensen MS, Baare WF, Paulson OB, Passingham RE (2007) Is the prefrontal cortex necessary for establishing cognitive sets? *J Neurosci* 27:13303–13310. [CrossRef Medline](#)
- Rowe JB, Hughes L, Nimmo-Smith I (2010) Action selection: a race model for selected and non-selected actions distinguishes the contribution of premotor and prefrontal areas. *Neuroimage* 51:888–896. [CrossRef Medline](#)
- Rowe J, Hughes L, Eckstein D, Owen AM (2008) Rule-selection and action-selection have a shared neuroanatomical basis in the human prefrontal and parietal cortex. *Cereb Cortex* 18:2275–2285. [CrossRef Medline](#)
- Rushworth MF, Walton ME, Kennerley SW, Bannerman DM (2004) Action sets and decisions in the medial frontal cortex. *Trends Cogn Sci* 8:410–417. [CrossRef Medline](#)
- Sakagami M, Tsutsui K (1999) The hierarchical organization of decision making in the primate prefrontal cortex. *Neurosci Res* 34:79–89. [CrossRef Medline](#)
- Sakai K (2008) Task set and prefrontal cortex. *Annu Rev Neurosci* 31:219–245. [CrossRef Medline](#)
- Sakai K, Passingham RE (2003) Prefrontal interactions reflect future task operations. *Nat Neurosci* 6:75–81. [CrossRef Medline](#)
- Sakai K, Passingham RE (2006) Prefrontal set activity predicts rule-specific neural processing during subsequent cognitive performance. *J Neurosci* 26:1211–1218. [CrossRef Medline](#)
- Serences JT, Saproo S, Scolari M, Ho T, Muftuler LT (2009) Estimating the influence of attention on population codes in human visual cortex using voxel-based tuning functions. *Neuroimage* 44:223–231. [CrossRef Medline](#)
- Shulman GL, Ollinger JM, Akbudak E, Conturo TE, Snyder AZ, Petersen SE, Corbetta M (1999) Areas involved in encoding and applying directional expectations to moving objects. *J Neurosci* 19:9480–9496. [Medline](#)
- Soon CS, Brass M, Heinze HJ, Haynes JD (2008) Unconscious determinants of free decisions in the human brain. *Nat Neurosci* 11:543–545. [CrossRef Medline](#)
- Thimm M, Weidner R, Fink GR, Sturm W (2012) Neural mechanisms underlying freedom to choose an object. *Hum Brain Mapp* 33:2686–2693. [CrossRef Medline](#)
- van Eimeren T, Wolbers T, Münchau A, Büchel C, Weiller C, Siebner HR (2006) Implementation of visuospatial cues in response selection. *Neuroimage* 29:286–294. [CrossRef Medline](#)
- Wallis JD, Miller EK (2003) From rule to response: neuronal processes in the premotor and prefrontal cortex. *J Neurophysiol* 90:1790–1806. [CrossRef Medline](#)
- Wallis JD, Anderson KC, Miller EK (2001) Single neurons in prefrontal cortex encode abstract rules. *Nature* 411:953–956. [CrossRef Medline](#)
- Walton ME, Devlin JT, Rushworth MF (2004) Interactions between decision making and performance monitoring within prefrontal cortex. *Nat Neurosci* 7:1259–1265. [CrossRef Medline](#)
- Woolgar A, Thompson R, Bor D, Duncan J (2011a) Multi-voxel coding of stimuli, rules, and responses in human frontoparietal cortex. *Neuroimage* 56:744–752. [CrossRef Medline](#)
- Woolgar A, Hampshire A, Thompson R, Duncan J (2011b) Adaptive coding of task-relevant information in human frontoparietal cortex. *J Neurosci* 31:14592–14599. [CrossRef Medline](#)
- Yantis S, Schwarzbach J, Serences JT, Carlson RL, Steinmetz MA, Pekar JJ, Courtney SM (2002) Transient neural activity in human

- parietal cortex during spatial attention shifts. *Nat Neurosci* 5:995–1002. [CrossRef Medline](#)
- Zhang J, Kourtzi Z (2010) Learning-dependent plasticity with and without training in the human brain. *Proc Natl Acad Sci U S A* 107:13503–13508. [CrossRef Medline](#)
- Zhang J, Meeson A, Welchman AE, Kourtzi Z (2010) Learning alters the tuning of functional magnetic resonance imaging patterns for visual forms. *J Neurosci* 30:14127–14133. [CrossRef Medline](#)
- Zhang J, Hughes LE, Rowe JB (2012) Selection and inhibition mechanisms for human voluntary action decisions. *Neuroimage* 63:392–402. [CrossRef Medline](#)