Compositionality of Rule Representations in Human Prefrontal Cortex

Carlo Reverberi[1,2], Kai Görgen[1] and John-Dylan Haynes[1,3]

[1]Bernstein Centre for Computational Neuroscience, Charité—Universitätsmedizin Berlin, 10115 Berlin, Germany, [2]Department of Psychology, Università Milano—Bicocca, 20126 Milano, Italy and [3]Max Planck Institute for Cognitive and Brain Sciences, 04103 Leipzig, Germany

Address correspondence to Carlo Reverberi, Department of Psychology, Università Milano-Bicocca, 20126 Milano, Italy. Email: carlo.reverberi@unimib.it.

Rules are widely used in everyday life to organize actions and thoughts in accordance with our internal goals. At the simplest level, single rules can be used to link individual sensory stimuli to their appropriate responses. However, most tasks are more complex and require the concurrent application of multiple rules. Experiments on humans and monkeys have shown the involvement of a frontoparietal network in rule representation. Yet, a fundamental issue still needs to be clarified: Is the neural representation of multiple rules compositional, that is, built on the neural representation of their simple constituent rules? Subjects were asked to remember and apply either simple or compound rules. Multivariate decoding analyses were applied to functional magnetic resonance imaging data. Both ventrolateral frontal and lateral parietal cortex were involved in compound representation. Most importantly, we were able to decode the compound rules by training classifiers only on the simple rules they were composed of. This shows that the code used to store rule information in prefrontal cortex is compositional. Compositional coding in rule representation suggests that it might be possible to decode other complex action plans by learning the neural patterns of the known composing elements.

Keywords: control functions, decoding, fMRI, frontal lobe, neural code

Introduction

A major challenge for human cognition is to successfully pursue relevant goals in a continuously changing environment. This implies the ability to select and implement the rules allowing to appropriately deal with the current situation (Passingham 1993; Shallice and Burgess 1996; Wise and Murray 2000; Duncan 2001; Miller and Cohen 2001; Bunge and Wallis 2007). For example, people, at a dinner party, can easily select and implement a simple rule, such as "if Jake prepared pizza, I should not eat it." Humans, however, can also successfully manage the situations in which set of rules more complex than that are required. A common occurrence is the concurrent application of more than one simple rule so that multiple rules need to be combined to form "compound rules." Even in the case of the dinner party, it might be not only important to know that you should avoid Jake's pizza but also that you should definitely try to get Ann's cake twice.

In recent years, several studies have been devoted to the understanding of which neural structures are involved in representation of rules. Neurophysiological studies on monkeys have shown neurons representing active rules in a wide frontoparietal network including prefrontal cortex, orbitofrontal cortex, striatum, and parietal cortex (Hoshi et al. 1998; White and Wise 1999; Asaad et al. 2000; Wallis et al. 2001; Wallis and Miller 2003; Stoet and Snyder 2004; Genovesio et al. 2005;

Muhammad et al. 2006). Neuroimaging studies on humans emphasized the role of both the ventrolateral prefrontal cortex and the superior parietal cortex in rule representation and implementation (Sakai and Passingham 2003, 2006; Bunge et al. 2003; Cavina-Pratesi et al. 2006; Reverberi et al. 2007; Bengtsson et al. 2009; Bode and Haynes 2009; Woolgar et al. 2011; Reverberi et al. 2010).

Despite the evidence already available, it has still remained unclear how the human brain encodes rules consisting of multiple simple rules in working memory. Here, we aimed to investigate how the human brain represents such compound rule sets (Fig. 1). Most importantly, we were interested whether the encoding of rules in the human brain follows the principle of "compositionality": Can the neural code for rules consisting of multiple simple rules be decomposed into the neural representation of the constituent simple rules? For example, is the neural pattern representing a rule composed by 2 simple rules such as 1) "if Jake's pizza then do not eat" and 2) "if Ann's cake eat twice" (the rule set for "parties at Jake's") equal to the combination of the neural patterns of the 2 simple component rules taken alone? Or does the brain use a new independent representation for the compound rules as studies on monkeys seem to suggest (Warden and Miller 2007, 2010; Sigala et al. 2008)? Answering these questions would provide key evidence on how different brain areas encode more complex sets of instructions and on how this impact our overt behavior. Importantly, any evidence for compositionality would provide an explanation for how the brain codes the vast number of different rules required for everyday behavior. On this point, however, it is important to realize that the ways in which 2 or more simple rules can be combined together in a compound rule exceed the one we considered in the present study (e.g., Koechlin et al. 2003; Badre et al. 2009). For example, one might be required to apply the 2 simple rules composing a compound rule in a specific order (first Rule 2 and then Rule 1) or to apply them depending on the status of another, hierarchically higher, layer of rules. Whether the brain uses a compositional code in all these cases was not directly tested in the present study, which represent a first step in that direction. Nevertheless, a principle of parsimony would suggest considering the evidence provided on the use of a compositional code as applicable to other composing rules.

The precise format of neural representations in the human brain has been difficult to address due to limitations in neuroimaging methodology. Most conventional brain imaging techniques analyze activity averaged or smoothed across extended regions of cortex. This has lead to an "activation-based view" of the human brain, where different modules are differentially involved in different processes, independent of the specific contents being processed. In contrast, the
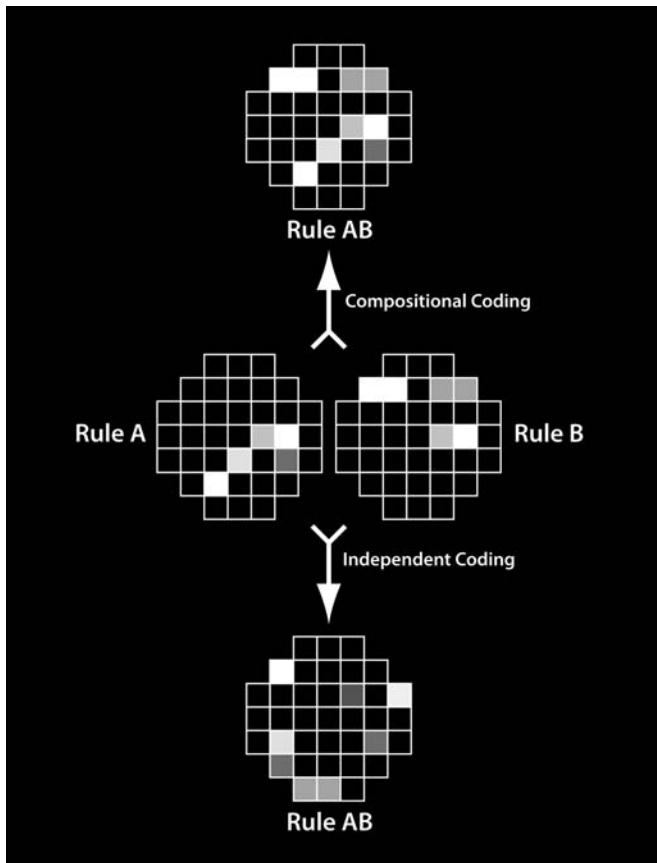
Figure 1. Alternative ways of coding complex rules. In the middle, examples of activity patterns coding for 2 simple rules: Rule A and Rule B. On top and bottom row, 2 alternative coding possibilities for the compound rule formed by Rule A and Rule B. The first possibility (top) is to use a compositional code so that the pattern coding for the Rule AB would be the combination of the patterns for Rule A and Rule B alone. An alternative possibility (bottom) is to use an independent code so that the pattern for Rule AB would be unrelated to the patterns for Rule A and Rule B.

complementary "representational view" of the human brain attempts to identify the nature of neural representations of different types of items (e.g., visual stimuli, rules). Several experiments suggest that most neural representations follow a distributed code that is based on the specific profile of conjoint activation of a larger population of neurons (Schoenbaum and Eichenbaum 1995; van Duuren et al. 2008, 2009; Kahnt et al. 2010). It has recently emerged that functional magnetic resonance imaging (fMRI) can to some degree be used to access information coded in such distributed populations, possibly even when they are encoded at a finer scale than the resolution of the measurement grid (Passingham 1993; Kamitani and Tong 2005; Haynes and Rees 2006; Norman et al. 2006; Swisher et al. 2007; Pereira et al. 2009; Kamitani and Sawahata 2010; Kriegeskorte et al. 2010; Op de Beeck 2010, but see also Op de Beeck 2010). Multivariate pattern recognition has been successfully used to extract information from fine-grained multi-voxel patterns in fMRI data. For example, it has been shown that it is possible to decode low-level visual features (Haynes and Rees 2005; Kamitani and Tong 2005) but also more abstract high-level representations, such as future intentions (Haynes and Rees 2005; Haynes et al. 2007) or concepts (Polyn et al. 2005; Mitchell et al. 2008) from the fine-grained spatial activity patterns in human cortex.

In the present experiment, we used fMRI in combination with multivariate pattern recognition. Specifically, we assessed whether the neural representation of compound rules can be predicted from the neural representation of their simple constituent rules. For this, our paradigm used 2 types of rule sets, "simple" and "compound." The compound rule sets are formed by adding 2 simple rules (Fig. 2). If the brain uses a compositional code in representing compound rule sets, it should be possible to use the relevant simple rules to adjudicate which of 2 alternative compound rules is currently represented.

Materials and Methods

Participants and Experimental Procedure

Thirteen healthy subjects (average age 27 years; 9 males) participated in the experiment. All participants gave written informed consent. The study was approved by the local ethics committee. All participants were right-handed, had normal or corrected to normal vision, no neurological history, and no structural brain abnormalities. Our task required subjects to retrieve, maintain, and then apply a set of conditional rules to a series of target stimuli (Fig. 2). The cues and their associated rules had been learned by the subjects in separate training sessions (see below). The assignment of cues was randomized across subjects, so that each subject had relied on different cue-rule associations. At the beginning of each trial, a cue was presented for 500 ms at the center of the screen. The cue informed the subjects of the rule for the current trial. A delay of variable duration (3500–9500 ms) followed the cue presentation. Finally, 2 target images were presented for 750 ms. Subjects had to apply the current rules to the target stimuli and derive the correct response as fast as possible. Subjects were informed that they had to respond before the targets disappeared. The short time available during target presentation forced the subjects to retrieve the relevant rule set as soon as the cue was presented. Six rule sets belonging to 2 different classes were used (Fig. 2, Supplementary Fig. S1). "Single rules" consisted of a unique conditional rule linking one visual stimulus (house or face) to one specific action (left or right button press). Compound rules consisted of a "pair of single rules" to be applied at the same time. Four simple rules (S1–S4) and 2 compound rules (C1–C2) were used, that is, S1: "If there is a house press left"; S2: "If there is a face press right"; S3: "If there is a face press left"; and S4: "If there is a house press right." The 2 compound rules were C1: S1 + S2, C2: S3 + S4. During scanning, subjects performed 288 trials divided into 6 runs. In each run, 48 trials, 8 per each rule set, were administered in random order. Twelve cues coding for 6 different rules were used, so that 2 visually unrelated cues corresponded to the same rule. As targets, we used gray-scale images of faces (F), houses (H), or unrelated objects (O) presented in pairs (Kriegeskorte et al. 2003). Six target pairs were considered: FO, FH, HO, HF, OF, and OH (the first letter refers to the target image presented on the left of the screen). The distribution of target pairs was equal across rule sets so that in each run, each target pair was presented once for every single rule set and twice for every compound rule set. Given the type of rules and targets allowed in the task, 3 responses could be appropriate: left, right, or no button press. At most one answer had to be produced: In case both left and right button presses were compatible with the target and the active rule, subjects could choose any of the 2.

Training was delivered in 2 separate sessions lasting about 1 h each. In the first session, subjects learnt the correspondence between the 12 cues and the 6 rules. Subjects were first instructed to read and learn a first set of 6 cues and the corresponding rules. Following this, a testing procedure started: subjects were presented with a cue and, after a delay of 3 s, required to "write" the corresponding rule. Writing was performed by clicking with the mouse on the relevant basic components of the target rules, presented on a computer screen along with distractors. The basic components were for example: "If there is a house," or "then press right," or "If there is a face." Feedback was provided in case of incorrect answers. This training phase ended when
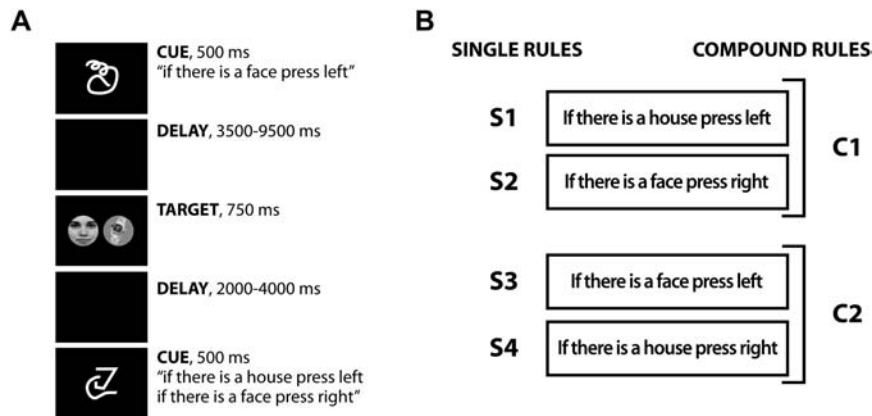
Figure 2. Schema of the experimental paradigm. (*A*) Timeline of a single trial. At the beginning of each trial, a cue was presented. The cue informed the subjects which rule had to be followed in the current trial. A delay of variable duration followed the cue presentation. A blank screen was shown during delay. After the delay, the target image was presented. Subjects had to apply the active rules to the target stimuli and derive the correct response as fast as possible. The short time available during target presentation forced the subjects to retrieve and represent the relevant rule set as soon as the cue was presented. Following target presentation and before the succeeding trial, a blank screen was presented for a variable time. (*B*) Rules used in the experiment. Four single rules (left side, S1–S4) and 2 compound rules (C1–C2) were used. Compound rules are formed by 2 single rules, for example, C1 = S1 + S2.

the subjects reached 100% accuracy over the last 15 trials they performed. After having completed this training phase, a second set of 6 cues was presented for learning. In the following testing procedure, the whole cue set (12 cues) was checked. Again, this training phase ended when the subject reached perfect accuracy over the last 15 trials. In the second session, the training procedure was similar to the procedure used for the fMRI experiment. Two notable differences were present. First, feedback was delivered on incorrect trials. Second, the delay between cue presentation and target presentation was long at the beginning (6.5 s), but it progressively reduced following the improvement of the performance. The training procedure stopped when subject attained 100% accuracy in the last 15 trials with a short delay (1.5 s). The whole training procedure assured that subjects would perform the task during fMRI scanning with high proficiency.

Image Acquisition

Functional imaging was conducted on a 3-T Siemens Trio (Erlangen, Germany) scanner equipped with a 12-channel head coil. In each of the 6 scanning sessions, 305 T_2^*-weighted gradient-echo echo-planar images (EPI) containing 32 slices (3 mm thick) separated by a gap of 0.75 mm were acquired. Imaging parameters were as follows: repetition time 2000 ms, echo time 30 ms, flip angle 90°, matrix size 64 × 64, and a field of view of 192 mm, thus yielding an in-plane voxel resolution of 3 mm², resulting in a voxel size of 3 by 3 by 3.75 mm.

Data Analysis

Functional data were analyzed using SPM8. Preprocessing included rigid-body transformation (realignment) and slice timing. Normalization was applied after completion of the decoding analyses.

Analysis 1: Decoding Visual Cues

A finite impulse response (FIR) model was applied to the realigned and slice-timed images (Henson 2004). Each condition was modeled using 16 time bins of 2 s. Four conditions were modeled in this analysis: They corresponded to the 4 different cues coding for the 2 compound rules. Only correct trials were used for the estimation of the parameters. The time of onset was the presentation of the cue. A decoding analysis using a "searchlight" approach was then applied to the first 4 time bins (Kriegeskorte et al. 2006). In this way, we could estimate in which time bin the information about the cue was first available (Bode and Haynes 2009). The following procedure was repeated for each time bin (Supplementary Fig. S2): A spherical cluster centered on a voxel v_i and with a radius of 4 voxels was defined. The FIR parameters for the voxels included in the sphere v_i were then extracted. They were then transformed into *n*-dimensional pattern vectors for each condition and for each run. Twenty-four vectors (6 runs × 4 conditions) were available for each subject. These vectors represented the spatial response patterns to the 4 conditions from the cluster of voxels v_i. These pattern vectors were divided into 2 sets. Two independent decoding analyses were applied to each vector set. Each set contained 12 vectors, corresponding to 2 of the 4 original conditions. The 12 vectors in one set belonged from trials on which the same compound rule was active: Six vectors corresponded to a cue and 6 vectors corresponded to the other cue introducing the same compound rule. A linear support vector pattern classifier (Muller et al. 2001) with a fixed regularization parameter $C = 1$ was used. In each set, pattern classifiers were trained to distinguish between pattern vectors coding for the 2 alternative cues. A 6-fold cross-validation procedure was implemented to assess the performance of the 2 classifiers. Each classifier was trained on pattern vectors of 5 of 6 fMRI runs to identify patterns corresponding to each cue. The classifier was then tested on the data from the fMRI run not used for training. This procedure was repeated 6 times, with each run acting as the independent test data set once. The average accuracy of all 6 iterations was assigned to the central voxel of the cluster v_i. Classification accuracy significantly above chance implied that the local cluster of voxels spatially encoded cue information. By repeating this procedure for every voxel in the brain, a 3D map of decoding accuracies for each position could be created. The same procedure was independently applied to each time bin. The accuracy maps were normalized using the parameters estimated during preprocessing. The accuracy maps were then submitted to a second-level analysis in SPM8, to test whether the decoding accuracy for voxel differed from chance level (50%) across all subjects at group level. In particular, we subtracted 0.5 to each accuracy map so that the chance level (50%) was represented in the modified maps by zero. The modified accuracy maps for each of the 4 time bins were entered in an analysis of variance (ANOVA) analysis in SPM8. Any significant positive deviation from 0 was interpreted as showing the presence of information in the voxel(s). First, each time bin was tested independently in order to identify in which time bin the information about the cue was first available. The information was already available in occipital cortex starting from the second time bin (Only for exploratory purposes, we performed also a time-resolved analysis for decoding of compound rules. If a relatively liberal statistical threshold is allowed ($P < 0.005$), information is already present in lateral frontal cortex from the second time bin, as it is the case for cue decoding. Presumably, the temporal resolution of our experiment was not sensitive enough to reveal the short delays (about 1 s) between the cue presentation and the rule representation). Such starting point was then used to define the relevant time window for the analyses devoted

to assess where in the brain the information is available. Thus, for all the following analyses, we considered the time window between the second and the fourth time bin, starting from cue presentation. This time window would effectively cover the cue phase and the delay phase. On the contrary, it should effectively discard activity related to target presentation. The target appeared at variable delays from the cue presentation: Only in trials with the shortest delay (one-third of the cases), the target could have been included in the time window we considered. The combined effect of the substantial jitter in delay duration and the presence of target processing in only 1 of 3 trials considered should have effectively prevented the target-related activations to contribute to the estimation of the used FIR time bins. We considered the main effect of accuracy as significantly different from chance at $P < 0.05$, corrected for multiple comparisons at cluster level.

Analysis 2: Decoding Compound Rules

The analysis of compound rules largely followed the first analysis, the main difference consisting on how the same pattern vectors were arranged in the training and testing set. Multivariate pattern classification was used to assess whether information about the compound rules was encoded in the spatial response patterns. A pattern classifier was trained on 2 of 4 conditions. Each of the 2 training conditions corresponded to 1 of the 2 alternative compound rules and associated cues. The classifier was then applied to an independent test data set comprising the 2 conditions not included in the training data set. Each of the 2 conditions of the test data set corresponded to 1 of the 2 alternative compound rules, just as in the training data set. Critically, however, the cues associated with the compound rules in the test data set were different from the cues in the training data set. In this way, it was possible to identify patterns of brain activation encoding "specifically" for the "meaning" of the cue—that is, for the compound rules—and not for the cue visual features. This procedure was repeated 4 times for all the possible combinations of training and testing sets. All relevant time bins (i.e., from the second to the fourth time bin) were considered in the same decoding analysis. This means that all time bins from 2 to 4 belonging from the same condition were considered as equivalent instances of the same class. This is a plausible assumption given that the same representation (i.e., the specific rule active in the trial) needs to be maintained during the whole time window (Bengtsson et al. 2009). As in the preceding analysis, the 4 accuracy maps per subject were submitted to a second-level ANOVA in SPM8, to test whether the decoding accuracy differed from the chance level (50%) at group level.

Analysis 3: Decoding Compound Rules from Simple Rules—Compositionality Analysis

Thirty-six pattern vectors for each time bin were extracted from the FIR parameter estimates corresponding to the 6 conditions (4 single and 2 compound rules) and the 6 fMRI runs. Four related multivariate analyses were run, using 4 different pairs of training data sets. Namely, the training set of the first multivariate analysis comprised pattern vectors belonging to the simple rules (Fig. 2): S1 and S3; in the second multivariate analysis, we used the pattern vectors belonging to S1 and S4; in the third, S2 and S3; and finally in the fourth, S2 and S4. For all 4 multivariate analyses, the pattern vectors of the 2 compound rules were used as test data set. In this way, 4 accuracy maps for each subject could be obtained. These accuracy maps estimated how well the information contained in the vector patterns of single rules allowed to classify pattern vectors of the compound rules. As in the preceding decoding analysis, all relevant time bins were considered in the same decoding analysis. The normalized accuracy maps were finally submitted to a second-level analysis to test whether the decoding accuracy for each voxel significantly differed from chance level (50%).

Results

During scanning, subjects applied the relevant rules to the target with a high accuracy. The average proportion of correct responses was 94% (standard deviation [SD] = 3.3%). Subjects were also very fast in generating the responses: The average reaction time from target appearance was 505 ms (SD = 35 ms). Such short reaction times, combined with the high accuracy, confirm that subjects represented the relevant rule sets during the delay before target presentation. Overall, we performed 3 main sets of decoding analyses, 2 of them exploring the neural representation of cues and rules and the last addressing the neural code of rule representation. Three accessory analyses were also performed to further specify our main findings and to constrain their interpretation.

First, we investigated that brain regions represent the cues themselves irrespectively of their meaning (Fig. 3 and Supplementary Table S1). Multivariate pattern classifiers were trained and tested on instances of cues that had the same meaning. In this way, only the visual features specific to each individual cue would be the source of a successful classification. We were able to decode cue identity from the left inferior occipital gyrus (Brodmann Area [BA] 18, $P < 0.05$ corrected, peak accuracy 60%) and from the right middle occipital gyrus (BA 19, $P < 0.05$
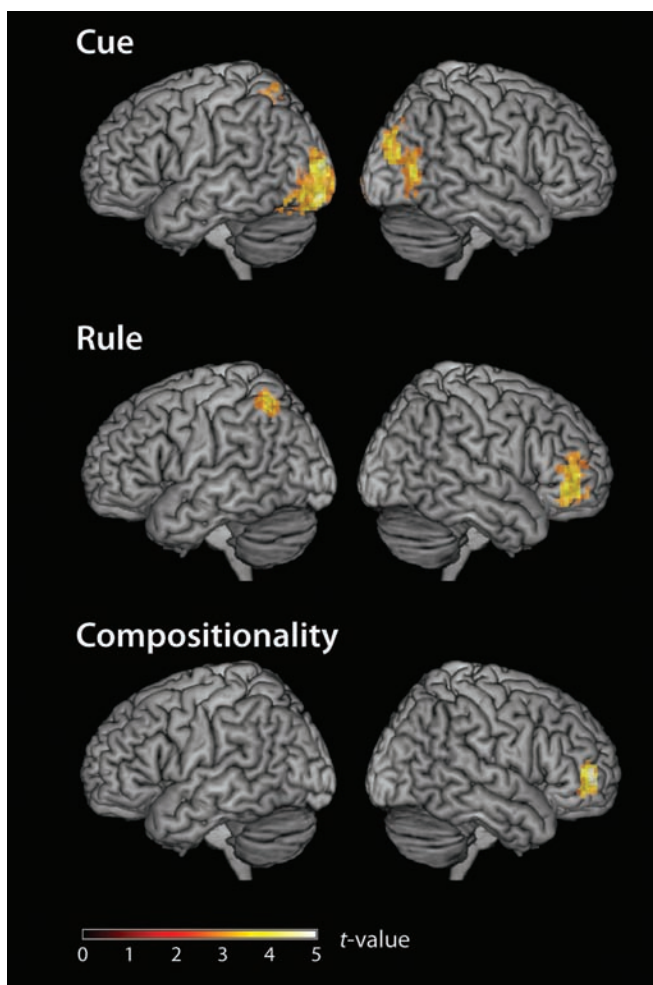

Figure 3. Regions where it was possible to decode from local patterns of activation (*A*) which cue was presented and (*B*) which compound rule was active in a specific trial. (*C*) Compositionality analysis: Regions in which it was possible to decode which compound rule was active by only relying on information extracted from the local patterns of activation associated with single rules. Informative brain regions are rendered onto a T_1-weighted brain image. The reported effects are all significant at $P < 0.05$, corrected for multiple comparisons. The color scale (*t* values) is reported on the bottom left.

corrected, peak accuracy 60%). Furthermore, it was possible to decode cue identity from a region in the left superior parietal lobe (BA 7, $P < 0.05$ corrected, peak accuracy 60%).

Next, we investigated where in the brain the active compound rule is represented. To identify brain areas specifically coding for the rule, it is necessary to control for other correlated dimensions of both the stimuli and the task. First of all, it is important to disentangle between the cue (e.g., its visual features) and its meaning (i.e., the rule). Second, it is necessary to disentangle the representation of the rule linking between stimulus and action from the representation of stimuli and actions alone. The experimental paradigm and the implemented decoding analysis were devised to comply with the above requirements. First, each of the rules could be instructed by 1 of 2 visually different cues. In this way, we could train the pattern classifiers on rule instances that were introduced by 1 of the 2 cues with the same meaning, and then test on other instances of the same rules, but introduced by the alternative different cue (Supplementary Figs. S1 and S2). Second, for this analysis, we considered compound rules only. Either compound rule is constituted by the same basic elements. In both compound rules right and left index presses, houses and faces are involved. The only feature changing between the 2 compound rules is the rule linking the basic elements (Fig. 2). This also assures that the same cognitive resources (e.g., working memory) are needed to represent the 2 compound rules. We were able to decode the identity of the active compound rule from local activation patterns in the left parietal lobe (BA 7/40, $P < 0.05$ corrected, peak accuracy 59) and in the right inferior frontal lobe (mainly BA 47, $P < 0.05$ corrected, peak accuracy 57%). There was a partial overlap of the brain areas representing the superficial features of the cues and the brain areas representing the rules in the left parietal lobe, with the area related to rule representation peaking more laterally than the area related to cue representation (Supplementary Fig. S3). In contrast, cue identity information could not be detected in the right lateral frontal cortex even allowing very liberal statistical threshold ($P < 0.01$, uncorrected for multiple comparisons).

The preceding analyses showed that areas in the right frontal and left parietal represent compound rules. However, which code do those brain areas use to represent compound rules? Do they use a "compositional code" (Fig. 1) so that the neural coding of a conditional rule such as "if there is a face press left" is the same when the rule is represented alone (simple rule) and when the same rule is part of a wider rule set (compound rule)? We trained multivariate pattern classifiers on pairs of simple rules, each of which was always part of a different compound rule. The performance of the classifiers trained on simple rules was then tested on the activation patterns of the 2 compound rules. Decoding of compound rules by information extracted from single rules was only possible in the right lateral inferior frontal cortex (mainly BA 47 and 46, $P < 0.05$ corrected, peak accuracy 56%). This is the same lateral frontal area that was shown to be informative in the preceding "independent" analysis on rule decoding. The overlap between the 2 is large: namely, 60 of a total of 104 voxels in the compositionality analysis overlap with those in the rule analysis (Supplementary Fig. S3). By contrast, it was not possible to detect rule information with the compositionality procedure in the left parietal region even allowing very liberal statistical threshold ($P < 0.01$, uncorrected).

If a compositional code is used in right lateral prefrontal cortex for representing single and compound rules, it should be possible not only to decode compound rules from single rules but also to decode single rules from activity patterns estimated on compound rules. We refer to this analysis as the "inverse compositionality analysis" to distinguish it from the main compositionality analysis described above. Convergence of findings between the indirect and the main compositionality analysis would further corroborate the claim of compositionality of the neural code. We trained a multivariate pattern classifier on the pair of compound rules available (i.e., C1 vs. C2). The classifier was then tested on the activation patterns of the 4 relevant pairs of single rules, namely S1 versus S3, S1 versus S4, S2 versus S3, and S2 versus S4. Decoding of single rules by information extracted from compound rules was possible in the right lateral inferior frontal cortex ($P < 0.05$ corrected, peak accuracy 55%, Fig. 4). This is the same region found in the main analyses on compositionality and on compound rule decoding. No brain region showed a difference in accuracy when the indirect and the main compositionality decoding performance were compared.

To further assess the relative performance of the 3 main decoding analyses considered in this study (cue, rule and compositionality), we ran a "region of interest (ROI) analysis" (Fig. 5). We considered as ROI the 2 brain areas, inferior right frontal and left parietal, which were involved in rule representation (Fig. 3). We then computed, per each subject, the average accuracy attained by the 3 classifiers across all voxels in the ROI. Those values were then submitted to a one-sample t-test analysis. Notice that the selection criteria of the ROI, based uniquely on the rule representation analysis, are independent from the compositionality and cue representation analyses. Thus, no risks of circular analyses are introduced when only the latter are considered. The findings in the ROI analysis confirm the whole-
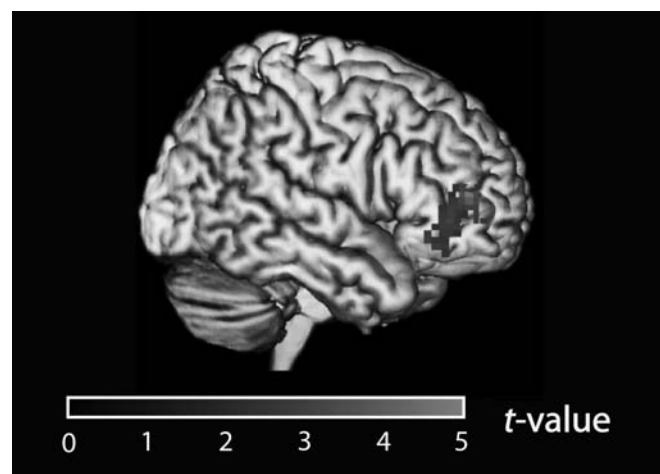

Figure 4. Inverse compositionality analysis: Regions in which it was possible to decode which simple rule was active by only relying on information extracted from the local patterns of activation associated with compound rules. We trained a multivariate pattern classifier on the pair of compound rules available (i.e., C1 vs. C2). The performance of the classifier trained on compound rules was then tested on the activation patterns of the 4 relevant pairs of single rules, namely S1 versus S3, S1 versus S4, S2 versus S3, and S2 versus S4. Informative brain regions are rendered onto a T_1-weighted brain image. The reported effects are all significant at $P < 0.05$, corrected for multiple comparisons. The color scale (t-values) is reported on the bottom.

brain analyses (Fig. 5). In particular, information on cue identity is represented in the parietal lobe but not in the frontal lobe (parietal: $t_{12} = 1.95$, $P = 0.038$; frontal: $t_{12} = -0.948$, $P = 0.82$). The opposite is true for the compositionality analysis (parietal: $t_{12} = 0.47$, $P = 0.32$; frontal: $t_{12} = 3.29$, $P = 0.003$). The interaction information-type (cue vs. composition) × brain area (parietal vs. frontal) was significant (repeated measures ANOVA: $F_{1,12} = 7.24$, $P = 0.02$), meaning that the information represented in the 2 areas was different.

In the main compositionality analysis, we showed that it was possible to decode compound rules by information extracted from single rules in the right lateral inferior frontal cortex. We interpreted this finding as showing that the right lateral inferior frontal cortex uses a compositional code to represent compound rules. However, an alternative hypothesis could explain this finding. During training, the subjects might have created strong associative links between the simple rules and the compound rules in which these are embedded, so that whenever a simple rule is represented, the associated compound rule is automatically retrieved and represented as well. For example, when subjects are required to represent the simple rule 1 (i.e., S1, see Fig. 2), they not only do that but they also automatically represent the related compound rule 1 (C1). This would explain our findings without assuming the compositionality of the neural code. Decoding in right lateral inferior frontal cortex, which is also involved in compound rule representation, would be possible because subjects represent compound rules even when they are only required to represent simple rules.

We devised an accessory analysis to assess this hypothesis. We chose an approach that makes opposite predictions depending on whether the decoding uses information about simple rules or about associated compound rules. To understand this, take for example the following classifier:

TRAINING: S1 (house, left) versus S4 (house, right).

When applying this classifier to the remaining 2 simple rules:

TEST: S3 (face, left) versus S2 (face, right).

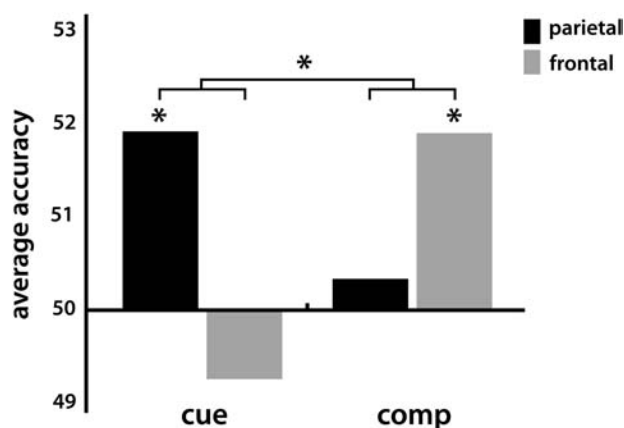Two different outcomes are possible:

1. The classifier learns a fact about the 2 simple rules that is the different motor response between S1 and S4 and classifies S3 and S2 according to their motor responses. In that case, it would classify S3 like S1 and S2 like S4.
2. The classifier learns the compound rules associated with each of the simple rules (C1 for S1 and C2 for S4) and classifies S3 and S2 according to the compound rules associated with them. In that case it would classify S3 like S4 and S2 like S1.

Please note that the 2 predictions are exactly opposite. Given how we implemented the decoding analysis, if the classifier learns information on simple rules, this would lead to a performance significantly higher than chance (i.e., >50%). By contrast, if the classifier learns information on the associated compound rules, then the performance should be significantly lower than chance (i.e., <50%). There are in total 4 different combinations of simple rules that can be tested this way (train S1–S4, test S3–S2; train S3–S2, test S1–S4; train S1–S3, test S2–S4; train S2–S4, test S1–S3). The Figure 6 shows the average performance of the above-mentioned analyses. No brain region showed a performance of the classifiers reliably lower than chance. On the contrary, we found higher than chance accuracy in several brain regions, including the right lateral prefrontal region, which we already observed in the compositionality analysis. Overall, this analysis does not support the hypothesis that the compositionality analysis reflects the automatic representation of compound rules associated to simple rules.

Discussion

In this study, we used fMRI in combination with multivariate pattern recognition to understand where and, most importantly, how the brain represents and organizes "multiple rules" active at the same time in working memory. Specifically, we were interested whether the encoding of rules in the human brain follows the principle of compositionality (Fig. 1). We found that both lateral parietal and inferior lateral frontal cortex are involved in representing compound rules. However, the 2 areas use a different neural code to represent the same information: while lateral parietal cortex codes simple and compound rules by using an independent coding, inferolateral prefrontal cortex uses a compositional code.
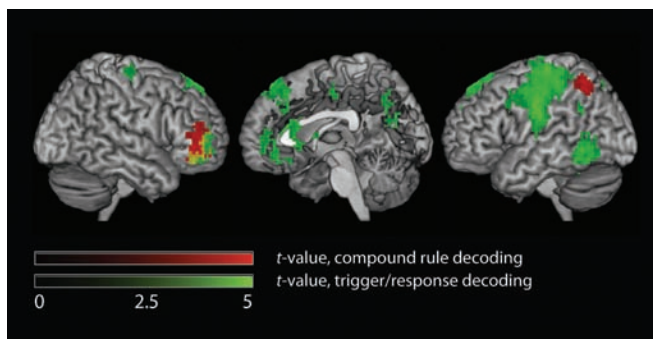

Figure 5. Average decoding performance in the 2 brain areas carrying information on compound rules (Fig. 3). The bars represent the average decoding accuracy in the 2 decoding analyses independent from compound rule decoding: the cue decoding ("cue") and the compositionality analysis ("comp"). Please notice that the plots represent the average accuracy over all voxels belonging to the ROI. Because of this, the average accuracy level is expected to be lower than the reported peak accuracy. *$P < 0.05$ Cue: cue decoding. Comp: compositionality analysis. Black bars: parietal ROI. Gray bars: frontal ROI.

Figure 6. Assessment of the automatic association hypothesis. Green: Regions where it was possible to decode information on triggers (i.e., house vs. faces) and motor responses (i.e., right vs. left button presses) involved in simple rules. No brain region showed a systematically wrong performance (see text for explanation). Red: Findings of the compositionality analysis. Regions in which it was possible to decode which compound rule was active by only relying on information extracted from the local patterns of activation associated with single rules.

Representation of Compound Rules: Location

Two brain areas were found to store the information on the currently active compound rules: the inferolateral prefrontal cortex and the lateral parietal cortex. The information on compound rules could not be attributed to irrelevant low-level features of the stimuli or to different cognitive resources (e.g., working memory) involved for representation. Instead, the information was specifically related to the core meaning of the rule: the representation of the link between a triggering condition (e.g., seeing a house) to its consequences (e.g., pressing the left button).

Overall, our findings are compatible with and further specify the growing evidence on the representation of task sets and rules. Studies on monkeys have shown that, besides other brain structures, both prefrontal cortex and lateral parietal cortex are involved in the representation of the current task (Asaad et al. 2000; Hoshi et al. 2000; Bussey et al. 2001; Wallis et al. 2001; Wallis and Miller 2003; Stoet and Snyder 2004, 2009; Genovesio et al. 2005; Buckley et al. 2009; White and Wise 1999). Neuroimaging evidence on humans confirmed that both ventrolateral prefrontal cortex and lateral parietal cortex were active while maintaining rules in working memory (Bunge et al. 2003; Sakai and Passingham 2003, 2006; Rowe et al. 2007; Bengtsson et al. 2009), particularly when rules as complex as compound rules are considered (Bunge and Zelazo 2006). Compared with evidence on the monkey brain, our study showed that the brain areas involved in the temporary representation of rules are more confined in humans. This might be because we did not compare the representations of tasks, which involve different cognitive operations (e.g., a spatial task vs. an object matching task). Instead, our study focused on the representation of 2 alternative compound rules involving exactly the same cognitive operations (house and face recognition, left and right button presses, and rule processing). The only difference between the examined conditions is how the same cognitive operations are organized or connected by arbitrary rules. Previous neuroimaging studies are compatible with the findings presented here (Bunge et al. 2003; Sakai and Passingham 2003, 2006; Bunge and Zelazo 2006; Rowe et al. 2007; Bengtsson et al. 2009). Our study goes beyond this previous work by examining whether brain regions involved in rule processing encode information about the "specific" rules that are currently active. A recent study (Bode and Haynes 2009) filled this gap by applying multivariate decoding techniques. However, in that study, it was not possible to determine whether the information about the active task was due to the representation of the cue or the rules with which the cues were associated. Here, we solved the ambiguity by showing that both the lateral parietal and ventrolateral frontal areas encode information on the active rule that cannot be explained by the concurrent encoding of the visual features of the associated cues.

Our findings also test one of the major tenets of representation theories on prefrontal lobe function (Dehaene et al. 1998; Duncan 2001; Miller and Cohen 2001). These theories hold that prefrontal cortex acts as a temporary store only for those facts that are "relevant" for the course of thinking or action, in this way biasing the activity of lower level brain areas. Following this hypothesis, prefrontal cortex should represent which rule is active but not which specific cue was presented, that is, decoding should be at chance for low-level visual features. This is indeed the pattern we found in inferior lateral frontal cortex. This is in agreement with a previous study using multivariate analysis techniques applied to a visual categorization task showing that lateral prefrontal cortex flexibly changes the information represented following task needs (Li et al. 2007). Furthermore, it has been claimed that also lateral parietal cortex is part of an extended network of brain areas supporting the temporary storage of relevant information (Dehaene et al. 1998; Duncan 2001; Assad 2003). Interestingly, however, the lateral parietal cortex stored both relevant (the rule) and irrelevant (the cue identity) information. Is parietal cortex a nonspecific store bound to maintain any kind of information, relevant or not? Not necessarily. Cue identity is actually relevant for an accessory, but critical, task: the assignment of the meaning to an identified cue. One might speculate that the parietal cortex is not involved in the main task but only in the enabling accessory task of translating the visual cues into their meaning. This would explain why both cues and rules are represented in parietal cortex. Furthermore, it may also explain why the peaks in the accuracy maps for cue and rule decoding are not fully overlapping (Fig. 3), with the former being more medial than the latter. Overall, this hypothesis is consistent with other studies (Bunge et al. 2002; Brass and von Cramon 2004; Cavina-Pratesi et al. 2006; Gail and Andersen 2006; Bode and Haynes 2009).

Representation of Compound Rules: Compositionality

Once we established where the brain represents the active compound rules, our further goal was to understand how the brain encodes them. A straightforward prediction of the compositionality hypothesis is that it should be possible to adjudicate which compound rule a subject is representing by only relying on the knowledge of the neural patterns associated with each composing simple rule. We found that this was indeed possible in the right ventrolateral frontal cortex, the very same area in which we found information on compound rules (Fig. 3, Supplementary Fig. S3).

These findings are interesting under several respects. First, by generating results very similar to our previous independent analyses, they replicate the finding that the right lateral frontal cortex is involved in compound rule representation. Second, they show that the right ventrolateral frontal cortex uses a compositional code to temporarily store multiple rules. Thus, the same code is used in this brain area to represent simple rules irrespective of the fact that they are represented alone or as part of a more complex rule. This conclusion was also confirmed by the assessment of a possible alternative interpretation, the "automatic association hypothesis," which could be discarded (Fig. 6).

To date, limited evidence is available on how representations of multiple items are organized in working memory. In a recent paper, Li et al. (2007) applied multivariate decoding to simple and "complex" rules, similarly to the present experiment. The simple rule was categorization of a test movie based on spatial similarity to the prototype, and the complex rule was categorization based on spatial and temporal similarity to the prototype. They showed that prefrontal cortex is involved in the representation of both rule types. Since they employed only one simple rule and one complex rule, they could not explore the relation between the neural representations of specific contents of the 2 rule types. Most neurophysiological studies

focused on the preliminary question of how neurons represent individual memory items, so that only one study explored an issue related to ours (Warden and Miller 2007, 2010). Warden and Miller found evidence for independent coding in monkey prefrontal cortex. They had monkeys remember 2 objects sequentially presented. Two delays were introduced between the first and the second object and after the second object. They found that the spiking activity of single neurons after the presentation of the first object did not relate in a straightforward way to the spiking activity of the same single neurons after the presentation of the second object. Thus, at least for a simple short-term memory task, the monkey brain seems to use an independent code in lateral prefrontal cortex: The neural code of exactly the same object changes depending on the context (first vs. second delay). Further evidence is needed to establish whether this is a real cross species difference, whether it is related to the different scale explored by the 2 techniques, or whether it is related to differences in the specific task at hand. In favor of the last interpretation, we may notice that the representation of the first object has been explored in 2 different task phases (first and second delay). The use of an independent code in this case would be consistent with recent evidence showing that different task phases of the same trial are associated with approximately orthogonal patterns of activity across the population of neurons in the lateral prefrontal cortex (Sigala et al. 2008). Conversely, in our task, we tested the "same" task phase belonging to different trials. A further element to consider is that in our task, the use of a compositional code is beneficial for performance. A compositional code not only allows the brain to temporarily store the multiple rules involved in the active rule set but also to keep track of the similarity relations of the currently active rules with the other rules used in the task. This in turn allows exploiting the expertise acquired in managing the simple rules to boost the performance of the compound rules.

Please note that, as shown by the cross-classification procedure between different cue sets, the rule representations we identified are cue invariant and thus have a certain degree of abstraction. However, cue invariance, in our study, does not imply the use of a compositional code. One area could use a unique code to represent a specific compound rule, irrespectively to the triggering cue, and still it could use a noncompositional code. This seems to be the case of the lateral parietal area.

In previous fMRI studies, it has been proven difficult to functionally dissociate lateral prefrontal and parietal cortices (e.g., Brass and von Cramon 2002, 2004; Crone et al. 2006; but see also Montojo and Courtney 2008 for previous positive evidence; Rowe et al. 2008; Stoet and Snyder 2009). In this study, the use of a compositional code was observed in the prefrontal but not in the parietal cortex. This difference across the 2 brain areas was also formally confirmed by the ROI analysis. This corroborates our hypothesis about their distinct, even though strongly related functional roles. We argued that the parietal cortex is mainly involved in translating the cues into their specific meaning. If this were the case, then the "parietal task" could be effectively redescribed as the application of a different rule set compared with the main (frontal) task. This "accessory" set would map each cue to the appropriate rule, such as, for example, "If cue 1, then rule 3." Critically, in this accessory rule set, no whole-part relation is present. None of the accessory rules can be derived from the combination of any other since each of them contains a different cue. The representation of such rules, not being compositional at the conceptual level, should not be compositional at the neural level as well, even in brain areas that use compositional coding as a general organizing principle.

As mentioned in the introduction, the ways in which 2 or more simple rules can be combined together in a more complex rule set exceed the one we considered in the present study. Further studies are thus needed to directly test whether a compositional code is used also in those cases, as it could be extrapolated from the presented evidence. Most interestingly, further studies are warranted to understand how the brain encodes the different ways of combining the same set of simple rules.

Conclusions

Overall, the findings on cue decoding, rule decoding, and compositionality analyses suggest that the lateral parietal cortex is involved in translating the cues into their meaning, that is, the currently active rules. Only the latter information is then made available to the inferolateral frontal cortex, which will maintain it during delay, and supposedly, will be also involved in its final implementation (Buckley et al. 2009).

Our study shows for the first time that the code used to temporarily store rule information in lateral prefrontal cortex (but not in parietal cortex) is compositional. Further evidence is needed to understand in which other contexts, for which material and in which brain areas this type of code is used. The definition of the code used by a brain area in relation to a specific task will also represent a further powerful tool to better characterize the functional role of brain regions, even when they are part of highly integrated and interconnected networks. Finally, our findings also address an important question related to "mental state decoding" (Haynes and Rees 2006). It suggests that it might be possible to decode other complex compound thoughts by learning the neural patterns of the known composing elements.

Supplementary Material

Supplementary material can be found at: http://www.cercor. oxfordjournals.org/.

Funding

Notes

References

Assad JA. 2003. Neural coding of behavioral relevance in parietal cortex. Curr Opin Neurobiol. 13(2):194–197.

Asaad WF, Rainer G, Miller EK. 2000. Task-specific neural activity in the primate prefrontal cortex. J Neurophysiol. 84(1):451–459.

Badre D, Hoffman J, Cooney JW, D'Esposito M. 2009. Hierarchical cognitive control deficits following damage to the human frontal lobe. Nat Neurosci. 12(4):515-522.

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(8):1929-1936.

Bode S, Haynes JD. 2009. Decoding sequential stages of task preparation in the human brain. Neuroimage. 45(2):606-613.

Brass M, von Cramon DY. 2002. The role of the frontal cortex in task preparation. Cereb Cortex. 12(9):908-914.

Brass M, von Cramon DY. 2004. Decomposing components of task preparation with functional magnetic resonance imaging. J Cogn Neurosci. 16(4):609-620.

Buckley MJ, Mansouri FA, Hoda H, Mahboubi M, Browning PG, Kwok SC, Phillips A, Tanaka K. 2009. Dissociable components of rule-guided behavior depend on distinct medial and prefrontal regions. Science. 325(5936):52-58.

Bunge SA, Hazeltine E, Scanlon MD, Rosen AC, Gabrieli JD. 2002. Dissociable contributions of prefrontal and parietal cortices to response selection. Neuroimage. 17(3):1562-1571.

Bunge SA, Kahn I, Wallis JD, Miller EK, Wagner AD. 2003. Neural circuits subserving the retrieval and maintenance of abstract rules. J Neurophysiol. 90(5):3419-3428.

Bunge SA, Wallis JD. 2007. Neuroscience of rule-guided behavior. New York: Oxford University Press.

Bunge SA, Zelazo PD. 2006. A brain-based account of the development of rule use in childhood. Curr Dir Psychol Sci. 15(3):118-121.

Bussey TJ, Wise SP, Murray EA. 2001. The role of ventral and orbital prefrontal cortex in conditional visuomotor learning and strategy use in rhesus monkeys (Macaca mulatta). Behav Neurosci. 115(5):971-982.

Cavina-Pratesi C, Valyear KF, Culham JC, Kohler S, Obhi SS, Marzi CA, Goodale MA. 2006. Dissociating arbitrary stimulus-response mapping from movement planning during preparatory period: evidence from event-related functional magnetic resonance imaging. J Neurosci. 26(10):2704-2713.

Crone EA, Wendelken C, Donohue SE, Bunge SA. 2006. Neural evidence for dissociable components of task-switching. Cereb Cortex. 16(4):475-486.

Dehaene S, Kerszberg M, Changeux JP. 1998. A neuronal model of a global workspace in effortful cognitive tasks. Proc Natl Acad Sci U S A. 95(24):14529-14534.

Duncan J. 2001. An adaptive coding model of neural function in prefrontal cortex. Nat Rev Neurosci. 2(11):820-829.

Gail A, Andersen RA. 2006. Neural dynamics in monkey parietal reach region reflect context-specific sensorimotor transformations. J Neurosci. 26(37):9376-9384.

Genovesio A, Brasted PJ, Mitz AR, Wise SP. 2005. Prefrontal cortex activity related to abstract response strategies. Neuron. 47(2):307-320.

Haynes JD, Rees G. 2005. Predicting the orientation of invisible stimuli from activity in human primary visual cortex. Nat Neurosci. 8(5):686-691.

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

Haynes JD, Sakai K, Rees G, Gilbert S, Frith C, Passingham RE. 2007. Reading hidden intentions in the human brain. Curr Biol. 17(4):323-328.

Henson R. 2004. Analysis of fMRI time series. In: Frackowiak RSJ, Friston KJ, Frith CD, Dolan RJ, Price CJ, Zeki S, Ashburner JT, Penny WD, editors. Human brain function. 2nd ed. London: Academic Press. p. 793-822.

Hoshi E, Shima K, Tanji J. 1998. Task-dependent selectivity of movement-related neuronal activity in the primate prefrontal cortex. J Neurophysiol. 80(6):3392-3397.

Hoshi E, Shima K, Tanji J. 2000. Neuronal activity in the primate prefrontal cortex in the process of motor selection based on two behavioral rules. J Neurophysiol. 83(4):2355-2373.

Kahnt T, Heinzle J, Park SQ, Haynes JD. 2010. The neural code of reward anticipation in human orbitofrontal cortex. Proc Natl Acad Sci U S A. 107(13):6010-6015.

Kamitani Y, Sawahata Y. 2010. Spatial smoothing hurts localization but not information: pitfalls for brain mappers. Neuroimage. 49(3):1949-1952.

Kamitani Y, Tong F. 2005. Decoding the visual and subjective contents of the human brain. Nat Neurosci. 8(5):679-685.

Koechlin E, Ody C, Kouneiher F. 2003. The architecture of cognitive control in the human prefrontal cortex. Science. 302(5648):1181-1185.

Kriegeskorte N, Cusack R, Bandettini P. 2010. How does an fMRI voxel sample the neuronal activity pattern: compact-kernel or complex spatiotemporal filter? Neuroimage. 49(3):1965-1976.

Kriegeskorte N, Goebel R, Bandettini P. 2006. Information-based functional brain mapping. Proc Natl Acad Sci U S A. 103(10):3863-3868.

Kriegeskorte N, Sorger B, Naumer M, Schwarzbach J, van den Boogert E, Hussy W, Goebel R. 2003. Human cortical object recognition from a visual motion flowfield. J Neurosci. 23(4):1451-1463.

Li S, Ostwald D, Giese M, Kourtzi Z. 2007. Flexible coding for categorical decisions in the human brain. J Neurosci. 27(45):12321-12330.

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

Mitchell TM, Shinkareva SV, Carlson A, Chang K-M, Malave VL, Mason RA, Just MA. 2008. Predicting human brain activity associated with the meanings of nouns. Science (New York, NY). 320(5880):1191-1195.

Montojo CA, Courtney SM. 2008. Differential neural activation for updating rule versus stimulus information in working memory. Neuron. 59(1):173-182.

Muhammad R, Wallis JD, Miller EK. 2006. A comparison of abstract rules in the prefrontal cortex, premotor cortex, inferior temporal cortex, and striatum. J Cogn Neurosci. 18(6):974-989.

Muller KR, Mika S, Ratsch G, Tsuda K, Scholkopf B. 2001. An introduction to kernel-based learning algorithms. IEEE Trans Neural Netw. 12(2):181-201.

Norman KA, Polyn SM, Detre GJ, Haxby JV. 2006. Beyond mind-reading: multi-voxel pattern analysis of fMRI data. Trends Cogn Sci. 10(9):424-430.

Op de Beeck HP. 2010. Probing the mysterious underpinnings of multi-voxel fMRI analyses. Neuroimage. 50(2):567-571.

Passingham RE. 1993. Frontal lobes and voluntary action. New York: Oxford University Press, USA.

Pereira F, Mitchell T, Botvinick M. 2009. Machine learning classifiers and fMRI: a tutorial overview. Neuroimage. 45(Suppl 1):S199-S209.

Polyn SM, Natu VS, Cohen JD, Norman KA. 2005. Category-specific cortical activity precedes retrieval during memory search. Science. 310(5756):1963-1966.

Reverberi C, Cherubini P, Frackowiak RS, Caltagirone C, Paulesu E, Macaluso E. 2010. Conditional and syllogistic deductive tasks dissociate functionally during premise integration. Hum Brain Mapp. 31(9):1430-1445.

Reverberi C, Cherubini P, Rapisarda A, Rigamonti E, Caltagirone C, Frackowiak RS, Macaluso E, Paulesu E. 2007. Neural basis of generation of conclusions in elementary deduction. Neuroimage. 38(4):752-762.

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(10):2275-2285.

Rowe JB, Sakai K, Lund TE, Ramsoy T, Christensen MS, Baare WF, Paulson OB, Passingham RE. 2007. Is the prefrontal cortex necessary for establishing cognitive sets? J Neurosci. 27(48):13303-13310.

Sakai K, Passingham RE. 2003. Prefrontal interactions reflect future task operations. Nat Neurosci. 6(1):75-81.

Sakai K, Passingham RE. 2006. Prefrontal set activity predicts rule-specific neural processing during subsequent cognitive performance. J Neurosci. 26(4):1211-1218.

Schoenbaum G, Eichenbaum H. 1995. Information coding in the rodent prefrontal cortex. II. Ensemble activity in orbitofrontal cortex. J Neurophysiol. 74(2):751-762.

Shallice T, Burgess P. 1996. The domain of supervisory processes and temporal organization of behaviour. Philos Trans R Soc Lond B Biol Sci. 351(1346):1405-1411; discussion 1411-1412.

Sigala N, Kusunoki M, Nimmo-Smith I, Gaffan D, Duncan J. 2008. Hierarchical coding for sequential task events in the monkey prefrontal cortex. Proc Natl Acad Sci U S A. 105(33):11969-11974.

Stoet G, Snyder LH. 2004. Single neurons in posterior parietal cortex of monkeys encode cognitive set. Neuron. 42(6):1003-1012.

Stoet G, Snyder LH. 2009. Neural correlates of executive control functions in the monkey. Trends Cogn Sci. 13(5):228-234.

Swisher JD, Halko MA, Merabet LB, McMains SA, Somers DC. 2007. Visual topography of human intraparietal sulcus. J Neurosci. 27(20):5326-5337.

van Duuren E, Lankelma J, Pennartz CM. 2008. Population coding of reward magnitude in the orbitofrontal cortex of the rat. J Neurosci. 28(34):8590-8603.

van Duuren E, van der Plasse G, Lankelma J, Joosten RN, Feenstra MG, Pennartz CM. 2009. Single-cell and population coding of expected reward probability in the orbitofrontal cortex of the rat. J Neurosci. 29(28):8965-8976.

Wallis JD, Anderson KC, Miller EK. 2001. Single neurons in prefrontal cortex encode abstract rules. Nature. 411(6840):953-956.

Wallis JD, Miller EK. 2003. From rule to response: neuronal processes in the premotor and prefrontal cortex. J Neurophysiol. 90(3): 1790-1806.

Warden MR, Miller EK. 2007. The representation of multiple objects in prefrontal neuronal delay activity. Cereb Cortex. 17(Suppl 1): i41-i50.

Warden MR, Miller EK. 2010. Task-dependent changes in short-term memory in the prefrontal cortex. J Neurosci. 30(47):15801-15810.

White IM, Wise SP. 1999. Rule-dependent neuronal activity in the prefrontal cortex. Exp Brain Res. 126(3):315-335.

Wise SP, Murray EA. 2000. Arbitrary associations between antecedents and actions. Trends Neurosci. 23(6):271-276.

Woolgar A, Thompson R, Bor D, Duncan J. 2011. Multi-voxel coding of stimuli, rules, and responses in human frontoparietal cortex. Neuroimage. 56(2):744-752.