OXFORD

Cerebral Cortex, 2016, 1–15

doi: 10.1093/cercor/bhw032
Original Article

ORIGINAL ARTICLE

Neural Coding for Instruction-Based Task Sets in Human Frontoparietal and Visual Cortex

Paul S. Muhle-Karbe[1,2], John Duncan[3], Wouter De Baene[1,4], Daniel J. Mitchell[3] and Marcel Brass[1]

[1]Department of Experimental Psychology, Ghent University, Gent, Belgium, [2]Center for Cognitive Neuroscience, Duke University, Durham, USA, [3]MRC Cognition and Brain Sciences Unit, Cambridge University, Cambridge, UK and [4]Department of Cognitive Neuropsychology, Tilburg University, Tilburg, The Netherlands

Address correspondence to Paul S. Muhle-Karbe, Department of Experimental Psychology, Ghent University, Henri Dunantlaan 2, B-9000 Ghent, Belgium. Email: paul.muhlekarbe@gmail.com

Abstract

Task preparation has traditionally been thought to rely upon persistent representations of instructions that permit their execution after delays. Accumulating evidence suggests, however, that accurate retention of task knowledge can be insufficient for successful performance. Here, we hypothesized that instructed facts would be organized into a task set; a temporary coding scheme that proactively tunes sensorimotor pathways according to instructions to enable highly efficient "reflex-like" performance. We devised a paradigm requiring either implementation or memorization of novel stimulus–response mapping instructions, and used multivoxel pattern analysis of neuroimaging data to compare neural coding of instructions during the pretarget phase. Although participants could retain instructions under both demands, we observed striking differences in their representation. To-be-memorized instructions could only be decoded from mid-occipital and posterior parietal cortices, consistent with previous work on visual short-term memory storage. In contrast, to-be-implemented instructions could also be decoded from frontoparietal "multiple-demand" regions, and dedicated visual areas, implicated in processing instructed stimuli. Neural specificity in the latter moreover correlated with performance speed only when instructions were prepared, likely reflecting the preconfiguration of instructed decision circuits. Together, these data illuminate how the brain proactively optimizes performance, and help dissociate neural mechanisms supporting task control and short-term memory storage.

Key words: cognitive control, frontoparietal cortex, MVPA, task preparation, visual cortex, working memory

Introduction

A distinctive aspect of adaptive human cognition concerns the ability to rapidly transform symbolic instructions into novel goal-directed behaviors (Duncan et al. 1996, 2008; Cole, Laurent, et al. 2013). Whether you are visiting a foreign city for the first time, building a piece of furniture, or learning how to use a new application on your computer, your first steps will likely be guided by instructions that specify how to act in order to achieve a given goal. Yet, although learning from instruction is ubiquitous in daily life and unique in its efficiency, little is known about the precise mechanisms that give rise to this capacity (Wenke and Frensch 2005; Wenke et al. 2007; Meiran et al. 2014).

Extant evidence suggests that the lateral prefrontal cortex (LPFC) is critically involved in such flexible control of novel behaviors. Lesions of the LPFC can perturb the ability to learn novel tasks while leaving pre-learned behaviors unaffected (Luria 1966; Walsh 1978; Fuster 1980; Petrides 1985; Duncan 1986; Duncan et al. 1997). Likewise, in functional magnetic resonance imaging (fMRI) studies the LPFC becomes active when novel task instructions are given, but rapidly disengages after only few applications (e.g., Ruge and Wolfensteller 2010; Dumontheil et al. 2011; Hartstra et al. 2011). In line with popular "dual system" views, these findings suggest that LPFC is involved in the initial construction of task representations for the first controlled

applications, followed by the gradual buildup of more pragmatic representations in premotor-basal-ganglia loops that establish behavioral routines (Ramamoorthy and Verguts 2012; Wolfensteller and Ruge 2012). However, despite this well-documented general importance of the LPFC in the assembly of novel task representations, the precise nature of its contribution remains controversial.

Classic accounts of LPFC function have emphasized the role of persistent neuronal firing in the active maintenance of task-relevant information (Fuster 2001; Miller and Cohen 2001). Unlike sensory brain areas that typically exhibit transient bursts of activity time-locked to the instruction cue, the LPFC can remain active throughout the entire delay separating the cue from a contingent target stimulus (Fuster and Alexander 1971; Kubota and Niki 1971; Fuster 1973; Cohen et al. 1997; Courtney et al. 1997). Such sustained activity is often specific to a particular type of task-relevant information such as stimuli, locations, rewards, and even abstract task rules (Watanabe 1996; Asaad et al. 1998; Rainer et al. 1998; Wallis et al. 2001). Accordingly, the LPFC has been thought to actively maintain a stable representation of the task cue, thereby making it accessible for later selection of task-appropriate behavior when it is no longer available in the environment (Miller et al. 1996; Miller and Cohen 2001).

Several lines of evidence, however, suggest a more complex picture. A recent neurophysiological study has revealed that context-dependent activity patterns in LPFC neurons evolve dynamically within the delay phase and are virtually orthogonal to the initial representation of the task cue (Stokes et al. 2013). In parallel, behavioral studies in humans have shown that accurate retention of task knowledge can be insufficient for behavioral control during performance. As noted above, frontal lobe patients often have difficulties in obeying novel task instructions. Strikingly, however, the very same patients can often properly recall the demands of a task, despite their inability to perform it (Milner 1963; Konow and Pibram 1970; Duncan et al. 1996). Such cases of "goal neglect" have also been documented in the normal population (Duncan et al. 1996, 2008, 2012; Dumontheil et al. 2011; Bhandari and Duncan 2014). Further studies have shown that active preparation, but not memorization, of task instructions creates vulnerability to response priming when the instructed stimuli are encountered in a nested secondary task (Cohen-Kdoshay and Meiran 2009; Liefooghe et al. 2012, 2013; Meiran et al. 2012, 2014). In conjunction, these findings indicate that task readiness constitutes a specific cognitive state that is distinct from the mere maintenance of instructed task demands.

In the present study, we wanted to compare the neural underpinnings of these 2 states. Based on the aforementioned findings, we hypothesized that preparation would elicit the formation of a task set in the sense of a "prepared reflex," that is, a temporary configuration that tunes sensorimotor systems toward goal-dependent processing of anticipated input. Such a configuration may proactively optimize performance by binding perceptual and motor codes into a compound representation, so that encountering the instructed stimulus will automatically trigger the associated response. In contrast, memorization of task instructions may be achieved by more persistent activation of an initial semantic instruction representation. To address this question, we designed a paradigm that required participants either to implement newly instructed stimulus–response (SR) mappings or to memorize them for a later recognition test. An initial behavioral study confirmed that only prepared, yet not memorized, stimuli are capable of priming the associated responses in a secondary task (see Supplementary material). Subsequently, we used multivoxel pattern analysis (MVPA) of fMRI data to reveal the neural representation of to-be-implemented and to-be-memorized instructions in the human brain. MVPA assesses the neural coding of particular task parameters by identifying fine-grained local activity patterns that permit their discrimination (Haynes and Rees 2006; Norman et al. 2006). This methodology has been successfully applied in studying the neural representation of both task rules (Bode and Haynes 2009) and working memory contents (Serences et al. 2009), and in revealing changes in the coding strength of such variables as a function of the task context (Woolgar, Hampshire, et al. 2011; Waskom et al. 2014, 2016). Here, we used MVPA to compare the neural basis of semantic task knowledge and behaviorally effective task sets.

Materials and Methods

Participants

Twenty-seven right-handed and neurologically healthy subjects (11 female) participated in the fMRI experiment. One participant was removed from analysis due to excessive head motion during the scanning, exceeding voxel size. Furthermore, 3 participants had to terminate the experiment prematurely, as they were unable to perform the tasks above chance level. Thus, the final sample contained 23 participants (10 female; mean age = 24.09; SD = 5.06). Approval for the study was obtained from the local ethics committee, and all participants gave written informed consent prior to participation.

Experimental Rationale

Our experimental task was motivated by recent behavioral studies that have combined a primary SR mapping task with a nested secondary task (see Supplementary material). These studies have demonstrated that the instructed SR mapping modulates performance in the secondary task (i.e., target stimuli prime their associated responses), but only when it is actively prepared and not when it is memorized for recognition or recall (Liefooghe et al. 2012, 2013). Here, instead of employing a secondary task, we sought to measure brain activity during the pretarget phase (i.e., the delay between task instruction and execution) to probe the differential nature of the cognitive states that are established under preparation and memorization demands. Previous studies comparing task preparation with short-term memory storage have used univariate analysis techniques and observed very similar brain activity under both types of demand, especially within the LPFC (e.g., Ikkai and Curtis, 2011; Hartstra et al. 2011). However, an equivalent increase in the average signal intensity of this region mainly indicates a comparable engagement of cognitive control, while leaving it open whether or not the underlying functional states are similar. In the present study, we therefore utilized MVPA as a means to quantify the neural coding of instruction information during the pretarget phase. This permitted us to compare the strength of representations between implementation and memorization conditions, to establish links between such coding properties and subsequent performance, and to compare the similarity and stability of the neural code in which task instructions are represented throughout the pretarget phase.

To be able to measure the coding of instruction information via classification analyses, we included 2 types of SR mapping instructions that involved different types of stimuli, namely faces and houses. The rationale for this procedure was twofold: First, by presenting novel items on each trial, this procedure permitted us to avoid repetitions of SR mappings, while using 2 constant

and separable categories of instructions. Ensuring instruction novelty was critical, as previous studies have shown that the neural mechanisms supporting task control change substantially with only few repetitions (Cole et al. 2010; Ruge and Wolfensteller 2010; 2013; Hartstra et al. 2011; Stocco et al. 2012). Second, the use of distinct stimulus categories was also intended to provide the classification with a strong signal that would allow us to measure a modulatory effect of pretarget demands on instruction coding.

Apparatus and Stimulus Material

MR images were acquired using a 3T Trio MR scanner (Siemens Medical Systems, Erlangen, Germany) with a standard 32-channel radio-frequency head coil. Stimulus presentation was controlled via Presentation software (Neurobehavioral Systems Inc., Albany, CA), and the stimulus displays were projected on a screen, located at a distance of approximately 120 cm (frame rate = 60 Hz). All stimuli were presented on a gray background (RGB = 100, 100, 100). The stimulus material comprised images of faces and houses as well as text captions. A total of 252 face images and 252 house images were collected by accessing publicly available databases and by downloading further images form the Internet. These images were converted to gray scale and displayed at a scale of 150 × 200 pixels (equivalent to 3.77 × 5.03° visual angle). Text captions were displayed in white font (font type = Arial, font size = 22, equivalent to 4.49 × 0.65°).

Task Design

Our task was comprised of 2 different block types that required implementation and memorization of SR mapping instructions respectively. In implementation blocks, trials began with the presentation of an instruction screen in which either 2 faces or 2 houses were presented along with the letters "+ WIJS" and "+ MIDDEL" (Dutch for "+ index" and "+ middle"; see Fig. 1A for an illustration). This procedure mapped one of the images to the bimanual index fingers, and the other image to the bimanual middle fingers. The position of each set of fingers on the instructions screen (i.e., upper vs. lower half) was counterbalanced across trials. Instruction screens were presented for a total of 2000 ms. Pilot studies had revealed that this duration is necessary to ensure effective stimulus encoding. Moreover, we opted for bimanual rather than lateralized responses because we wanted to avoid participants employing imagery strategies wherein they imagine one image on the left side and another on the right side (see Hartstra et al. 2011 for similar reasoning). The instruction phase was followed by a delay interval, during which a fixation point (diameter = 0.24°; color = white) was presented centrally on the screen. The duration of the delay was jittered to permit separation of delay-related brain activity from the preceding instruction phase and from the subsequent target phase. Durations were varied between 800 and 5200 ms in steps of 400 ms following a distribution of pseudologarithmic density (mean duration = 2467 ms). Afterward, in the target phase, one

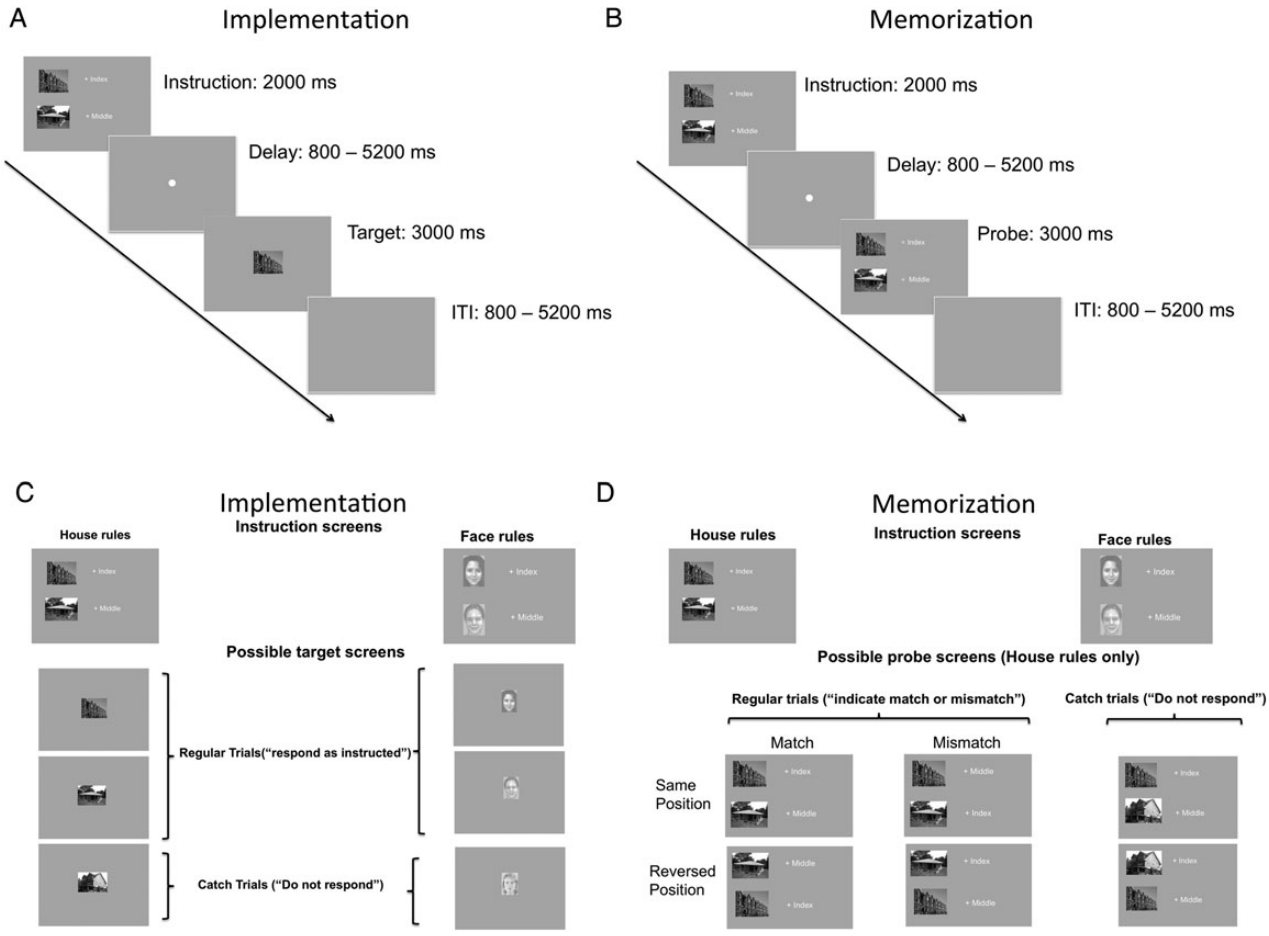

Figure 1. The upper panel illustrates the single-trial structure in implementation blocks (*A*) and memorization blocks (*B*). The lower panel illustrates the possible target configurations in implementation blocks (*C*) and memorization blocks (*D*).

of the 2 instructed images was presented centrally for 3000 ms requiring the response dictated by the preceding instruction. Occasionally, the target image was a novel image from the same stimulus category that was not presented on the preceding instruction screen (see Fig. 1C). These "catch trials" required response omissions and were included to encourage participants to encode both SR mappings and not only one of them. The target phase was followed by an intertrial interval (ITI) of variable length in which a blank screen was presented. The length of the ITI was jittered randomly, based on the same distribution as the delay phase.

Trials in memorization blocks followed the same structure (see Fig. 1B). The only difference was the design of the target phase. Here, a second instruction screen was presented (probe screen) and participants were required to indicate whether or not it conveyed the same SR mapping as the initial instruction screen. Probe screens were designed in the same manner as the instruction screen, except that the letters defining the response set were accompanied by a "=" sign rather than by a "+" sign. This change was implemented to ensure that participants could always distinguish instruction screens from probe screens. Half of the probe screens displayed the same SR mapping as the instruction screen (matches), and the other half displayed the reversed mapping (mismatches). Furthermore, we also varied the item position on the probe screen, independently of the probe validity. That is, the position of the 2 images (top vs. bottom) on the probe screen could be either the same, compared with the instruction screen, or reversed. In both cases, the probe could be either a match or a mismatch (see Fig. 1D for an illustration). We included this manipulation of the item position to en sure that participants relied on the information conveyed by the SR mapping instruction rather than on a mere visual image thereof (i.e., participants could maintain a visual image of the instruction screen and any deviance from this template on the probe screen would be taken as evidence for a mismatch). Matches and mismatches were indicated with bimanual index fingers and middle fingers (mapping counterbalanced across participants). Moreover, as in implementation blocks, catch trials were included. In these trials, one of the 2 items on the probe screen was replaced by a novel uninstructed image, requiring response omissions. In half of these trials the upper images was replaced, and in the other half the lower image.

Overall, participants performed five runs of each block type. Each run contained 22 regular trials and 2 catch trials. For both regular trials and catch trials, half of the instructions displayed faces and the other half houses. Images were randomly drawn from the whole set of images and combined into a unique set of instructions for each participant. Every image was presented only once throughout the entire experiment. The only constraints were that face instructions always displayed 2 images from the same gender, and that distractors on catch trials were taken from the same gender as the 2 instructed images. Based on pilot studies, we included this constraint to match face instructions and house instructions in terms of accuracy and reaction time (RT). Moreover, it should prevent participants from encoding face items merely by virtue of their gender. Trial transitions were random with the constraints that repetitions and alternations of the instruction category occurred equally often in each run, and that all relevant trial elements—that is, probe validity (match vs. mismatch), target position on the instruction screen (top vs. bottom), distractor position on the probe screen (top vs. bottom)—were not repeated more than 2 times. Finally, based on pilot studies, we suspected that frequent alternation between implementation and memorization demands might introduce transfer effects, for example, by encouraging participants to utilize one general strategy. Therefore, participants first completed five runs of one block type and then five runs of the other (sequence counterbalanced across participants).

Functional Localizer

After completing the experimental runs, participants performed a localizer task designed to identify functionally dedicated visual brain regions that preferentially process face and house stimuli. This task comprised 3 types of mini-blocks, in which a series of 16 images were rapidly presented (duration = 800 ms; interstimulus interval = 200 ms). Images were either faces, or houses, or scrambled images (i.e., random permutations of pixels from both image sets). Participants were required to attend to the image stream and to indicate stimulus repetitions via a button press, which occurred 2 times in each mini-block. The localizer task consisted of four runs, each of which contained all 3 different mini-blocks (sequence counterbalanced across participants). Within runs, mini-blocks were separated by 16-s fixation periods.

Scanning Procedure

After participants were placed headfirst and supine in the scanner bore, a high-resolution anatomical image was acquired using a T_1-weighted 3D MPRAGE sequence (time repetition [TR] = 1550 ms, time echo [TE] = 2.39 ms, time to inversion = 900 ms, acquisition matrix = 256 × 256 × 176, sagittal field of view (FOV) = 220 mm, flip angle = 9°, voxel size resized to 1 × 1 × 1 mm). Functional images during the experimental tasks and the localizer tasks were acquired using a T_2^*-weighted echo planar imaging (EPI) sequence, sensitive to BOLD contrast (TR = 2000 ms, TE = 35 ms, image matrix = 64 × 64, FOV = 224 mm, flip angle = 80°, slice thickness = 3 mm, distance factor = 17%, voxel size resized to 3.5 × 3.5 × 3.5 mm^3, 30 axial slices). Overall the scanning time lasted about 60 min per subject.

Behavioral Data Analysis

RT and accuracy of regular trials were analyzed by means of general linear models (GLMs) with the factors BLOCK TYPE (implementation vs. memorization) and INSTRUCTION CATEGORY (face vs. house; see Supplementary Material for an additional analysis using linear mixed effects models). In memorization blocks, performance was averaged across match and mismatch trials. Error trials and trials subsequent to those were discarded from the RT analysis. Moreover, the accuracy of catch trials was analyzed in a separate GLM of analogous design. We also computed inverse efficiency scores (IES) by dividing each participant's mean RT of a design cell by the percentage of accurate responses (see Townsend and Asby 1983). This score serves the integration of speed and accuracy into a single index, and was used in some analyses to rule out the possibility of speed-accuracy trade-offs.

fMRI Data Analysis

Data preprocessing was performed using SPM 8 software (Wellcome Department of Cognitive Neurology, London, UK) and data visualization was performed using caret software (Van Essen 2005). The first four volumes of each run were excluded to allow for T_1 relaxation. The remaining volumes were realigned to their mean image and corrected for differences in slice-time acquisition. Each participant's anatomical image was coregistered with the mean functional image, and normalized to the template brain provided by the Montreal Neurological Institute (MNI).

Transformational parameters of the anatomical images were then applied to the EPI images, and motion parameters were estimated, separately for each run. The time series data at each voxel were processed using a high-pass filter with a cut-off of 128 s to remove low-frequency artifacts. For univariate analyses, data were smoothed using an 8 mm full-width half-maximum (FWHM) Gaussian kernel. Multivariate decoding analyses were performed on normalized but unsmoothed data.

Statistical analyses of the experimental task were performed separately for implementation and memorization blocks. Time series data were modeled based on a series of events. We defined six vectors based on the different trial phases (i.e., instruction phase, delay, and target phase), and instruction categories (i.e., face instructions and house instructions). An additional vector of no-interest was defined that contained all trial phases of error trials, and the target phases of catch trials. The durations of all events (from start to finish) were convolved with a haemodynamic response function and entered into the regression model, which contained additional regressors to account for variance related to head motion. For univariate analyses, the model included another regressor for the ITI, which was used as a low-level baseline for preparation-related brain activity. The localizer task was modeled with separate regressors for the different types of mini-blocks (i.e., faces, houses, scrambled images, and fixation), and regressors to account for head motion. In all models, statistical parameter estimates were computed separately for all columns in the design matrix.

Definition of Regions of interest

Prior to performing whole-brain analyses, we selected a number of candidate regions that we expected to be involved in the assembly of novel task representations. As noted above, previous studies have primarily emphasized the importance of the LPFC, though the exact location of foci has varied across investigations and may depend upon the type of instruction that is given. Specifically, processing of instructions that convey specific SR mappings has typically been associated with caudal LPFC sections in the vicinity of the inferior frontal junction area (IFJ; Ruge and Wolfensteller 2010, 2013; Hartstra et al. 2011), whereas more abstract and/or complex task instructions might be processed in more rostral sections along the inferior frontal sulcus (IFS; Cole et al. 2010; Cole, Laurent, et al. 2013). Beyond the LPFC, novel instructions typically also yield activity in the parietal lobe in and around the intraparietal sulcus (IPS; Ruge and Wolfensteller 2010; Dumontheil; et al. 2011; Stocco et al. 2012). Coactivation of LPFC and IPS has been documented across a variety of paradigms that require flexible instantiation, updating, or reinforcement of task parameters (Cole and Schneider 2007; Duncan 2010; Cole, Reynolds, et al. 2013). Accordingly, these regions are likely candidates for the construction of task sets based on instruction. We defined 3 sets of bilateral ROIs covering the IFJ, the IFS, and the IPS (see Fig. 2A for an illustration). Spherical ROIs pertaining to the IFJ were centered at the peak coordinates of a meta-analysis on frontal lobe contributions to task control (Derrfuss et al. 2005; radius = 10 mm; size = 515 voxels; left IFJ: −40, 4, 30; right IFJ: 44, 10, 34). Regions of interest (ROIs) pertaining to the IFS and the IPS were based on a study by Fedorenko et al. (2013; IFS: size = 2666 voxels; center of gravity: ±38, 39; 23; IPS: size = 8520 voxels; center of gravity = ±29; −56; 46). In this study, an average statistical t-map was computed across 6 different types of cognitive demand (spatial working memory, verbal working memory, arithmetic, 2 types of multi-source interference, Stroop interference) that were all revealed by contrasting a difficult task condition with an easier baseline condition. The full unthresholded map and the parcellation into its components is available for download at http://imaging.mrc-cbu.cam.ac.uk/imaging/MDsystem.

In addition to frontoparietal regions, we were also interested in the representation of instruction information within functionally dedicated visual regions. Specifically, the fusiform face area (FFA) and the parahippocampal place area (PPA) have been documented as relatively specialized processors of faces and scenes (including houses), respectively (Kanwisher et al. 1997; Epstein and Kanwisher 1998; Epstein et al. 1999; Kanwisher and Yovel 2006). Recent MVPA studies from the working memory literature moreover indicate that such feature- or category-selective regions may contain "high-fidelity" representations of visual memoranda during delay intervals (Sreenivasan, Curtis, et al. 2014). Such representations appear to reflect the sensory features of memoranda (Harrison and Tong 2009; Lee et al. 2013; Sreenivasan, Vytlacil, et al. 2014), persist throughout the delay even if average activity returns to baseline (Serences et al. 2009), and are tied with the precision of WM performance (Emrich et al. 2013; Ester et al. 2013). We therefore assumed that, in our task, pattern separation in the FFA and the PPA during the delay would index the precision or vividness of target representation. Spherical ROIs (radius = 10 mm, size = 515 voxels) pertaining to these regions were centered at the peak coordinates resulting from the localizer task (left FFA: −42, −55, −20; right FFA: 42, −49, −14; left PPA: −27, −43, −8; right PPA: 30, −58, −8; see Fig. 2B). To reveal category-selective regions, activity

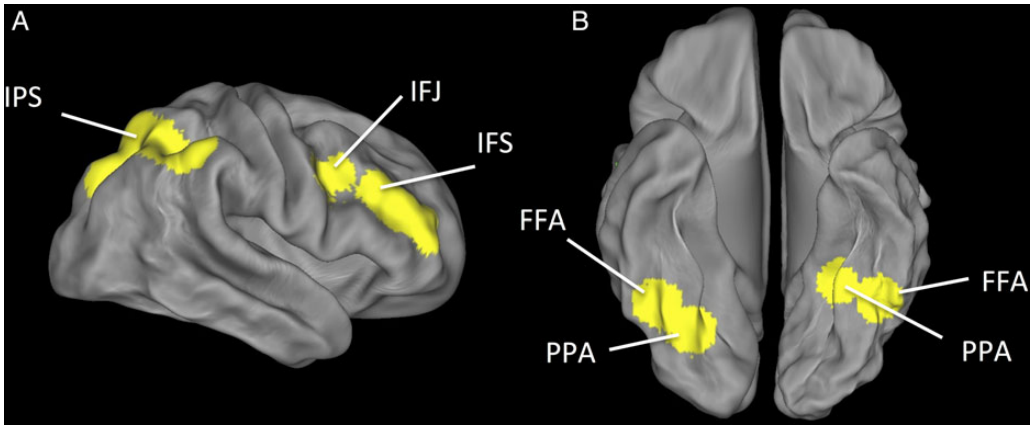

Figure 2. Illustration of the ROI within frontoparietal cortex (*A*) and functionally defined dedicated visual areas (*B*).

during face blocks and house blocks was separately contrasted with fixation at the first level. Thereafter, at the second-level, fixation-corrected activity from both block types were contrasted with each other ($P < 0.001$, uncorrected for multiple comparisons).

Multivariate fMRI Analyses

Multivariate decoding analyses were performed with the PyMVPA toolbox (Hanke et al. 2009). We employed a searchlight analysis to reveal local activity patterns that carry information about the instruction category (Kriegeskorte; et al. 2006) using a spherical searchlight with a radius of 3 mm. Normalized but unsmoothed β images were subjected to the analysis and a linear support vector machine (SVM; cost parameter $C = 1$) was used for classification. Our aim was to reveal the neural representation of instruction categories during the pretarget phase. To this end, we trained the SVM to distinguish between face instructions and house instructions within each block type. These analyses were performed separately for the β images pertaining to the instruction phase and for those images pertaining to the delay. In each analysis, we used a leave-one-run-out cross-validation procedure. That is, the classifier was trained on the data of four runs and subsequently tested on its accuracy at classifying the data of the remaining run. This process was repeated five times using all possible combinations of training and test data. Classification accuracies were averaged across all five iterations, yielding a mean decoding accuracy map for each participant. Prior to the second-level analyses, decoding maps were smoothed with an 8 mm FWHM kernel.

MVPA Within ROIs

We first extracted decoding accuracies within voxels of the predefined ROIs for each design cell. These scores were averaged across left and right hemisphere and analyzed in a GLM with the factors TRIAL PHASE (instruction phase vs. delay), BLOCK TYPE (implementation vs. memorization), and ROI (IFJ vs. IFS vs. IPS vs. FFA vs. PPA). We assumed that initial processing stages would be largely similar under implementation and memorization demands, and that differences in instruction coding should be strongest toward the end of the pretarget phase. Specifically, we reasoned that the instruction phase primarily reflects the perceptual analysis of the instruction screen and encoding processes, so we did not expect substantial differences between block types during this trial phase (though it is conceivable that encoding may already differ between block types). The delay, on the other hand, should capture the distinct states of task readiness and maintenance, and thus differentiate most clearly between block types. Given the absence of visual stimulation in the delay, decoding results from this trial phase should also yield a more precise index of internal task representation than the preceding instruction phase, which includes perceptual differences between face and house stimuli.

Whole-brain Decoding Analysis

We also explored instruction coding outside of the ROIs by contrasting the respective whole-brain decoding maps. Again, analyses were conducted separately for the instruction phase and for the delay. First, the decoding map of each block type was contrasted with chance level of accuracy (50%) to reveal significant coding of instruction category in implementation and memorization blocks. Thereafter, the 2 maps were contrasted with one another via paired-samples *t*-tests to reveal differences in the coding strength between the 2 block types. For these analyses, we used a peak threshold of $P < 0.001$ (uncorrected for multiple comparisons) in combination with a cluster threshold of 22 contiguous voxels.

Univariate fMRI Analyses

To identify further candidate regions for the decoding analyses that may also contribute to preparatory adjustments, but are not detected in the rather conservative whole-brain searchlight analysis, we contrasted pre-target brain activity with activity during the ITI. This was done separately for the instruction phase and for the delay. Significance of these comparisons was established using a family-wise error corrected cluster threshold of $P < 0.01$, and a minimum cluster extent of 100 contiguous voxels. Note that the ITI should be considered a "low-level baseline," as participants were not engaging in task performance during this trial phase. Accordingly, contrasting pretarget activity against the ITI provides a sensitive but not a specific index of preparatory control processes.

Internal Validation

An important complication in the interpretation of decoding accuracy is that this measure only considers the discriminability of experimental conditions and is agnostic toward the source and directionality of these differences (see Todd et al. 2013 for extensive discussion). This renders decoding data more vulnerable to confounds than univariate analyses, because such factors may contaminate a classifier's discrimination success even if they are unsystematic in their directionality across individuals and cancel each other out at the group level. In the context of task rule decoding in frontoparietal cortex, the most salient confound is difficulty. Frontoparietal regions are sensitive to task difficulty across a variety of cognitive domains (Fedorenko et al. 2012, 2013). Hence, successful decoding of task variables within these regions, in the presence of difficulty confounds, may reflect differential amounts of effort or attention to the task, rather than genuine neural coding of the respective task variable. To examine whether and to what extent the decoding of instruction categories might reflect differential task difficulty, we employed 2 types of validation analyses. First, within each participant, we compared RT between face and house instructions using one-way analysis of variances (ANOVAs). This analysis should identify participants who displayed different performance for face instructions and house instructions. Second, across participants, we correlated performance differences between face instructions and house instructions (recoded as absolute values) and the decoding accuracies within frontoparietal ROIs. This analysis should reveal if better pattern separation in these regions relies on differential performance between instruction categories. Note, however, that our study examined brain activity prior to task execution. Accordingly, these validation analyses are concerned with the role of "anticipated" task difficulty rather than actual performance demands. This is also why we chose not to include RT as an additional factor in our regression model (see Waskom et al. 2014; Woolgar et al. 2014), as it is unclear to what extent preparatory BOLD signal translates into RT on a trial-by-trial basis.

External Validation

In the next set of analyses, we sought to probe the relevance of the decoding results for task performance, and to further delineate the contributions of the different ROIs in establishing task readiness.

To reiterate, we expected that active preparation would promote direct associations between sensory representations of the instructed targets and motor representations of the linked responses, to enable efficient task execution. In memorization blocks, less binding should occur, as the SR mapping information was merely maintained for comparison with the probe screen. Further assuming that delay-related decoding accuracy in the FFA and the PPA indexes the fidelity of target representation (see above), we reasoned that pattern separation in these regions might be associated with the speed of performance, especially in implementation blocks. Such a link could be considered a neural signature of a "prepared reflex" and mirror response-priming effects observed in behavioral studies. To test this assumption, we calculated partial correlations across participants. Specifically, we correlated delay-related decoding accuracies in visual areas (averaged across FFA and PPA) with RT (averaged across face instructions and house instructions), while controlling for variance related to decoding accuracies in the same visual areas during the instruction phase. The latter was done to eliminate potential perception-related signal bleed from the instruction phase.

Pattern Similarity Analysis

We conducted a final set of analyses to further specify the differential nature of delay activity between the 2 block types. Specifically, we wanted to adjudicate between 2 alternative explanations of enhanced instruction coding in frontoparietal areas in implementation blocks. One possibility, mentioned above, is that preparation promotes a change in the way that instructions are internally represented via binding of perceptual codes and motor codes into a compound action plan (transformation hypothesis). Alternatively, however, enhanced instruction coding could also reflect the mere amplification of a common representational template, for example, via enhanced focus (amplification hypothesis). To pit these 2 hypotheses against each other, we analyzed the similarity of activation patterns within frontoparietal regions between the different trial phases (instruction phase vs. delay) and block types (implementation vs. memorization) with the following rationale: Under the transformation hypothesis, one would predict pattern stability (i.e., the similarity of patterns from the instruction phase and the delay of the same block type) to be lower in implementation blocks than in memorization blocks, reflecting the putative conversion of semantic facts into task sets. In addition, cross-conditional similarities (i.e., the similarity between instruction-related patterns of one block type and delay-related patterns of the other block type) should be asymmetric with greater similarity between delay-related patterns from memorization blocks and instruction-related patterns from implementation blocks than vice versa (reflecting greater persistence of an initially shared representation in memorization blocks). In contrast, the amplification hypothesis would predict the opposite pattern of results with greater pattern stability in implementation blocks and greater cross-conditional similarity between delay-related patterns of implementation blocks and instruction-related patterns of memorization blocks than vice versa (reflecting the greater perpetuation of a common representational scheme in implementation blocks). To address this question, we calculated average β images across the five experimental runs pertaining to each block type (implementation vs. memorization), trial phase (instruction phase vs. delay), and instruction category (face instructions vs. house instructions). We then extracted voxels corresponding to a composite volume covering all of the frontoparietal ROIs (i.e., bilateral IFJ, IFS, and IPS), and estimated the similarity between the different conditions via Pearson correlations. Correlation values were averaged within subjects across face and house instructions, normalized via Fisher's transformation, and compared via paired-samples *t*-tests.

Results

Behavioral Results

The RT analysis revealed a significant main effect of BLOCK TYPE ($F_{1,22} = 355.457$, $P < 0.001$, $\eta^2 = 0.942$), reflecting faster responses in implementation blocks than in memorization blocks (see Table 1). Importantly, neither the main effect of INSTRUCTION CATEGORY ($F_{1,22} = 2.550$, $P = .125$, $\eta^2 = 0.104$) nor the interaction term was significant ($F_{1,22} = 0.064$, $P = 0.803$, $\eta^2 = 0.003$). The analysis of performance accuracy revealed non-significant main effects of BLOCK TYPE ($F_{1,22} = 0.173$, $P = 0.681$, $\eta^2 = 0.008$) and INSTRUCTION CATEGORY ($F_{1,22} = 0.042$, $P = 0.840$, $\eta^2 = 0.002$). The interaction term reached significance ($F_{1,22} = 4.731$, $P = 0.041$, $\eta^2 = 0.177$), but post hoc comparisons revealed that this merely reflected the converse directionality of non-significant differences within both block types. That is, numerically there were fewer errors with face instructions than with house instructions in implementation blocks, while the reverse was observed in memorization blocks (see Table 1). However, neither of these within-block differences reached statistical significance (*P*-values of both pairwise comparisons >0.212). Next, we compared RT between face and house instructions within each participant to evaluate the matching of instruction categories at an individual level (see Materials and Methods section). Within each block type, one-way ANOVAs between RT of correct responses with face and house instructions reached significance at $P < 0.05$ for only 3 out of the 23 participants (removing these subjects from the decoding analyses did not alter the pattern of results). We consider this as confirmation that instruction categories were successfully matched in terms of difficulty for both block types.

Accuracy on catch trials was also high in both block types (see Table 1). The GLM revealed a main effect of INSTRUCTION CATEGORY ($F_{1,22} = 4.62$, $P < .001$, $\eta^2 = 0.321$) indicating that face distractors were detected less often than house distractors (see Table 1). Moreover, there was a significant interaction between the 2 factors ($F_{1,22} = 5.55$, $P < 0.001$, $\eta^2 = 0.158$), reflecting that the accuracy difference between face instructions and house instructions was stronger in implementation blocks ($t_{22} = 4.49$, $P < 0.001$, $d = 0.936$) than in memorization blocks ($t_{22} = 1.61$, $P = 0.12$, $d = 0.254$).

ROI-based Decoding Analyses

The GLM of decoding accuracies within the predefined ROIs revealed a significant main effect of TRIAL PHASE ($F_{1,22} = 183.033$,

Table 1 Behavioral performance (values are means and standard errors)

	Implementation		Memorization	
	RT (ms)	% Correct	RT (ms)	% Correct
Regular trials				
Mean	1043 (44)	89.3 (1.1)	1794 (41)	89.4 (0.9)
Face	1027 (42)	90.8 (1.2)	1780 (44)	88.4 (1.1)
House	1062 (49)	89.1 (1.5)	1807 (43)	90.5 (1.1)
Catch trials				
Mean		80.8 (2.9)		78.8 (3.9)
Face		71.7 (4.5)		75.8 (5.8)
House		90.0 (2.5)		81.7 (3.5)

$P < 0.001$, $\eta^2 = 0.893$), reflecting greater accuracies during the instruction phase than during the delay ($t_{22} = 13.529$, $P < 0.001$, $d = 2.821$; instruction phase = 68%, delay = 53%) across ROIs. As mentioned above, this is likely due, at least partly, to the absence of visual stimulation in the delay that eliminates perceptual confounds between instruction categories. The main effect of ROI was significant as well ($F_{1,22} = 71.110$, $P < 0.001$, $\eta^2 = 0.764$), indicating differential degree of pattern separation across the employed ROIs (IFS = 53%, IFJ = 56%, IPS = 57%, FFA = 61%, PPA = 75%). The main effect of BLOCK TYPE was also significant ($F_{1,22} = 8.479$, $P = 0.008$, $\eta^2 = 0.278$), reflecting greater decoding accuracies in implementation blocks than in memorization blocks ($t_{22} = 2.912$, $P = 0.008$, $d = 0.612$; implementation = 62%, memorization = 59%). Importantly, the main effects of TRIAL PHASE and BLOCK TYPE interacted ($F_{1,22} = 10.893$, $P = 0.003$, $\eta^2 = 0.331$). Inline with our expectations, decoding accuracies did not differ between block types during the instruction phase ($t_{22} = 0.485$, $P = 0.633$, $d = 0.102$; implementation = 68%, memorization = 68%), but in the delay they were significantly greater and only above-chance level in implementation blocks ($t_{22} = 3.479$, $P = 0.002$, $d = 0.731$; implementation = 56%, memorization = 50%; see Fig. 3). The nonsignificant 3-way interaction indicated that this pattern of results occurred relatively uniformly across the different ROIs ($F_{1,22} = 1.405$, $P = 0.248$, $\eta^2 = 0.058$; see Fig. 3 for an illustration). Note that a more detailed GLM that contained the additional factor HEMISPHERE (left vs. right) only revealed a significant main effect of HEMISPHERE ($F_{1,22} = 31.58$, $P < .001$, $\eta^2 = 0.482$; left hemisphere = 59%; right hemisphere = 62%), but no significant interaction term involving the factors HEMISPHERE and BLOCK TYPE (all F-values <1).

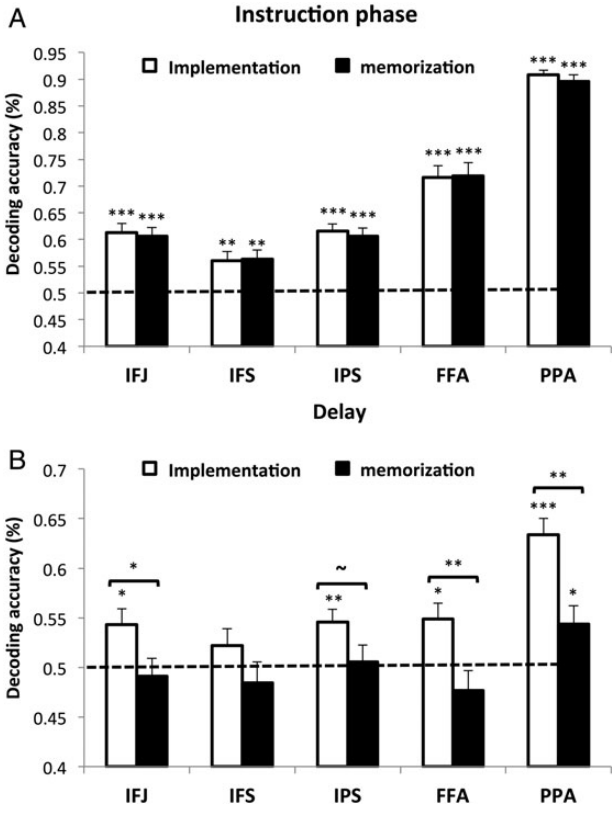

Figure 3. Decoding accuracies within ROIs separately for the instruction phase (upper panel) and for the delay (lower panel). ~$P < 0.10$, *$P < 0.05$, **$P < 0.01$, ***$P < 0.001$.

Accordingly, the block type manipulation did not affect ROIs in the left and right hemisphere differentially.

After establishing this delay-specific effect of the block type manipulation on instruction coding, we sought to further rule out that the differential classification was driven by differences in task difficulty (see part on internal validation in the Materials and Methods section). To this end, we correlated RT-differences between face instructions and house instructions (recoded as absolute values) with delay-related decoding accuracies of each ROI, separately for both block types. None of these correlations reached significance, despite using uncorrected thresholds (implementation: all P-values >0.531, memorization: all P-values >0.127). Notably, the same pattern was observed when performance accuracy (implementation: all P-values >0.367, memorization: all P-values >0.243), IES (implementation: all P-values >0.090, memorization: all P-values >0.279), or performance accuracy on catch trials (implementation: all P-values >0.145, memorization: all P-values >0.122) were subjected to the analysis. These findings emphasize that decoding of instruction category information is very unlikely to result from subtle performance confounds.

Whole-brain Decoding Analyses

Overall, the whole-brain decoding analyses largely confirmed our selection of ROIs and revealed only few additional involved areas (see Fig. 4 and Tables 2 and 3). During the instruction phase, decoding accuracy was above chance most strongly in posterior cortices, centered in the parahippocampal gyri and spreading widely across occipital, temporal, and parietal lobes. In addition, instruction category could be decoded from activity in the frontal lobe. Bilateral clusters were located along the IFS and extended onto the precentral gyri. An additional cluster was located in the dorsal frontomedian cortex (dFMC). These results were very similar for both block types, and only few clusters were identified that exhibited increased instruction category information during implementation blocks. These clusters were located in the right premotor cortex, the cingulate gyrus/caudate, the right anterior PFC, and the lingual gyrus. No area was identified as carrying more information during memorization blocks. As in the ROI analysis, the results differed more clearly between block types during the delay. In memorization blocks, decoding accuracy was above chance only in bilateral clusters within mid-occipital and posterior parietal cortices. In contrast, in implementation blocks, decoding accuracy was above chance throughout most of the occipital lobe and in several frontal and parietal clusters. Contrasts between block types confirmed that instruction category was represented more strongly during implementation blocks in a number of occipital, parietal, and frontal regions, among them the left IPS and the right IFJ. No region was found to carry stronger information about instruction categories in memorization blocks (Fig. 4).

Decoding in Additional Regions With Univariate Effects

We performed follow-up univariate analyses of delay activity to identify additional brain areas that show delay-related increases in average activation, relative to the ITI (see Materials and Methods section). Contrasts within each block type and the conjunction analysis across both block types consistently revealed a number of further areas, among them the pre-SMA (6, 5, 52), the anterior insula (left: −30, 20, 4; right: 33, 26, 4), the right premotor cortex (33, −13, 64), and 2 subcortical clusters centered at the caudate heads and extending caudally to the thalamus

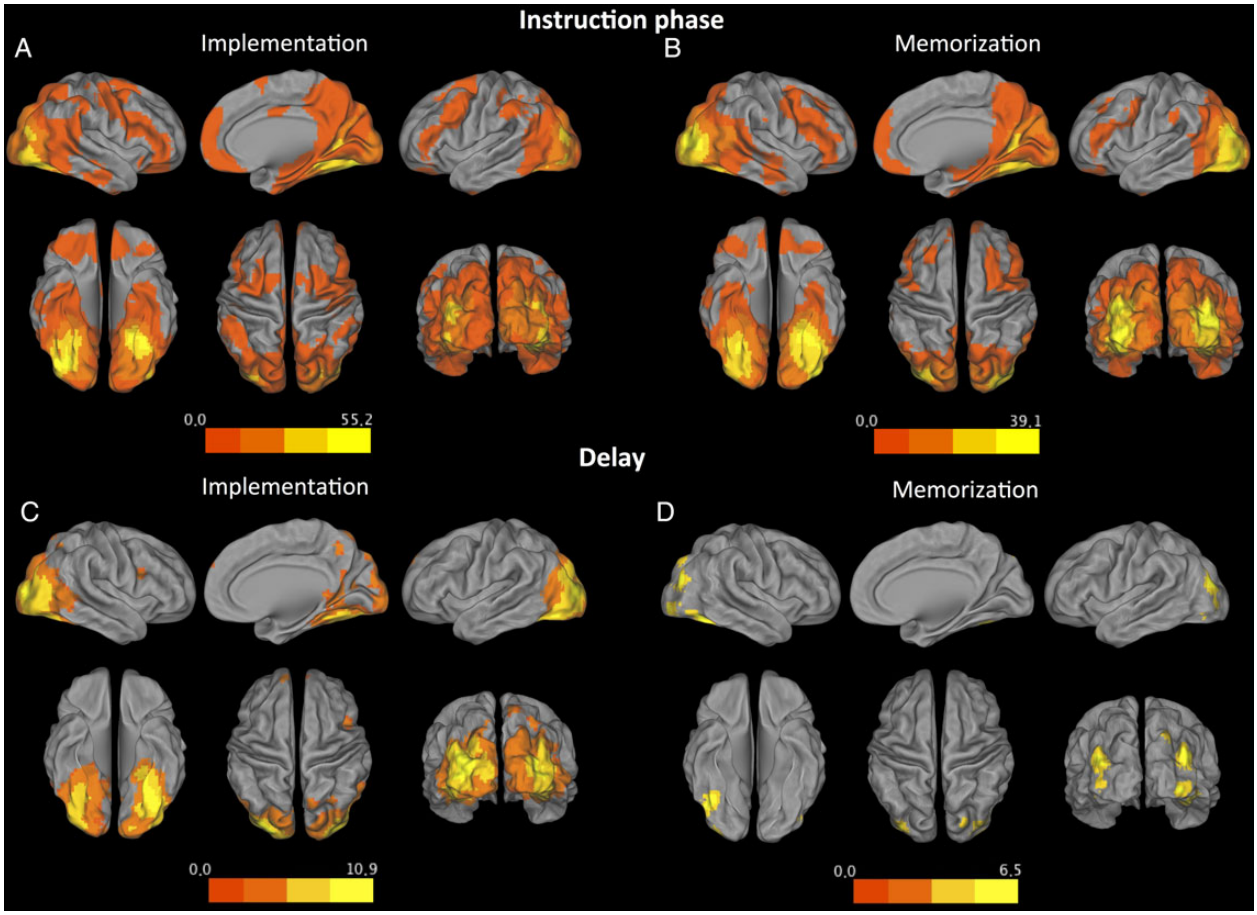

Figure 4. The upper panel illustrates the whole-brain decoding results from the instruction phase, separately for implementation blocks (*A*) and memorization blocks (*B*). The lower panel illustrates the whole-brain decoding results from the delay, separately for implementation blocks (*C*) and memorization blocks (*D*). Note: All contrasts are based on a voxel threshold of $P < .001$, and a minimal cluster extent of 22 contiguous voxels.

Table 2 Results of the whole-brain decoding analysis during the instruction phase

Area	Hemisphere	Peak (MNI)	T_{Max}	Extent (voxels)
Implementation				
OTC	L/R	36 −67 −8	57.06	22 177
PMD/IFJ/IFS	R	39 −4 49	8.08	1774
	L	−42 14 28	5.44	1081
dFMC	L/R	−3 50 34	4.89	126
Memorization				
OTC	L/R	45 −73 −7	39.07	23 126
PMD/IFJ/IFS	R	42 23 22	8.28	1825
	L	−45 26 19	6.04	1163
dFMC	L/R	6 59 13	8.15	390
Implementation > memorization				
Cingulate/caudate	R	9 2 25	4.54	69
PMD	R	36 −13 49	4.22	50
Anterior PFC	R	18 35 −8	3.95	36
Lingual gyrus	L	−3 −82 −5	3.82	24

OTC, occipitotemporal cortex; PMD, dorsal premotor cortex; IFJ, inferior frontal junction; IFS, inferior frontal sulcus; dFMC, dorsal frontomedian cortex.

(left: −12, 2, 4; right: 12, 5, 4; see Supplementary Figs 2 and 3 for details). Follow-up decoding analyses, based on spherical ROIs (radius = 10 mm, size = 515 voxels) centered at the peaks of each cluster, showed that activity in none of these regions carried above-chance information about instruction category (all *P*-values >0.242). This indicates that these regions might be involved in content-unspecific preparatory processes.

External Validation

Next, we examined the relevance of the MVPA results for task performance by means of correlational analyses (see Materials and Methods section). Consistent with our hypothesis that preparation elicits direct associations between target representations and response codes, there was a significant correlation between delay-related decoding accuracies in dedicated visual regions and RT in implementation blocks ($r = -0.441$, $P = 0.029$, $R^2 = 0.188$), while controlling for instruction-related decoding accuracy in visual regions. The negative magnitude reflects that better discrimination of instruction categories in these regions was associated with faster responses. No such correlation was found in memorization blocks ($r = -0.001$, $P = 0.999$, $R^2 < 0.001$), despite equivalent variance in terms of both RT and decoding accuracy in both block types (Mauchly tests of all pairwise comparisons were non-significant). We tested the difference between these 2 correlations for significance using Steiger's method (see Steiger 1980) that applies an asymptotic *z*-test after normalizing correlations via Fisher's transformation. This confirmed that the difference between the 2 correlation coefficients was significant ($z = 2.045$; $P = 0.023$; see Fig. 5).

Following up on this finding, we also wanted to explore the putative role of frontoparietal regions in the genesis of this state. Given our assumption that these regions are involved in constructing task sets based on instruction, one could expect a similar association between neural specificity in these areas during the delay and RT. We tested this idea by means of a partial correlation analysis, analogously to the one described above (i.e., we correlated delay-related decoding accuracy in one large volume covering all frontoparietal areas with RT, while controlling for variance related to decoding accuracies from the instruction phase). However, no significant correlations between decoding results and performance were observed (both *P*-values >0.433). We therefore addressed an alternative possibility, namely that frontoparietal areas contribute to performance indirectly by proactively adjusting sensorimotor pathways during the encoding stage (see Ruge and Wolfensteller (2010) for evidence that frontal activity during encoding can contribute to the quality of later task performance). To examine this idea, we correlated decoding accuracies in frontoparietal regions during the instruction phase with delay-related decoding accuracies in visual areas, while controlling for variance related to instruction-related decoding accuracies. In partial support for this hypothesis, there was a marginally significant correlation in implementation blocks ($r = 0.395$, $P = 0.069$, $R^2 = 0.155$), but not in memorization blocks ($r = -0.014$, $P = 0.988$, $R^2 = 0.001$), and a post hoc comparison revealed a marginally significant difference between these 2 correlations ($z = 1.823$; $P = 0.068$).

Table 3 Results of the whole-brain decoding analysis during the delay

Area	Hemisphere	Peak (MNI)	T_{Max}	Extent (voxels)
Implementation				
OTC	L/R	33 −85 10	16.54	12 422
IFJ	R	42 8 25	5.09	60
SFS	L	−33 35 52	4.61	36
	R	27 44 52	4.28	28
PCC	L/R	6 −49 43	4.57	28
dFMC	L/R	−12 53 37	4.28	121
Memorization				
OTC	L	−36 −85 19	6.90	578
	R	36 −79 22	5.85	1026
Fusiform gyrus	L	−45 −76 −14	3.95	28
Implementation > memorization				
OTC	L/R	30 −88 10	6.18	2980
PMD	R	33 −7 37	4.65	50
IPS	L	−51 −49 49	4.62	77
SFS	L	−27 32 52	4.25	93
IFJ	R	42 11 25	4.10	38
MTG	L	−45 −70 25	3.94	34
Caudate	R	21 −28 19	3.88	85
IPL	R	57 −46 25	3.74	28
SPL	R	12 −58 64	3.62	48

OTC, occipitotemporal cortex; PMD, dorsal premotor cortex; IFJ, inferior frontal junction; SFS, superior frontal sulcus; PCC, posterior cingulate cortex; dFMC, dorsal frontomedian cortex; IPS, intraparietal sulcus; MTG, middle temporal gyrus; IPL, inferior parietal lobule; SPL, superior parietal lobule.

Pattern Similarity Analysis

Finally, to further elucidate the nature of differential instruction coding, we compared the similarity of frontoparietal activation patterns from the instruction phase and the delay of both block types (see Materials and Methods section). As shown in Fig. 6, similarity between block types was generally greater during the instruction phase than during the delay ($t_{22} = 6.148$, $P < 0.001$, $d = 1.282$). Pattern stability tended to be greater in memorization blocks than in implementation blocks ($t_{22} = 2.041$, $P = 0.053$, $d = 0.426$). Finally, cross-conditional similarities differed significantly with greater similarity between delay-related patterns from memorization blocks and instruction-related patterns from implementation blocks than for the reverse cross-correlation ($t_{22} = 3.648$, $P < 0.001$, $d = 0.761$). This pattern of results clearly favors the transformation hypothesis and suggests that the state of task readiness that is established in the delay phase of implementation blocks reflects the emergence of a distinct cognitive state.

Of note, an analogous analysis of pattern similarity within dedicated visual areas revealed only a significant difference between the 2 trial phases with greater similarity between block types during the instruction phase than during the delay ($t_{22} = 9.566$, $P < 0.001$, $d = 1.995$). No significant differences were found in terms of pattern stability within-block types ($t_{22} = 1.396$, $P = 0.177$, $d = 0.291$) or in terms of cross-conditional similarity ($t_{22} = 1.113$, $P = 0.278$, $d = 0.232$). In line with results from recording studies in monkeys, these findings suggest that the tuning of frontoparietal regions is highly flexible and can

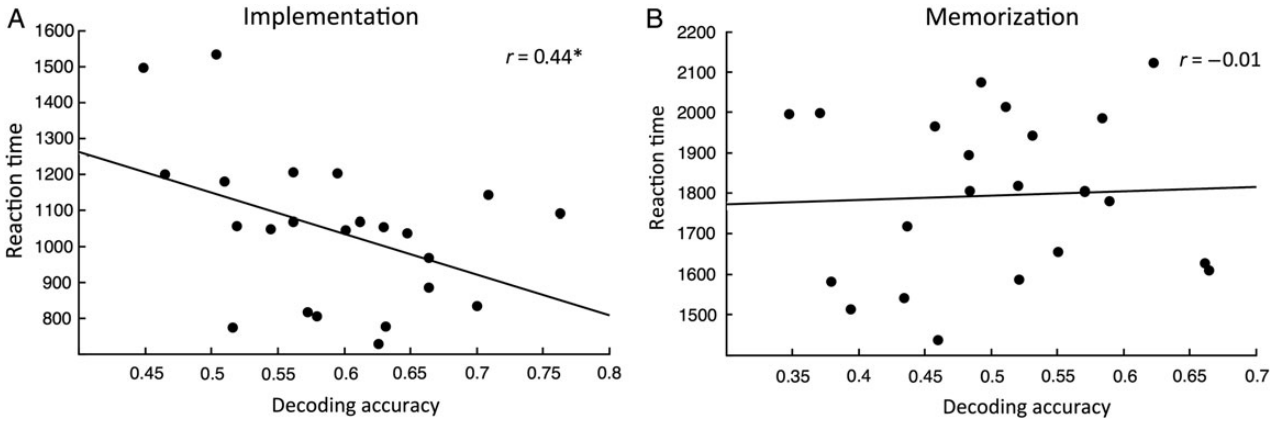

Figure 5. Scatter plots displaying the partial correlations between decoding accuracies in dedicated visual areas (average across FFA and PPA), and reaction time (average across face instructions and house instructions), while controlling for variance related to decoding accuracies in the instruction phase, separately for implementation blocks (*A*) and memorization blocks (*B*).

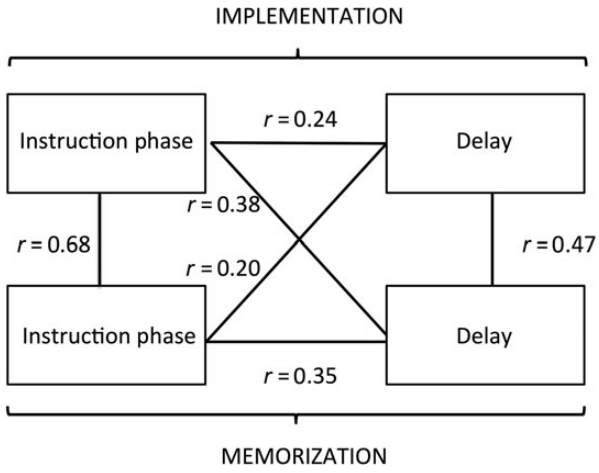

Figure 6. Results of the pattern similarity analysis, indicating the similarity of frontoparietal activation patterns between the different trial phases (instruction vs. delay) and block types (implementation vs. memorization). Values reflect z-normalized Pearson correlations (see Materials and Methods section).

represent the same task parameters in dynamically changing neural codes (Crowe et al. 2010; Stokes et al. 2013, Stokes 2015), whereas the response profile in visual regions is more fixed and stimulus-grounded.

Discussion

We compared the neural representation of novel SR mapping instructions that were either prepared for execution or memorized for recognition. Although the instructed information was successfully maintained in both cases, preparation enhanced the discriminability of instruction categories in a set of frontal, parietal, and visual regions during the delay between task instruction and application. Moreover, only during preparation was the coding strength in visual regions, implicated in processing the instructed target stimuli, associated with the speed of performance. Frontoparietal regions, on the other hand, appeared to be involved in creating this state during both encoding and delay phases. Below, we discuss the implications of our findings along with possible directions for further inquiry.

A Mechanism for Proactive Control Over Novel Tasks

Patients with frontal lobe damage often exhibit a striking mismatch between intact formal knowledge regarding how to behave in a novel task environment and a severely perturbed capability to bring this knowledge into control of behavior (Duncan et al. 1996, 1997; Luria 1966). While such cases imply that implementation of novel tasks is supported by some sort of knowledge transformation, they leave open how exactly this process should be conceived. Our results indicate that proactive control over novel task demands can take the form of a "prepared reflex" (Hommel 2000; Meiran et al. 2014). That is, task preparation appears to promote the integration of perceptual target representations and motor response codes into a compound conditional action plan that proactively facilitates performance. Behaviorally, this was evidenced by the observation that instructed stimuli gained the power to automatically activate the associated responses in a secondary task, only when the instruction was prepared, but not when it was memorized (similar to Liefooghe et al. 2012). At the brain level, we observed that the neural specificity within functionally dedicated visual areas during the delay was tied to the speed of task application. This correlation was found only in implementation blocks and thus likely reflects a neuronal signature of preconfigured sensorimotor pathways wherein the fidelity of the target representation during the delay determines the automaticity of subsequent performance.

How is this state of task readiness established? Evidently, frontoparietal regions, play a central role in this respect. Previous studies have shown that these regions are reliably activated when novel task instructions are received (Ruge and Wolfensteller 2010; Dumontheil et al. 2011; Hartstra et al. 2011; Cole, Reynolds, et al. 2013) and some further evidence exists that this activity is related to the quality of later task implementation (Ruge and Wolfensteller 2010). Our study extends these findings by pointing toward a mechanism by which these regions may facilitate task performance. First, we observed that early coding of instruction categories in these regions tended to be associated with better pattern separation in visual areas during the subsequent delay (which was in turn associated with performance). Hence, frontoparietal regions may facilitate novel task performance as early as during encoding, by selectively tuning representations in sensory regions based on the instructed task demands. There is indeed copious evidence that frontoparietal regions are critically involved in stimulus encoding (e.g., Gazzaley et al. 2007; Mayer et al. 2007; Chadick and Gazzaley 2011). However, the precise nature of this contribution remains difficult to infer. Discrimination of instruction-categories during encoding could reflect either specificity for the stimulus content (Golby et al. 2001; Johnson et al. 2003) or instead the utilization of domain-specific control processes such as different amounts of semantic vs. spatial analysis, verbal rehearsal, etc. (Kuhl et al. 2012). Beyond encoding, there was also robust category discrimination within frontoparietal areas during the delay in implementation blocks but not in memorization blocks. This finding is of particular interest in light of current debates on the neural mechanisms supporting short-terms memory storage (Sreenivasan, Curtis, et al. 2014; D'Esposito and Postle 2015). Frontoparietal regions typically exhibit robust activation increases throughout memory delays, yet MVPA studies indicate that this activity carries only weak information about to-be-memorized contents (see D'Esposito and Postle 2015), while coding more robustly for abstract, goal-related variables such as task rules that are used for stimulus classification (Bode and Haynes 2009; Woolgar, Hampshire, et al. 2011; Woolgar, Thompson, et al. 2011; Zhang et al. 2013; Waskom et al. 2014; Etzel et al. 2015). Hence, compared with the representation of specific task elements such as particular stimuli or responses, frontoparietal activity seems to be more diagnostic about the abstract, context-based connections among those elements (e.g., Bode and Haynes 2009; Woolgar, Hamsphire, et al. 2011; Woolgar, Thompson, et al. 2011). This resonates well with our interpretation of the differential decoding results during the delay phase; depending on the block type, the same instruction screens appeared to be represented either procedurally as an SR mapping rule, or semantically as a visual memorandum. Evidently, the former is represented more distinctly in frontoparietal regions.

The Nature of Semantic Task Representations

Given the considerations above, an important question concerns the exact nature of instruction coding within memorization blocks. In these blocks, a striking data pattern was observed: on the one hand, the absence of response priming in the pilot study and the drop in delay-related decoding accuracy consistently indicated that no active SR mapping representations were

established. On the other hand, participants were clearly able to maintain the instructed information with equivalent accuracy and frontoparietal regions were robustly activated during the delay phase. This suggests that memorization blocks similarly drew upon cognitive control processes, but that participants were employing a different strategy with a different allocation of frontoparietal resources.

How should this strategy be conceived? It has been argued that preparatory coding schemes critically depend upon the information conveyed by the task cue, and upon the nature of anticipated targets and distractors (Stokes 2011). Even though identical instruction screens were used in both block types, they provided participants with differential information about the upcoming target phase. Implementation blocks allowed for relatively specific predictions to be made, as only three different outcomes were possible (2 targets and one distractor), each of which was associated with a contingent response. This setting clearly encourages instruction coding as a conditional action plan and "motor imagery" as a suitable rehearsal strategy (Jeannerod and Decety 1995; Jeannerod and Frak 1999; Jeannerod 2001). In memorization blocks, the same facts about the instruction had to be maintained, but it was far less predictable how these facts would have to be applied. The larger number of possible probe screens (four on regular trials) and their greater complexity may have called instead for a stable memory representation of the instruction screen to be matched with the probe screen. The fact that the instructed effectors were also used to indicate the match–mismatch decision may further have discouraged motor imagery of the SR mappings as a maintenance strategy. It is surprising nevertheless, that in the delay of memorization blocks decoding accuracies fell to chance not only in frontoparietal but also in some of the visual areas implicated in visual short-term storage (significant decoding in PPA but not FFA). Stronger decoding in these regions might be achieved with individually tailored ROIs that are typically more sensitive than group-based localizers (Saxe et al. 2006), though the same localizer was appreciably accurate in implementation blocks. This, along with the absence of regions coding specifically for memorized instructions, suggests that visual instruction coding was generally weaker in memorization blocks. Clearly, great caution is warranted when interpreting a single negative result, but these findings could be due to a greater reliance of rehearsal processes on verbal codes that likely distinguish less between face and house instructions than percept-like sensory codes.

Limitations and Future Directions

Our study contains several limitations worth noting that should encourage further investigation. One central aspect concerns the blocking of implementation and memorization demands. As noted above, we deliberately chose this experimental design, based on behavioral pilot studies, to prevent participants from utilizing a single general strategy and thereby maximize the power of our manipulation. Now, after delineating the general signature of preparation on instruction coding, a promising next step could be to translate our paradigm into event-related designs where implementation and memorization demands vary on a trial-by-trial basis. A particularly informative approach could be to present task cues, signaling implementation versus memorization demands, after the initial encoding of SR mapping instructions. If feasible, such a procedure could be of great help in further constraining the time period at which the formation of task sets takes place, compared with our broader distinction between instruction and delay phases (we are grateful to an anonymous reviewer for this suggestion). Another powerful strategy to substantiate our findings could be to combine our decoding protocol with high temporal resolution techniques such as magnetoencephalography in order to trace gradual transitions through representational states over the time course of the pre-target phase (see Stokes et al. 2013, 2015; King and Dehaene 2014). Likewise, analyses of functional connectivity could be employed to further delineate the neural sources of top-down control under implementation and memorization demands (Zanto et al. 2010; Baldauf and Desimone 2014). Finally, future studies should also attempt to circumvent binary classifications between disparate stimulus categories such as faces and houses. Although the same stimulus categories were used in both block types, and diverse control analyses consistently indicated that our decoding results were not driven by performance differences, we cannot rule out with certainty that general differences between the stimulus materials (e.g., the greater biological significance of face stimuli) may have impacted the classification to some degree. Future studies could address this issue by using more similar stimulus types (e.g., different types of objects; similar to Lee et al. 2013) and/or by including more than 2 categories of instructions (see Cole, Reynolds, et al. 2013).

Conclusion

In summary, the present study suggests that the implementation of newly instructed goal-directed behavior is supported by the transformation of semantic task knowledge into a temporary task set that proactively tunes sensorimotor processing in line with anticipated demands. These task sets optimize subsequent performance and can be distinguished from semantic memory representations of the same instructions. Together, our findings shed new light on how instructions are converted into specific action plans and also help in disentangling the neural mechanisms that support memory storage and task control.

Supplementary Material

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

Funding

The authors would like to thank Maggie Lynn and Jan Derrfuss for helpful discussion, and two anonymous reviewers for thoughtful comments on an earlier version of the manuscript. *Conflict of Interest*: None declared.

Notes

This work was supported by grant B/09019/02 from the Flemish Research foundation (FWO) and grant P7/33 "COOL" from the Belgian Science Policy Office (Interuniversity Poles of Attraction Program).

References

Asaad WF, Rainer G, Miller EK. 1998. Neural activity in the primate prefrontal cortex during associative learning. Neuron. 21:1399–1407.

Baldauf D, Desimone R. 2014. Neural mechanisms of object-based attention. Science. 344:424–427.

Bhandari A, Duncan J. 2014. Goal neglect and knowledge chunking in the construction of novel behaviour. Cognition. 130:11–30.

Bode S, Haynes JD. 2009. Decoding sequential stages of task preparation in the human brain. NeuroImage. 45:606–613.

Chadick JZ, Gazzaley A. 2011. Differential coupling of visual cortex with default or frontal-parietal network based on goals. Nat Neurosci. 14:830–832.

Cohen JD, Perlstein WM, Braver TS, Nystrom LE, Noll DC, Jonides J, Smith EE. 1997. Temporal dynamics of brain activation during a working memory task. Nature. 386:604–608.

Cohen-Kdoshay O, Meiran N. 2009. The representation of instructions operates like a prepared reflex: flanker compatibility effects found in first trial following S-R instructions. Exp Psychol. 56:128–133.

Cole MW, Bagic A, Kass R, Schneider W. 2010. Prefrontal dynamics underlying rapid instructed task learning reverse with practice. J Neurosci. 30:14245–14254.

Cole MW, Laurent P, Stocco A. 2013. Rapid instructed task learning: a new window into the human brain's unique capacity for flexible cognitive control. Cogn Affect Behav Neurosci. 13:1–22.

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

Cole MW, Schneider W. 2007. The cognitive control network: Integrated cortical regions with dissociable functions. Neuroimage. 37:343–360.

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

Crowe DA, Averbeck BB, Chafee MV. 2010. Rapid sequences of population activity patterns dynamically encode task-critical spatial information in parietal cortex. J Neurosci. 30:11640–11653.

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

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

Dumontheil I, Thompson R, Duncan J. 2011. Assembly and use of new task rules in frontoparietal cortex. J Cogn Neurosci. 23:168–182.

Duncan J. 1986. Disorganization of behaviour after frontal lobe damage. Cogn Neuropsychol. 3:271–290.

Duncan J. 2010. The multiple-demand (MD) system of the primate brain: mental programs for intelligent behaviour. Trends Cogn Sci. 14:172–179.

Duncan J, Emslie H, Williams P, Johnson R, Freer C. 1996. Intelligence and the frontal lobe: the organization of goal-directed behavior. Cogn Psychol. 30:257–303.

Duncan J, Johnson R, Swale MJ. 1997. Frontal lobe deficits after head injury: unity and diversity of function. Cogn Neuropsychol. 14:713–741.

Duncan J, Parr A, Woolgar A, Thompson R, Bright P, Cox S, Bishop S, Nimmo-Smith I. 2008. Goal neglect and Spearman's g: competing parts of a complex task. J Exp Psychol Gen. 137:131–148.

Duncan J, Schramm M, Thompson R, Dumontheil I. 2012. Task rules, working memory, and fluid intelligence. Psychon Bull Rev. 19:864–870.

Emrich SM, Riggall AC, Larocque JJ, Postle BR. 2013. Distributed patterns of activity in sensory cortex reflect the precision of multiple items maintained in visual short-term memory. J Neurosci. 33:6516–6523.

Epstein R, Harris A, Stanley D, Kanwisher N. 1999. The parahippocampal place area: recognition, navigation, or encoding? Neuron. 23:115–125.

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

Ester EF, Anderson DE, Serences JT, Awh E. 2013. A neural measure of precision in visual working memory. J Cogn Neurosci. 25:754–761.

Etzel JA, Cole MW, Zacks JM, Kay KN, Braver TS. 2015. Reward motivation enhances task coding in frontoparietal cortex. Cereb Cortex. doi: 10.1093/cercor/bhu327.

Fedorenko E, Duncan J, Kanwisher N. 2013. Broad domain generality in focal regions of frontal and parietal cortex. Proc Natl Acad Sci. 110:16616–16621.

Fedorenko E, Duncan J, Kanwisher N. 2012. Language-selective and domain-general regions lie side by side within Broca's area. Curr Biol. 22:2059–2062.

Fuster JM. 1973. Unit activity in prefrontal cortex during delayed-response performance: neuronal correlates of transient memory. J Neurophysiol. 36:61–78.

Fuster JM. 1980. The prefrontal cortex: anatomy, physiology, and neuropsychology of the frontal lobe. New York: Raven.

Fuster JM. 2001. The prefrontal cortex—an update: time is of the essence. Neuron. 30:319–333.

Fuster JM, Alexander GE. 1971. Neuron activity related to short-term memory. Science. 173:652–654.

Gazzaley A, Rissman J, Cooney J, Rutman A, Seibert T, Clapp W, D'Esposito M. 2007. Functional interactions between prefrontal and visual association cortex contribute to top-down modulation of visual processing. Cereb Cortex. 17: 125–135.

Golby AJ, Poldrack RA, Brewer JB, Spencer D, Desmond JE, Aron AP, Gabrieli JD. 2001. Material-specific lateralization in the medial temporal lobe and prefrontal cortex during memory encoding. Brain. 124:1841–1854.

Hanke M, Halchenko YO, Sederberg PB, Hanson SJ, Haxby JV, Pollmann S. 2009. PyMVPA: a python toolbox for multivariate pattern analysis of fMRI data. Neuroinformatics. 7:37–53.

Harrison SA, Tong F. 2009. Decoding reveals the contents of visual working memory in early visual areas. Nature. 458: 632–635.

Hartstra E, Kühn S, Verguts T, Brass M. 2011. The implementation of verbal instructions: an fMRI study. Hum Brain Mapp. 32:1811–1824.

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

Hommel B. 2000. The prepared reflex: Automaticity and control in stimulus response translation. In: Monsell S, Driver J, editors. Attention and performance, 18: Control of cognitive processes. Cambridge, MA: MIT Press. pp. 247–273.

Ikkai A, Curtis CE. 2011. Common neural mechanisms supporting spatial working memory, attention and motor intention. Neuropsychologia. 49:1428–1434.

Jeannerod M. 2001. Neural simulation of action: a unifying mechanism for motor cognition. Neuroimage. 14:103–109.

Jeannerod M, Decety J. 1995. Mental motor imagery: a window into the representational stages of action. Curr Opin Neurobiol. 5:727–732.

Jeannerod M, Frak V. 1999. Mental imaging of motor activity in humans. Curr Opin Neurobiol. 9:735–739.

Johnson MK, Raye CL, Mitchell KJ, Greene EJ, Anderson AW. 2003. fMRI evidence for an organization of prefrontal cortex by both

type of process and type of information. Cereb Cortex. 13:265–273.
Kanwisher N, McDermott J, Chun MM. 1997. The fusiform face area: a module in human extrastriate cortex specialized for face perception. J Neurosci. 17:4302–4311.
Kanwisher N, Yovel G. 2006. The fusiform face area: a cortical region specialized for the perception of faces. Phil Trans Roy Soc Lond B Biol Sci. 361:2109–2128.
King JR, Dehaene S. 2014. Characterizing the dynamics of mental representations: the temporal generalization method. Trends Cogn Sci. 18:203–210.
Konow A, Pibram KH. 1970. Error recognition and utilization produced by injury to the frontal cortex in man. Neuropsychologia. 8:489–491.
Kriegeskorte N, Goebel R, Bandettini P. 2006. Information-based functional brain mapping. Proc Natl Acad Sci. 103:3863–3868.
Kubota K, Niki H. 1971. Prefrontal cortical unit activity and delayed alternation performance in monkeys. J Neurophysiol. 34:337–347.
Kuhl BA, Rissman J, Wagner AD. 2012. Multi-voxel patterns of visual category representation during episodic encoding are predictive of subsequent memory. Neuropsychologia. 50:458–469.
Lee SH, Kravitz DJ, Baker CI. 2013. Goal-dependent dissociation of visual and prefrontal cortices during working memory. Nat Neurosci. 16:997–999.
Liefooghe B, De Houwer J, Wenke D. 2013. Instruction-based response activation depends on task preparation. Psych Bull Rev. 20:481–487.
Liefooghe B, Wenke D, De Houwer J. 2012. Instruction-based task-rule congruency effects. J Exp Psychol Learn Mem Cogn. 38:1325–1335.
Luria AR. 1966. Higher cortical functions in man. London: Tavistock.
Mayer JS, Bittner RA, Nikolić D, Bledowski C, Goebel R, Linden DEJ. 2007. Common neural substrates for visual working memory and attention. Neuroimage. 36:441–453.
Meiran N, Cole MW, Braver TS. 2012. When planning results in loss of control: intention-based reflexivity and working-memory. Front Hum Neurosci. 6:1–7.
Meiran N, Pereg M, Kessler Y, Cole MW. 2014. The power of instructions: proactive configuration of stimulus—response translation. J Exp Psychol Learn Mem Cogn. 41:768–786.
Miller EK, Cohen JD. 2001. An integrative theory of prefrontal cortex function. Ann Rev Neurosci. 24:167–202.
Miller EK, Erickson CA, Desimone R. 1996. Neural mechanisms of visual working memory in prefrontal cortex of the macaque. J Neurosci. 16:5154–5167.
Milner B. 1963. Effects of different brain lesions on card sorting. Arch Neurol. 9:90–100.
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.
Petrides M. 1985. Deficits on conditional associative learning tasks after frontal- and temporal –lobe lesions. Neuropsychologia. 5:601–614.
Rainer G, Asaad WF, Miller EK. 1998. Selective representation of relevant information by neurons in the primate prefrontal cortex. Nature. 393:577–579.
Ramamoorthy A, Verguts T. 2012. Word and deed: a computational model of instruction following. Brain Res. 1439:54–65.
Ruge H, Wolfensteller U. 2013. Functional integration processes underlying the instruction-based learning of novel goal-directed behaviors. Neuroimage. 68:162–172.
Ruge H, Wolfensteller U. 2010. Rapid formation of pragmatic rule representations in the human brain during instruction-based learning. Cereb Cortex. 20:1656–1667.
Saxe R, Brett M, Kanwisher N. 2006. Divide and conquer: A defense of functional localizers. Neuroimage. 30:1088–1096.
Serences JT, Ester EF, Vogel EK, Awh E. 2009. Stimulus-specific delay activity in human primary visual cortex. Psychol Sci. 20:207–214.
Sreenivasan KK, Curtis CE, D'Esposito M. 2014. Revisiting the role of persistent neural activity during working memory. Trends Cogn Sci. 18:82–89.
Sreenivasan KK, Vytlacil J, D'Esposito M. 2014. Distributed and dynamic storage of working memory stimulus information in extrastriate cortex. J Cogn Neurosci. 26:1141–1153.
Steiger JH. 1980. Tests for comparing elements of a correlation matrix. Psychol Bull. 87:245–251.
Stocco A, Lebiere C, O'Reilly RC, Anderson JR. 2012. Distinct contributions of the caudate nucleus, rostral prefrontal cortex, and parietal cortex to the execution of instructed tasks. Cogn Affect Behav Neurosci. 12:611–628.
Stokes MG. 2011. Top-down visual activity underlying VSTM and preparatory attention. Neuropsychologia. 49:1425–1427.
Stokes MG. 2015. 'Activity-silent' working memory in prefrontal cortex: a dynamic coding framework. Trends Cogn Sci. 19:394–405.
Stokes MG, Kusunoki M, Sigala N, Nili H, Gaffan D, Duncan J. 2013. Dynamic coding for cognitive control in prefrontal cortex. Neuron. 78:364–375.
Stokes MG, Wolff MJ, Spaak E. 2015. Decoding rich spatial information with high temporal resolution. Trends Cogn Sci. 19:636–638.
Todd MT, Nystrom LE, Cohen JD. 2013. Confounds in multivariate pattern analysis: theory and rule representation case study. Neuroimage. 77:157–165.
Townsend JT, Asby FG. 1983. Stochastic modelling of elementary psychological processes. Cambridge: Cambridge University Press.
Van Essen DC. 2005. A population-average, landmark—and surface-based (PALS) atlas of human cerebral cortex. Neuroimage. 28:635–662.
Wallis JD, Anderson KC, Miller EK. 2001. Single neurons in prefrontal cortex encode abstract rules. Nature. 411:953–956.
Walsh KW. 1978. Neuropsychology: a clinical approach. New York: Churchill Livingstone.
Waskom ML, Frank MC, Wagner AD. 2016. Adaptive engagement of cognitive control in context-dependent decision making. Cereb Cortex. doi: 10.1093/cercor/bhv333.
Waskom ML, Kumaran D, Gordon AM, Rissman J, Wagner AD. 2014. Frontoparietal representations of task context support the flexible control of goal-directed cognition. J Neurosci. 34:10743–10755.
Watanabe M. 1996. Reward expectancy in primate prefrontal cortex. Nature. 382:629–632.
Wenke D, Frensch PA. 2005. The influence of task instruction on action coding: constraint setting or direct coding? J Exp Psychol Hum Percept Perform. 31:803–819.
Wenke D, Gaschler R, Nattkemper D. 2007. Instruction-induced feature binding. Psychol Res. 71:92–106.
Wolfensteller U, Ruge H. 2012. Frontostriatal mechanisms in instruction-based learning as a hallmark of flexible goal-directed behavior. Front Psychol. 3:192.
Woolgar A, Golland P, Bode S. 2014. Coping with confounds in multivoxel pattern analysis: What should we do about reaction time differences? A comment on Todd, Nystrom & Cohen 2013. Neuroimage. 98:506–512.

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

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

Zanto TP, Rubens MT, Bollinger J, Gazzaley A. 2010. Top-down modulation of visual feature processing: the role of the inferior fronal junction. Neuroimage. 53:736–745.

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