# Encoding, preparation and implementation of novel complex verbal instructions

Carlos González-García[a,b], Juan E. Arco[a,b], Ana F. Palenciano[a,b], Javier Ramírez[c], María Ruz[a,b,*]

[a] *Mind, Brain and Behavior Research Center, University of Granada, 18071 Granada, Spain*
[b] *Department of Experimental Psychology, 18071 Granada, Spain*
[c] *Department of Signal Theory, Telematics and Communications, CITIC-UGR, University of Granada, 18071 Granada, Spain*

## ARTICLE INFO

## ABSTRACT

Verbal instructions allow humans to acquire and implement complex novel rules in few seconds. A major question that remains elusive is how the brain represents this information prior to successful task execution. In this experiment, we studied the brain regions involved in representing categorical stimulus information during the encoding of novel instructions, their preparation and also their implementation, as well as the relation of the fidelity of these representations to observable behavior. To do so, we devised a novel instructions paradigm to delimitate these three stages. Using univariate and multivariate analyses of functional magnetic resonance data, our study revealed that the semantic content (faces or letters) of complex novel instructions can be decoded several seconds before the onset of a target, as soon as instructions are encoded. Crucially, the quality of the information represented in domain-general and category-selective regions correlated with subsequent behavioral performance. This suggests that the rapid transformation of novel instructions into coherent behavior is supported by control mechanisms that use available, relevant information about the current rule prior to its execution. In addition, our results highlight the relation between these control processes and others such as prospective memory and maintenance of future intentions.

## Introduction

The ability to implement verbal instructions allows humans to translate novel complex rules into behavior in mere seconds. How does the brain deal with new information in such a fast and efficient way? According to theoretical models, the path from instructions to overt behavior can be decomposed in different stages of processing (Bunge, 2004; Sakai, 2008). Initially, the content of the instructions has to be *encoded* in the system, employing representations of semantic rules that link specific stimulus features to concrete behaviors (Crone et al., 2006; Sakai, 2008). Once the target context (stimuli) appears, instructions are *implemented* by performing the appropriate actions according to the instructed rules. But before that, *preparation* entails a task set configuration (Meiran, 1996; Rubinstein et al., 2001). This stage of processing, understood as the adjustment to relevant task rules in anticipation of target stimuli (Rogers and Monsell, 1995) is a key component of complex task execution (Brass and von Cramon, 2002), and it is thought to be a cognitive state separable from related ones, such as the mere maintenance of task demands (Cohen-Kdoshay and Meiran, 2009; Liefooghe et al., 2013, 2012; Muhle-Karbe et al., 2014).

Crucially, the adequate configuration of cognitive resources achieved during the preparatory period enhances behavioral performance (Sakai, 2008).

Neuroimaging data suggest that task preparation relies on a frontoparietal network, which has been related to complex cognitive control operations (Duncan, 2010). More specifically, when switching between tasks the Inferior Frontal Junction (IFJ) updates the rule representation, whereas stimulus-response associations engage the intraparietal sulcus (IPS; Brass and von Cramon, 2004, 2002). De Baene and Brass (2014) proposed that the pre-supplementary motor area (pre-SMA) suppresses actions from previous tasks and enhances the appropriate response for the new stimulation. Previous studies also point to a dynamic interplay between the lateral prefrontal cortex (LPFC) and brain regions linked to category-specific processing (Sakai and Passingham, 2006, 2003). For instance, when participants prepare to perform semantic operations, task cues engage areas involved in effortful semantic processing, such as the left inferior frontal gyrus, prior to the target onset (e.g. González-García et al., 2016). Some other studies have also reported category-specific connectivity patterns in absence of changes in activity (Sakai, 2008). Although preparation

* Correspondence to: Department of Experimental Psychology, University of Granada, 18071 Granada, Spain.
  *E-mail address:* mruz@ugr.es (M. Ruz).

seems to play even a more important role when we have to implement novel instructions (Norman and Shallice, 1986), the neural architecture supporting this ability remains unclear.

Novelty brings larger requirements of control processes given the lack of pre-existing task-rule representations (Cole et al., 2013). However, in most studies of task preparation and switching, participants alternate between a small number of highly practiced tasks, and thus task sets are formed in advance and later retrieved from memory (Cole et al., 2013). Preparatory processes, however, seem to fulfill somewhat different demands with new tasks. When we face a novel situation, retrieval of previous full task sets does not suffice. Rather, new ones have to be generated from scratch: representations (e.g. of visual cues) have to be created for each new trial, including sensory, semantic and goal-related ones. Recent research has coined the term of Rapid Instructed Task Learning (RITL) to refer to the "ability to rapidly restructure one's behavior into novel configurations from instructions" (Cole et al., 2013). To differentiate this type of learning from others, such as trial-and-error learning, most of the RITL research focuses on the first time a given task set is presented. This strategy uses several new instructions together with practiced ones, which allows the comparison between the retrieval of previously practiced sets and the actual formation of novel ones. Some studies (Cole et al., 2010; Ruge and Wolfensteller, 2010) have assessed the *implementation* of novel instructions, suggesting a novelty-related gradient within the LPFC in which there is an anterior-to-posterior shift of activation as task sets transition from novel to practiced. Regarding *preparation* for novel tasks, the frontoparietal network also seems to be involved. For instance, Hartstra et al. (2011) highlighted the involvement of IFJ and IPS during the presentation of single words that encoded instructions. They also revealed the involvement of the dorsal pre-motor and M1 areas, a result that they interpreted as a correlate of motor imagery and the creation of memory codes for the instructions. Using also isolated words, a later study extended these results, revealing that the inferior frontal sulcus underpins the representation of task sets by creating a link between the stimulus and the motor response (Hartstra et al., 2012). Despite this suggestive evidence, these studies did not explore how the semantic content of the instructions was represented during their encoding and preparation. This is a crucial aspect, since relevant information is needed to create an accurate task configuration prior to task execution, as shown in task-switching. However, it is still unknown how and where this relevant information is represented during preparatory stages to support the translation of novel verbal representations into implemented rules.

Studies assessing the representation of specific information during stages of processing suggest that multivariate rather than univariate analysis of functional magnetic resonance images (fMRI) is more suitable for this matter. For instance, using single words as instructions, Cole et al. (2011) showed that a classifier could generalize to novel tasks when trained on practiced tasks with DLPFC fMRI patterns. Likewise, a classification algorithm was able to decode task sets within this region while the instructions were on the screen. However, there is an implicit difficulty to decode relevant preparatory information in classic RITL paradigms. Since participants have to encode a new instruction on every trial, it is hard to disentangle the activity specific to preparatory processes from encoding of the new verbal content of the instructions (Brass and von Cramon, 2002; Rogers and Monsell, 1995).

We designed a paradigm to isolate preparation from instruction encoding and later implementation. For this, we pseudorandomly manipulated the duration of the interval between novel instructions and novel target grids, and had participants prepare the novel instructions in only half of the trials. Similar strategies have been used to isolate the preparation component associated to novel instructions (Demanet et al., 2016). In addition, we employed complex verbal instructions rather than drawings or isolated words as used in previous studies (Cohen-Kdoshay and Meiran, 2009; Liefooghe et al., 2013,

2012). Although instruction implementation can be achieved via non-linguistic channels, verbal information is the most powerful means to convey novel instructions (Cole et al., 2013). Previous strategies, such as presenting novel symbols or images linked to specific responses, promote concrete stimulus-response pairings and visual imagery, which reduce the scope of the observations. Also, the instructions employed in the current study contained abstract rules, which entail less concrete commands than specific instructions (e.g. "If you see two squares, press A") and a larger number of potential perception-action scenarios (e.g. "If you see two vowels, press A"). Our verbal instructions posed yet another form of complexity as they allowed the combination of multiple rules (e.g. "If you see two contiguous green vowels of the same size, press A"), which is not easily achievable through non-linguistic instructions or single words.

In addition, our experiment employed a task designed to alleviate frequent confounds of task novelty and difficulty. As mentioned before, previous studies employ instructions practiced in advance as a contrast to novel ones (Cole et al., 2016, 2011; Stocco et al., 2012). However, this translates into practiced tasks that are easier to implement, since the mere presentation of the instruction elicits the adequate response, retrieved from memory, which leads to faster responses and higher accuracy scores. In our study, in contrast, every trial started with a *novel* complex verbal instruction. A subsequent cue indicated whether a novel or a practiced target grid would appear and prompted participants either to prepare to implement the previous instruction or to retrieve a response from memory upon later target presentation (see Section Design and procedure). This manipulation balanced demands across tasks and equated performance indexes across novel and practiced sets. Also, a secondary benefit was the increase in the number of novel instructions, which increased the power of the design to differentiate the encoding of instructions referring to different stimulus categories.

In sum, the main aim in our study was to advance our knowledge about how the brain uses new complex information to perform novel tasks. To do so, we used complex, fully grammatical verbal instructions referring to either faces or letters to assess which areas contained category-specific information during the encoding, preparation and implementation of novel rules. We predicted that partially differentiated patterns of regions would be involved in encoding vs. preparing for a new instruction, and that these would include areas related to cognitive control. Similarly, we expected that the semantic content of instructions would be decodable since their encoding, but that more regions would get involved when participant had to explicitly prepare, reflecting a finer tuning to relevant task information. In addition, we hypothesized that the degree of decodability of activity patterns of different categories would have a relation with observable performance, which would stress the relevance of these representations for actual behavior.

## Materials and methods

### Participants

Twenty-two students from the University of Granada (7 males; mean age: 23; range: 19–31) took part in the experiment and received 20€ in exchange. To encourage high performance during the task, participants were informed that the five of them with the highest scores (in terms of accuracy and reaction times) would receive 5 additional €. All participants reported normal or corrected-to-normal vision and no history of neurological disorders, and signed a consent form approved by the local Ethics Committee.

### Apparatus and stimuli

We created an initial pool of 210 different verbal instructions that referred to either face or letter-related features of grids of stimuli. Face
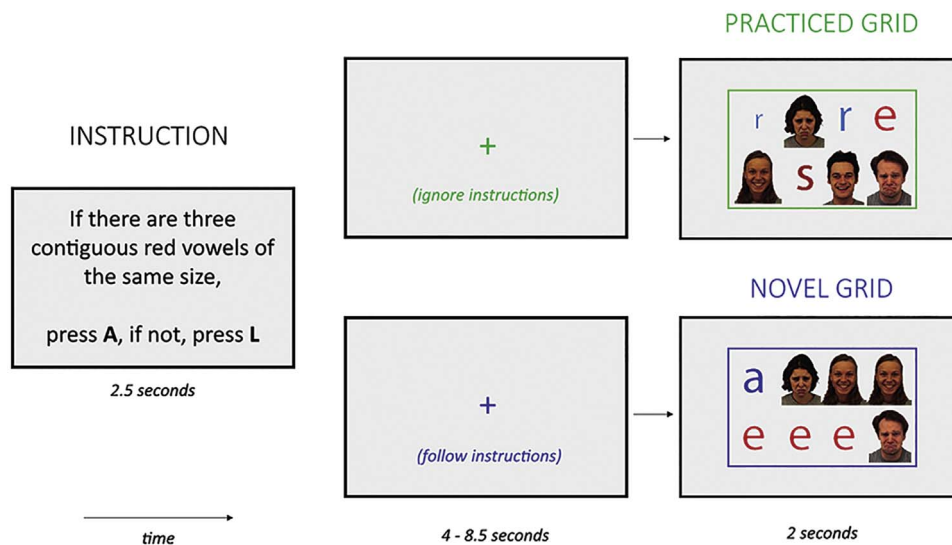
**Fig. 1.** Behavioral paradigm.

features were gender (male, female), emotion (happy, sad), size (large, small) amount (one, two, three or four faces) and contiguity (two, three or four contiguous faces). To create similar instructions for letter features, we equated face gender with type of letter (consonant, vowel) and face emotion with letter color (red, blue). Size, amount and contiguity were used in the same manner as in face-related instructions. Depending on the instruction, participants had to focus on one or several features of the same stimulus category. Instructions always used a conditional structure, such as "If there are three contiguous [e.g. female faces/red vowels] of the same size, press A, if not, press L". Two unique grids of 4 faces and 4 letters (one fulfilling the instruction and the other one not) were created for each of the 210 instructions. Grids were drawn from a pool of 8 faces (one happy male, one sad male, one happy female, one sad female; each in large and small sizes) and 8 letters (one blue consonant, one red consonant, one blue vowel, one red vowel; each in large and small sizes). To equate displays across categories, we created the equivalent of each grid for the opposite category, following the equivalences described above (e.g. gender [male, female]=type [consonant, vowel]).

A preliminary behavioral pilot study (*n*=14) detected that 17 instructions yielded consistently low accuracies (≤75%), and thus these were removed from the instruction pool together with their equivalent counterparts in the opposite category. The final pool contained 176 instructions (88 faces, 88 letters) equated in difficulty (see Supplementary Table 1).

For the fMRI experiment, a PC running E-Prime 2.0 projected the stimuli onto a screen located at the back of the scanner. A set of mirrors mounted on the head coil was used for participants to see the screen. During the whole experiment, they used the index fingers of their right and left hand to make speeded discrimination responses to target grids by pressing one of two buttons on a MRI-compatible button box.

### Design and procedure

To create a baseline to compare against novel trials that was equated in behavioral difficulty (as assessed by speed and accuracy of responses), the day before scanning participants practiced a set of eight instructions (half referring to faces and the other half to letters; extracted randomly without replacement from the instructions pool) and their associated target grids. This practice session lasted ~20 min (8 presentations of each instruction) and was repeated until participants achieved ≥90% of accuracy. To ensure that participants had learnt the response to the eight grids, afterwards they were prompted to answer to the same grids, this time *without* seeing the associated

instructions in advance. Therefore, by the end of the practice session participants had learned both grid-response and instruction-grid associations. This session also repeated until participants achieved ≥90% of accuracy, and lasted ~15 min (8 repetitions of each grid). If participants failed to complete the last phase after three repetitions, they had to repeat the complete learning session again (which happened to 3 of them). On average, participants practiced each instruction 20.6 times (SD=6.42), during approximately 43 min. Before the scanning session (which usually took place the day after the practice and never more than 9 days apart, with an average interval of 1.86 days), participants performed a ~5 min behavioral rehearsal session in which they again had to respond to the eight grids without their corresponding instruction. During the entire learning session, feedback was given after each trial to make participants aware of their performance and help them improve.

Scanning comprised a total of 160 trials. In each of these, regardless the type of trial, a novel instruction appeared. The color of the subsequent fixation cross (blue or green) signaled whether participants had to follow the instruction (80 trials) and thus prepare to implement it with a novel grid of stimuli or, alternatively, whether they had to ignore it (80 trials) and expect one of the eight practiced grids. For these practiced grids, participants had to respond based on the knowledge acquired during the learning session (see Fig. 1).

The associations between type of trial (novel, practiced), category and response options were counterbalanced across participants. The duration of the fixation cross indicating the type of trial, as well as inter-trial intervals, were jittered to allow the deconvolution of instruction- and grid-related signals. The pseudorandom duration of the preparation interval allowed the disambiguation of this stage from the encoding and implementation. Each trial comprised the following events (see Fig. 1): a 2.5 s instruction, a colored fixation cross (mean 6.25 s, range 4–8.5 s), a 2 s grid and an inter-trial interval displaying a black fixation cross (mean 6.25 s, range 4–8.5 s). On average, a trial lasted 10.750 s. The total fMRI task lasted 45 min approximately.

### Data acquisition

Magnetic resonance images were acquired using a 3 T Siemens Trio scanner at the Mind, Brain and Behavior Research Center (CIMCYC) in Granada (Spain). Functional images were obtained with a one-shot T2*-weighted echo planar imaging (EPI) sequence (time until echo [TE]=23 milliseconds (ms), flip angle=70 degrees, repetition time [TR] =2.21 s). Forty descending sagittal slices with a thickness of 2.3 mm (mm; gap of 20%) covered the entire brain (voxel size of $3 \times 3 \times 3$ mm$^3$).

The event-related experiment was performed in a run consisting of 1240 volumes. In addition, we acquired a standard structural image of each participant using a high-resolution T1-weighted sequence (TR=1900 ms; TE=2.38 ms; $1 \times 1$ mm$^2$ in-plane resolution and 1 slice thickness).

We used SPM8 (http://www.fil.ion.ucl.ac.uk/spm/) to preprocess and analyze the neuroimaging data. The first 4 volumes were discarded to allow for saturation of the signal. The images were then realigned and unwarped using a least-squares approach and a six-parameter (rigid body) spatial transformation to correct for motion artifacts. Then we used slice timing correction to account for differences in the time of slice acquisitions. Afterward images were normalized to the standard EPI template included in SPM8 and spatially smoothed using an 8 mm full-width at half-maximum isotropic Gaussian kernel. A 128 s high-pass filter was used to remove low-frequency artifacts.

*Univariate analyses*

Statistical analysis was performed with a General Linear Model (GLM) for each participant with corrections for serial autocorrelations using the AR(1) model. The model included regressors for Encoding (faces/letters instructions), Preparation (jitter of novel faces/letters task; jitter of practiced task), and Implementation (novel grid of faces/letters task; practiced grid faces/letters task). These regressors were convolved with the standard hemodynamic response function. Duration and onset vectors for the preparation interval were introduced into the GLM, whereas instructions and grids were modeled as events with zero duration. Trials with errors and missing responses were grouped together as separate events with an extended duration for the whole trial (encompassing instructions, fixation cross and grid). Contrasts of interest (i.e. Encoding vs. Baseline; Novel preparation vs. Practiced preparation; Novel implementation vs. Practiced implementation; Faces vs. Letters encoding; Faces vs. Letters novel preparation; Faces vs. Letters novel implementation) were obtained for each participant and then entered into a second-level analysis, where a $t$-test ($t_o$) was used to contrast conditions. To assess significance of each contrast at the population level, a non-parametric permutation-based approach was followed (Eklund et al., 2016), using the Statistical non-Parametric Mapping toolbox (SnPM13; http://warwick.ac.uk/snpm). On each permutation, the signs of the individual scores were randomly flipped and a new $t$-test was performed. This was repeated 5000 times, obtaining the true distribution of $t$-values, to which $t_o$ was empirically comparable. Cluster-wise inference with a cluster-defining threshold of $p < 0.001$ was later used to find significant clusters (FWE corrected, $p < 0.05$) on the resulting map. It is worth noting that on this as well as on the analyses that follow, results were almost identical to those obtained using a parametric cluster-wise FWE correction approach (obtained from an initial uncorrected $p < 0.001$) as implemented in SPM12.

After the identification of areas involved in the encoding and preparation stages, we conducted a conjunction analysis (Nichols et al., 2005) to look for brain regions shared by the two stages. For conjunction analyses, we performed one-way ANOVAs using first-level contrast images of interest, which allowed us to define which clusters were significantly active both during encoding and preparation. Only conjunction clusters surviving a MS/CN test (Minimum Statistic compared to the Conjunction Null; Nichols et al., 2005) are reported.

*Multivariate analyses*

We employed multivariate pattern analysis (MVPA) to study brain regions sensitive to different stimulus categories (face vs. letter) during the encoding, preparation and implementation of novel instructions. This decoding was performed on the non-normalized and non-smoothed images. For each participant, we used a Least-Squares Separate model (LSS; Mumford et al., 2012; Turner et al., 2012) to

reduce collinearity between the BOLD signal of consecutive events (Abdulrahman and Henson, 2016). Following this method, on each trial we fitted the standard hemodynamic response function to two regressors: (1) one of the events of the trial (e.g. novel grid of faces) and (2) the rest of events and trials. The output of this model was one beta image per event. To maximize the independence of training and test sets (Pereira et al., 2009), the total amount of betas obtained with the LSS model was split in 8 temporally distant chunks, each of which contained 5 betas per event of interest. For instance, for a given event (e.g. novel grid of faces), we grouped together the five betas closest in time (Turner et al., 2012). It is important to note that although these events were not always consecutive, since the different conditions were randomly interleaved, each chunk contained the trials closest in time. On average, each of these chunks comprised 340 s, being the average separation of train and test sets 510 s, and the distance between the two closest chunks 170 s.

To avoid biases in the selection of regions of interest (Kriegeskorte et al., 2009) we employed a searchlight approach (Kriegeskorte et al., 2006) across the whole brain. We used The Decoding Toolbox (Hebart et al., 2015) to create a spherical cluster of 4-voxel radius around a given voxel $V_1$, containing $C_{1...N}$ voxels. In order to cross-validate the performance of the decoder, accuracy was estimated following a leave-one-out scheme, with each chunk acting as test set once (Pereira et al., 2009). On each fold, a linear support vector machine ($C=1$) was trained to classify the patterns of each category in 7 of the 8 chunks (training set). Then, the algorithm's decoding accuracy was tested in the remaining chunk (test set). The averaged accuracy (see Results section) revealed the algorithm's ability to classify spatial patterns of each category in the cluster centered in $V_1$. This procedure was repeated for $V_{2...N}$ in each participant and for each event (encoding, novel preparation and novel implementation). The resulting accuracy maps for each participant and event were then normalized to a standard EPI image and smoothed with a 3-mm Gaussian kernel. The smoothed images were entered in a second-level analysis. Statistical correction was performed using the non-parametric approach described in the GLM section. The surviving clusters localized brain areas in which the algorithm's accuracy was above chance, that is, areas in which the spatial patterns for faces and letters were significantly different.

*Correlations between decoding accuracies and behavioral indices*

To test the behavioral relevance of the decoding results, we conducted a correlation analysis between individual behavioral indices (average reaction times [RTs] and accuracy on novel instruction trials) and decoding accuracies from significant clusters in the searchlight, using a permutation approach to correct for multiple comparisons (Groppe et al., 2011). For a given comparison between the behavioral index and a significant cluster across participants, we first calculated a correlation index, $r_{observed}$. We then shuffled participants' scores within one of the variables and performed the correlation test again. We carried out this step 10,000 times for each comparison, obtaining a null distribution of random $r_N$ values with the null hypothesis that the two variables were not correlated (i.e. no relationship between the decoding accuracy of a given cluster and the behavioral index). To prove this hypothesis wrong, the $r_{observed}$ should be greater than the 95% of $r_N$. Only correlations with $p$-values smaller than 0.05 are reported.

In addition, to rule out the potential confound of RT-related variance in the hemodynamic response, a control analysis was performed. Here, we estimated the betas for decoding again, but regressing out the specific RT of the trial to which the event belonged. We then performed the decoding with the new betas and the correlation between the decoder accuracy and behavioral measures. This approach yielded significant correlations in the same ROIs revealed by the original analysis.
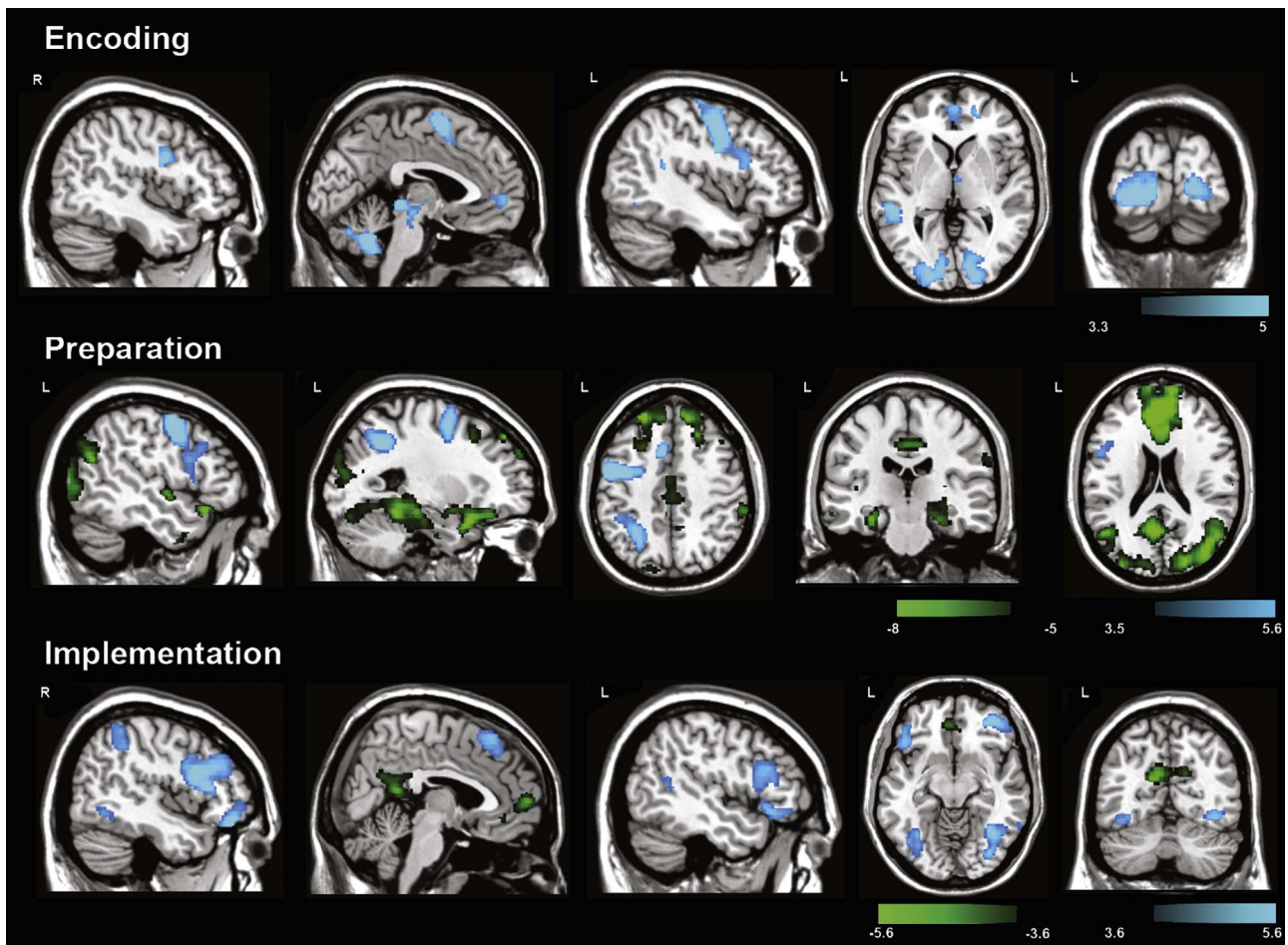
**Fig. 2.** Univariate results. GLM results for the encoding (novel instruction vs. baseline), preparation (follow instructions vs. ignore instructions) and implementation (novel grid vs. practiced grid). Blue colors indicate significant clusters for novel instructions, whereas green represents activity for the practiced condition. Scales reflect peaks of *t*-values.

## Results

### Behavioral

Differences between novel and practiced trials were not significant ($F < 1$ for accuracy scores and $F=1.33$, $p=.26$ for RTs). No significant differences were found between faces and letters for accuracy scores ($F < 1$), whereas there was a marginally significant effect for RTs ($F=4.12$, $p=.06$). These were nominally longer for faces ($M=1109.27$ ms; $SEM=24.73$ ms) than for letters ($M=1067.42$ ms; $SEM=25.77$ ms). Additionally, there were no significant interactions (all $ps > .1$).

### Univariate

We first looked for activity across categories in novel trials, along the three temporal events of interest (encoding, preparation and implementation of instructions; see Fig. 1). The GLM during encoding (see Fig. 2) yielded significant clusters of activity (Instructions vs. baseline) in the ventral rostral prefrontal cortex (RPFC; *x*, *y*, *z* MNI coordinates of peak voxel: 2, 54, 0; *k* (cluster size)=91 voxels), superior medial frontal gyrus (MFG; 12, 45, 28; *k*=147), bilateral inferior frontal gyrus near the IFJ (42, −2, 24; *k*=138; −34, 14, 20; *k*=112), pre-SMA (−6, 0, 64; *k*=520), left premotor cortex (PMC; −48, −2, 36; *k*=519), left middle temporal gyrus (MTG; −54, −36, 2; *k*=480), right and left precuneus (28, −38, 22; *k*=27; −28, −54, 14, *k*=55), bilateral visual association areas (20, −94, −2; *k*=99; −16, −94, −8; *k*=123) and cerebellum (0, −58, −32; *k*=289).

During preparation, the contrast of novel against practiced trials showed a different pattern of activations. Preparing to perform a novel instruction engaged the left inferior frontal gyrus (−44, 22, 32; *k*=331), pre-SMA (−6, 2, 62; *k*=733), left PMC (−44, 0, 42; *k*=900) and the left inferior parietal lobe, near the IPS (−28, −52, 38; *k*=206). The opposite contrast revealed the strong involvement of default mode network (DMN) regions, including RPFC (0, 44, 22; *k*=3570), posterior cingulate cortex (PCC) (−4, −60, 22; *k*=3281), and left (−48, −62, 22; *k*=764) and right angular gyrus (60, −52, 34; *k*=830).

During the implementation of instructions, responses to novel grids (vs. practiced ones) revealed the activation of the ventrolateral prefrontal cortex bilaterally (48, 42, −16; *k*=74; −46, 28, −12; *k*=136), bilateral dorsolateral prefrontal cortex, including the IFJ, (50, 15, 16; *k*=643; −50, 20, 20; *k*=325), the right inferior parietal lobe (32, −52, 48; *k*=799), left MTG (−60, −52, 2; *k*=364) and the bilateral fusiform gyrus (34, −58, −10; *k*=210; −34, −62, −10; *k*=131). The opposite contrast yielded the activation of part of the DMN, namely, the RPFC (−2, 56, 6; *k*=536) and precuneus (−6, −48, 12; *k*=783).

However, no clusters survived the statistical threshold when contrasting faces vs. letters during encoding and preparation for novel trials. During implementation, only a cluster located in visual areas near BA18 (25, −92, −2; *k*=136) was more active for face than for letter grids.

### Conjunction analysis

A one-way ANOVA with both the encoding and preparation regressors revealed a significant cluster involving the left inferior
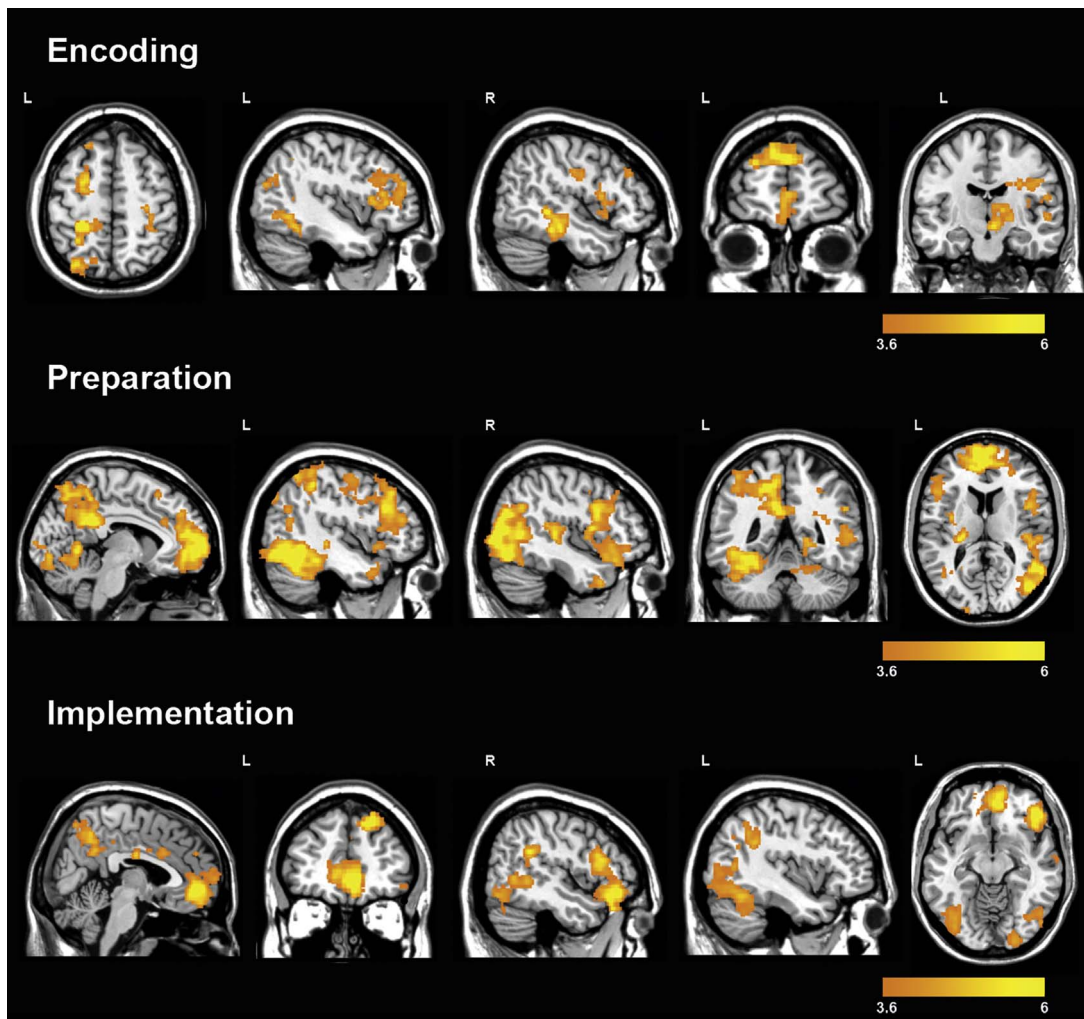
**Fig. 3.** Multivariate results. Differentiated spatial patterns for stimulus target categories (faces vs. letters) during encoding, preparation and implementation of novel instructions.

frontal gyrus (−42, 16, 24) and PMC (−40, −2, 44), and an additional cluster in the pre-SMA (−4, 2, 64). These results suggest that these three regions were active during both encoding and preparation of novel instructions. According to this results, the IPS seems to be the only region active exclusively during the preparation stage. To confirm the absence of IPS involvement during instruction encoding, we performed a region of interest-based comparison of instructions vs. baseline for the IPS cluster drawn from the preparation contrast. This did not yield significant results even after lowering the threshold at uncorrected $p < .1$.

*Multivariate*

While the instructions were on the screen, several regions showed differentiated spatial patterns of activation for encoding letter vs. face-related verbal instructions (see Fig. 3), including the ventral RPFC (−2, 58, −6; $k$=711; 54%), superior MFG (−2, 56, 28; $k$=546; 53.5%), left superior frontal gyrus (−24, 0, 56; $k$=135; 53.2%), left (−28, 32, 32; $k$=720; 54.1%) and right inferior frontal gyrus (48, 8, 12; $k$=41; 53%), right thalamus (16, −16, 6; $k$=352; 54%), left postcentral gyrus (−32, −38, 58; $k$=174; 53.3%), left superior parietal gyrus (−26, −42, 54; $k$=477; 54%), left angular gyrus (−36, −66, 26; $k$=65; 53.4%), left fusiform gyrus (−46, −58, −6; $k$=101; 53.4%), right lingual gyrus (18, −82, −10; $k$=388; 54.1%) and right cerebellum (28, −64, −42; $k$=567; 53.4%).

Preparation to perform either a letter or a face novel task elicited separable spatial patterns of activity in the ventral RPFC (2, 58, −2;

$k$=1244; 57.3%), a large cluster in the left dorsolateral prefrontal cortex including the IFJ (−30, 26, 50; $k$=2754; 56.2%), right inferior frontal gyrus (48, 10, 24; $k$=2191; 56%), left parahippocampal gyrus (−24, −14, −16; $k$=639; 55.5%), precuneus (−6, −44, 26; $k$=2953; 56.5%), right angular gyrus (52, −58, 14; $k$=351; 55.3%), left fusiform gyrus (−34, −52, −18; $k$=936; 57.5%), and right (46, −76, 8; $k$=1315; 57.1%) and left (−46, −78, −12; $k$=560; 56%) lateral occipital complex (LOC; see Fig. 3).

A conjunction analysis (Nichols et al., 2005) of the encoding and preparation stages revealed that the RPFC (0, 60, −2), left dorsolateral prefrontal cortex (−44, 24, 32), right inferior frontal (48, 8, 18), left superior parietal (−34, −76, 50) and left fusiform gyri (−50, −56, 10) represented relevant information in both stages of instructions processing.

Last, the stimulus category of novel tasks during implementation was encoded in the RPFC (6, 50, −4; $k$=545; 59%), superior frontal gyrus (18, 48, 52, $k$=343; 52.69%), right ventrolateral (42, 30, −16; $k$=261; 58.4%) and left dorsolateral prefrontal cortex (−46, 28, 24; $k$=196; 55.5%), anterior cingulate cortex (−8, 24, 24; $k$=186; 55.8%), right MTG (64, −10, −22, $k$=193; 55.5%), precuneus (−4, −64, 44; $k$=1885; 52.88%) and bilateral inferior parietal lobe (36, −50, 28; $k$=169; 55%; and −32, −56, 36; $k$=407; 56.4%, respectively).

*Correlations between decoding accuracies and behavioral indices*

We introduced the decoding accuracies of the peaks within significant clusters of the MVPA results and the average RTs and accuracy
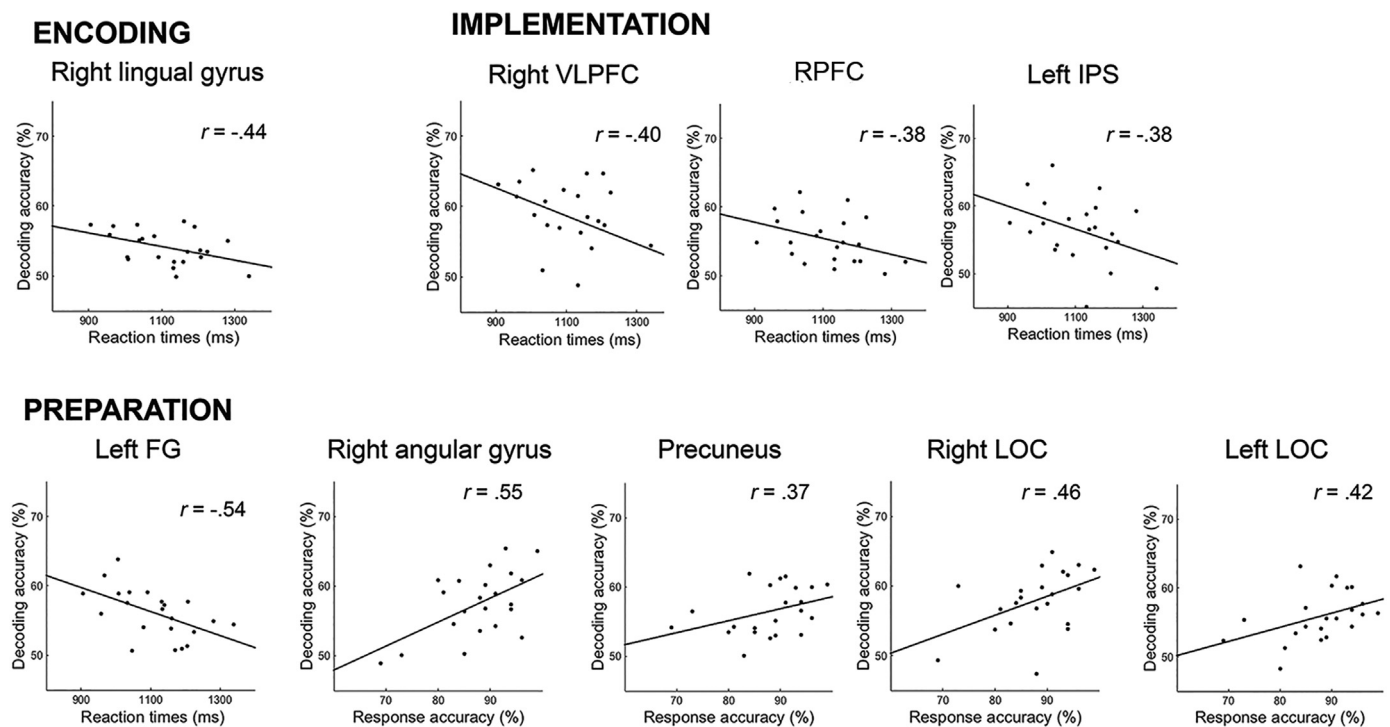
**Fig. 4.** Correlations between decoding and behavior. Scatter plots of significant correlations (corrected at $p < .05$) between decoding accuracies and behavioral indices of novel tasks. Initials stand for Fusiform Gyrus (FG), lateral occipital complex (LOC), ventrolateral prefrontal cortex (VPFC), RPFC (rostral prefrontal cortex) and intraparietal sulcus (IPS).

scores per participant into a correlation analyses. After correcting for multiple comparisons (see Section Correlations between decoding accuracies and behavior indices), this analysis revealed a significant negative correlation between the decoding of the instruction category (faces vs. letters) in the right lingual gyrus during the encoding period and RT ($r=-.44$, $p=.02$; see Fig. 4). During preparation for novel trials, RT negatively correlated with the decoding of the category in the left fusiform gyrus ($r=-.54$, $p=.005$). Moreover, during this stage response accuracy positively correlated with the decoding of the category in the precuneus ($r=.37$, $p=.04$), right angular gyrus ($r=.55$, $p=.005$), and right ($r=.46$ $p=.02$) and left ($r=.42$ $p=.02$) LOC. Regarding the implementation stage, RT negatively correlated with the decoding accuracy in the RPFC ($r=-.38$, $p=.04$), the right ventrolateral prefrontal cortex ($r=-.40$, $p=.03$) and the left IPS ($r=-.38$, $p=.04$).

## Discussion

In the present study, we assessed how the brain encodes the content of information conveyed by complex novel instructions. Using multivariate analysis, we have shown for the first time that the content of complex verbal instructions can be decoded during the encoding, preparation and implementation phases of novel tasks. Moreover, the decodability of category-specific information in high-order and selective processing brain regions during these stages has an impact on subsequent behavior.

As expected, our paradigm equated demands between novel and practiced tasks, as shown by behavioral measures, ruling out difficulty confounds. Moreover, performance was similar for faces and letters trials. Although we found a close to significance trend in RTs, a control analysis (see Section Correlations between decoding accuracies and behavior indices) confirmed that our decoding results, as well as the correlations of decoding accuracies with behavioral variables, were not affected by this speed of responses. The design of the paradigm also facilitated the separation of encoding, preparation and implementation stages. During the presentation of the verbal instructions on the screen, participants were asked to encode their content, which included the

perceptual category of the stimuli (faces vs. letters) that the instructions referred to. Crucially, at this stage, they did not know whether they would be required to later implement these instructions, which discouraged explicit preparation during this first, encoding period. The color of the subsequent fixation point carried this information, and in half of the trials participants had to prepare to perform the instruction just encoded, which would be implemented once the target grid appeared. On the other half of the trials, the color of the fixation point indicated that a practiced grid would appear and prompted participants to disregard the new rule and answer based on what they had learned during the practice session. This manipulation allowed us to study a component of explicit preparation, as well as differentiate it from the encoding of the verbal content, as discussed below. It is important to highlight that these instructions were complex abstract sentences, fully grammatical, rather than isolated words or pictures. Also, they prepared to respond to complex and variable target grids rather than to isolated stimuli. Hence, the results observed are hard to explain in terms of perceptual imagery.

Univariate results showed that encoding novel instructions engaged brain regions associated to instructed behavior, such as the IFJ, the pre-SMA and the PMC, which were also active during preparation. This suggests that even when participants do not know if they will be required to implement an instruction, its mere reading activates a new task set. This could reflect an automatic encoding of verbal instructions (Liefooghe et al., 2012) or, alternatively, a planned strategy of participants, by which they would willingly encode the new task set and decide whether to prepare or discard it later on. Either way, our results show larger involvement of the IPS during the preparatory stage compared to the encoding of instructions. This is in line with experimental models of executive control that propose that the activation of action-related codes occurs after completion of the task goal update (Rubinstein et al., 2001) and with empirical data that show the transmission of top-down representations from prefrontal to parietal neurons (Crowe et al., 2013, but see Bode and Haynes, 2009, for an alternative claim). Despite other regions related to action processing were active during the encoding phase, these have been previously

related to the suppression of previous task sets actions and the establishment of appropriate motor codes (De Baene and Brass, 2014; Hikosaka and Isoda, 2010). In general, our univariate results support the idea that instructions can foster the creation of S-R associations in a rather automatic manner (Liefooghe et al., 2012), but still a greater degree of preparation is needed for them to elicit action codes (Liefooghe et al., 2013; Meiran et al., 2012; Wenke et al., 2009). Similarly, Liefooghe et al. (2013) suggest that verbal instructions can be encoded in a declarative or a procedural format. According to these authors, only when the instruction has to be enacted in the future the declarative information is translated into an "action-based format". Wenke et al. (2009) propose that this translation entails the activation and binding of relevant features, which would take place during preparatory stages in our study. Likewise, Muhle-Karbe et al. (2014) used transcranial magnetic stimulation to show that the late disruption of IPS activity hindered the translation of abstract rules into specific motor commands. Hence, our results highlight the involvement of the IPS in binding stimulus and action features (Hartstra et al., 2012) and suggest a key role of this region while preparing to perform novel tasks. Last, during implementation, even with equated behavioral performance between novel and practiced trials, we found increased activation in lateral prefrontal and parietal cortices linked to novel targets. This result stresses the relevance of these regions in novelty processing, and not merely in more difficult contexts.

Univariate analyses, however, were not sensitive to the content of novel instructions. This lack of univariate sensitivity prior to target appearance is not uncommon in the literature (Sakai, 2008). Multivariate analyses, on the other hand, were more sensitive. During the encoding stage, content information could be decoded from high-order areas such as the IFJ, angular gyrus and the RPFC, as well as regions related to letter and word processing, such as the lingual (Borowsky et al., 2007; Leshikar et al., 2012; Vinckier et al., 2007) and fusiform gyri (Harris et al., 2016; McCandliss et al., 2003; Roberts et al., 2013). Our results show that these regions, which previous results link to lexical-semantic processing of sentences (e.g. Ye et al., 2011), are involved in encoding the semantic content of novel instructions. Crucially, prior to target onset, the content of instructions could also be decoded from category-selective regions, such as the fusiform gyrus or lateral occipital complex, key hubs for object processing (Eger et al., 2008a, 2008b; Grill-Spector et al., 2001). These results are coherent with the idea that prospective cognitive control processes engage not only higher-order, frontal areas but also specific processing regions involved in forthcoming stimulation (González-García et al., 2016; Sakai and Passingham, 2003). Previous studies that also reported results on the same line during preparatory stages (Muhle-Karbe et al., 2016) employed images of drawings to instruct new rules, which most likely engaged subsequent mental imagery retrieval of specific perceptual material. Our design, in contrast, employed complex verbal instructions composed of words and with no pictures, and therefore the information represented in perceptual regions had to be activated by the semantic content of the instructions *per se*, rather than the perceptual processing of images. Last, preparation also engaged a set of regions that has been previously related to episodic retrieval and recombination of related past events, such as the RPFC, PCC, parahippocampal cortex and angular gyrus. Interestingly, some of these regions are part of the DMN. We will return to this point later on the Discussion.

An important finding of the present study is the observed link between the quality of the representations during different stages of novel instruction processing and behavior. This relationship between decoding efficiency in novel tasks and behavior has been reported in previous studies. For instance, Etzel et al. (2015) revealed an increase in the quality of rule representations in the brain when participants received monetary incentives. Similarly, Cole et al. (2016) showed that task representations within the DLPFC are behaviorally relevant in the *implementation* of novel tasks, since the accuracy of the decoder

increased during correct in comparison with error trials. Results from the current experiment show that the discriminability of the perceptual categories referred to during the *encoding* and *preparation* of novel instructions also influence the efficiency of posterior behavioral responses. Interestingly, while during encoding and preparation there were significant correlations between behavior and decoding in some brain regions related to selective processing of objects, such as the lingual gyrus, fusiform gyrus and lateral occipital complex, during implementation these correlations were found in frontoparietal areas involved in cognitive control. Previous studies suggest that the IFJ drives category-specific regions involved in feature-based attention (Baldauf and Desimone, 2014). In this line, we hypothesize that before target onset, the updating of the task rule in the IFJ (Brass and von Cramon, 2004, 2002, Hartstra et al., 2012, 2011), replicated by our univariate analysis, is followed by the tuning of object processing areas such as the fusiform gyrus and the lateral occipital complex, as well as areas involved in episodic retrieval and recombination of past events (Lundstrom et al., 2005; Wagner et al., 2005). This tuning would enhance the integration of patterns of similar instructions and separation of irrelevant ones, which could explain the reported impact on subsequent behavior. During the implementation of the rule, this simultaneous pattern integration and separation would have the largest effect on cognitive control areas (Schlichting and Preston, 2015). Interestingly, decoding accuracy correlated with RT in some areas and with response accuracy in other areas, but not with both indexes. It is also noticeable that correlations with response accuracy were only found during preparation, whereas significant correlations with RT were present during the three phases. This set of results is however puzzling as previous literature did not led us to predict such differences, which would be interesting to explore in future studies.

Our multivariate results are coherent with the existence of a compositional mechanism underlying the ability to follow novel instructions to implement new tasks (Cole et al., 2013). This notion resonates with the *constructive episodic simulation hypothesis*, proposed in the field of prospective memory. This hypothesis predicts that the simulation of novel future events relies on a flexible recombination of small details of past events (Madore et al., 2014; Schacter and Addis, 2009; Szpunar et al., 2014), which is in line with the idea that instructed learning takes advantage of working memory resources to support "rapid updating, composionality, and combinatorics of the representations within the task sets" (Cole et al., 2013). The recombination of past events recruits different brain areas, including not only the prefrontal cortex, but also the RPFC, lateral temporal and temporopolar cortex, hippocampus, parahippocampal cortex, lateral parietal and PCC (Schacter and Addis, 2009). Some of these areas have a direct relation with behavior during constructive simulation of novel future events, suggesting their important role in our task and paradigms alike. Accordingly, some authors have proposed that rule and motor representations of novel instructions might entail mental simulation (Brass et al., 2009). Similar ideas have been suggested in the study of goal setting and intentions. Locke and Latham (2002) proposed that, when confronted with new stimulation, people retrieve a repertoire of skills used in similar contexts and apply them to attain a novel goal. Moreover, Lau et al. (2004) revealed that attending to the intention to perform a motor action involved brain regions closely related to novel instruction processing, such as the pre-SMA and the IPS. Several prospective memory studies assessing the correlates of future intentions have shown the involvement of brain areas recruited by our task, mainly the ventral RPFC (Gilbert, 2011; Landsiedel and Gilbert, 2015; Momennejad and Haynes, 2013, 2012). A related striking finding of the current dataset is the encoding of semantic information in DMN regions. These results hardly reflect mind wandering, since participants had to actively retrieve the response to the given target, or some sort of processing of social information, since the required responses were based on perceptual rather than social features. Although the role of these regions in task setting is not clear,

one possibility is that these areas are involved in the representation of the semantic content of internal speech and conscious thought (Huth et al., 2016). Accordingly, previous studies have highlighted the involvement of the DMN when large changes of the cognitive context take place (Crittenden et al., 2015), which potentially require prospective memory and active intentions. Also demanding prospective memory is the coding of task goals, which has been related to the frontal node of the DMN (Haynes et al., 2007) as well. Our results thus suggest that novel instruction processing is supported by some of these processes (such as internal speech, major revisions of cognitive context, recombination of past events, simulations of future events, encoding of intentions and attention to these). More research is needed to delimit the specific role of each of these processes in the implementation of instructions.

*Conclusions*

In sum, our results reveal that category-specific information of complex verbal sentences instructing novel tasks can be decoded from several brain regions. Moreover, we show for the first time that this information can be decoded starting from the encoding of verbal information in perceptual and semantic-related brain areas. Future research should address how the specific neural representation of the content of instructions varies across this large set of areas during different stages, as well as the associated pattern of connectivity. Moreover, future studies should aim at assessing different aspects of instructions, such as the level of compositionality of rules, abstraction, and relationships between concepts. This would provide useful insight about how specific semantic information is encoded in the brain (Huth et al., 2016). A further important issue to be resolved is how different cognitive control components, namely the maintenance of an overall task-set and adaptive task sets initiated on each trial, interact when we follow complex verbal instructions.

## Appendix A. Supplementary material

Supplementary data associated with this article can be found in the online version at http://dx.doi.org/10.1016/j.neuroimage.2017.01.037.

## References

Abdulrahman, H., Henson, R.N., 2016. Effect of trial-to-trial variability on optimal event-related fMRI design: implications for Beta-series correlation and multi-voxel pattern analysis. Neuroimage 125, 756–766. http://dx.doi.org/10.1016/j.neuroimage.2015.11.009.

Baldauf, D., Desimone, R., 2014. Neural mechanisms of object-based attention. Science 344, 424–427. http://dx.doi.org/10.1126/science.1247003.

Bode, S., Haynes, J.-D., 2009. Decoding sequential stages of task preparation in the human brain. Neuroimage 45, 606–613. http://dx.doi.org/10.1016/j.neuroimage.2008.11.031.

Borowsky, R., Esopenko, C., Cummine, J., Sarty, G.E., 2007. Neural representations of visual words and objects: a functional MRI study on the modularity of reading and object processing. Brain Topogr. 20, 89–96. http://dx.doi.org/10.1007/s10548-007-0034-1.

Brass, M., von Cramon, D.Y., 2004. Decomposing components of task preparation with functional magnetic resonance imaging. Cogn. Neurosci. J. 16, 609–620.

Brass, M., von Cramon, D.Y., 2002. The role of the frontal cortex in task preparation. Cereb. Cortex 12, 908–914. http://dx.doi.org/10.1093/cercor/12.9.908.

Brass, M., Wenke, D., Spengler, S., Waszak, F., 2009. Neural correlates of overcoming interference from instructed and implemented stimulus-response associations. J. Neurosci. 29, 1766–1772. http://dx.doi.org/10.1523/JNEUROSCI.5259-08.2009.

Bunge, S.A., 2004. How we use rules to select actions: a review of evidence from cognitive neuroscience. Cogn. Affect. Behav. Neurosci. 4, 564–579.

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. http://dx.doi.org/10.1027/1618-3169.56.2.128.

Cole, M.W., Bagic, A., Kass, R., Schneider, W., 2010. Prefrontal dynamics underlying rapid instructed task learning reverse with practice. J. Neurosci. 30, 14245–14254. http://dx.doi.org/10.1523/JNEUROSCI.1662-10.2010.

Cole, M.W., Etzel, J.A., Zacks, J.M., Schneider, W., Braver, T.S., 2011. Rapid transfer of abstract rules to novel contexts in human lateral prefrontal cortex. Front. Hum. Neurosci. 5. http://dx.doi.org/10.3389/fnhum.2011.00142.

Cole, M.W., Ito, T., Braver, T.S., 2016. The behavioral relevance of task information in human prefrontal cortex. Cereb. Cortex 26, 2497–2505. http://dx.doi.org/10.1093/cercor/bhv072.

Cole, M.W., 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. http://dx.doi.org/10.3758/s13415-012-0125-7.

Crittenden, B.M., Mitchell, D.J., Duncan, J., 2015. Recruitment of the default mode network during a demanding act of executive control. Elife 4, e06481. http://dx.doi.org/10.7554/eLife.06481.

Crone, E.A., Wendelken, C., Donohue, S.E., Bunge, S.A., 2006. Neural evidence for dissociable components of task-switching. Cereb. Cortex 16, 475–486. http://dx.doi.org/10.1093/cercor/bhi127.

Crowe, D.A., Goodwin, S.J., Blackman, R.K., Sakellaridi, S., Sponheim, S.R., MacDonald Iii, A.W., Chafee, M.V., 2013. Prefrontal neurons transmit signals to parietal neurons that reflect executive control of cognition. Nat. Neurosci. 16, 1484–1491. http://dx.doi.org/10.1038/nn.3509.

De Baene, W., Brass, M., 2014. Dissociating strategy-dependent and independent components in task preparation. Neuropsychologia 62, 331–340. http://dx.doi.org/10.1016/j.neuropsychologia.2014.04.015.

Demanet, J., Liefooghe, B., Hartstra, E., Wenke, D., De Houwer, J., Brass, M., 2016. There is more into 'doing'than 'knowing': the function of the right inferior frontal sulcus is specific for implementing versus memorising verbal instructions. NeuroImage 141, 350–356. http://dx.doi.org/10.1016/j.neuroimage.2016.07.059.

Duncan, J., 2010. The multiple-demand (MD) system of the primate brain: mental programs for intelligent behaviour. Trends Cogn. Sci. 14, 172–179. http://dx.doi.org/10.1016/j.tics.2010.01.004.

Eger, E., Ashburner, J., Haynes, J.-D., Dolan, R.J., Rees, G., 2008a. fMRI activity patterns in human LOC carry information about object exemplars within category. J. Cogn. Neurosci. 20, 356–370. http://dx.doi.org/10.1162/jocn.2008.20019.

Eger, E., Kell, C.A., Kleinschmidt, A., 2008b. Graded size sensitivity of object-exemplar-evoked activity patterns within human LOC subregions. J. Neurophysiol. 100, 2038–2047. http://dx.doi.org/10.1152/jn.90305.2008.

Eklund, A., Nichols, T.E., Knutsson, H., 2016. Cluster failure: why fMRI inferences for spatial extent have inflated false-positive rates. Proc. Natl. Acad. Sci. 113, 7900–7905. http://dx.doi.org/10.1073/pnas.1602413113.

Etzel, J.A., Cole, M.W., Zacks, J.M., Kay, K.N., Braver, T.S., 2015. Reward motivation enhances task coding in frontoparietal cortex. Cereb. Cortex 26, 1647–1659. http://dx.doi.org/10.1093/cercor/bhu327.

Gilbert, S.J., 2011. Decoding the content of delayed intentions. J. Neurosci. 31, 2888–2894. http://dx.doi.org/10.1523/JNEUROSCI.5336-10.2011.

González-García, C., Mas-Herrero, E., de Diego-Balaguer, R., Ruz, M., 2016. Task-specific preparatory neural activations in low-interference contexts. Brain Struct. Funct. 221, 3997–4006. http://dx.doi.org/10.1007/s00429-015-1141-5.

Grill-Spector, K., Kourtzi, Z., Kanwisher, N., 2001. The lateral occipital complex and its role in object recognition. Vision. Res. 41, 1409–1422. http://dx.doi.org/10.1016/S0042-6989(01)00073-6.

Groppe, D.M., Urbach, T.P., Kutas, M., 2011. Mass univariate analysis of event-related brain potentials/fields I: a critical tutorial review: mass univariate analysis of ERPs/ERFs I: review. Psychophysiology 48, 1711–1725. http://dx.doi.org/10.1111/j.1469-8986.2011.01273.x.

Harris, R.J., Rice, G.E., Young, A.W., Andrews, T.J., 2016. Distinct but overlapping patterns of response to words and faces in the fusiform gyrus. Cereb. Cortex 26, 3161–3168. http://dx.doi.org/10.1093/cercor/bhv147.

Hartstra, E., Kühn, S., Verguts, T., Brass, M., 2011. The implementation of verbal instructions: an fMRI study. Hum. Brain Mapp. 32, 1811–1824. http://dx.doi.org/10.1002/hbm.21152.

Hartstra, E., Waszak, F., Brass, M., 2012. The implementation of verbal instructions: dissociating motor preparation from the formation of stimulus–response associations. Neuroimage 63, 1143–1153. http://dx.doi.org/10.1016/j.neuroimage.2012.08.003.

Haynes, J.D., Sakai, K., Rees, G., Gilbert, S., Frith, C., Passingham, R.E., 2007. Reading hidden intentions in the human brain. Curr. Biol. 17 (4), 323–328. http://dx.doi.org/10.1016/j.cub.2006.11.072.

Hebart, M.N., Görgen, K., Haynes, J.-D., 2015. The Decoding Toolbox (TDT): a versatile software package for multivariate analyses of functional imaging data. Front. Neuroinform. 8. http://dx.doi.org/10.3389/fninf.2014.00088.

Hikosaka, O., Isoda, M., 2010. Switching from automatic to controlled behavior: cortico-basal ganglia mechanisms. Trends Cogn. Sci. 14, 154–161. http://dx.doi.org/10.1016/j.tics.2010.01.006.

Huth, A.G., de Heer, W.A., Griffiths, T.L., Theunissen, F.E., Gallant, J.L., 2016. Natural speech reveals the semantic maps that tile human cerebral cortex. Nature 532, 453–458. http://dx.doi.org/10.1038/nature17637.

Kriegeskorte, N., Goebel, R., Bandettini, P., 2006. Information-based functional brain mapping. Proc. Natl. Acad. Sci. U. S. A. 103, 3863–3868. http://dx.doi.org/10.1073/pnas.0600244103.

Kriegeskorte, N., Simmons, W.K., Bellgowan, P.S.F., Baker, C.I., 2009. Circular analysis in systems neuroscience: the dangers of double dipping. Nat. Neurosci. 12, 535–540. http://dx.doi.org/10.1038/nn.2303.

Landsiedel, J., Gilbert, S.J., 2015. Creating external reminders for delayed intentions: dissociable influence on "task-positive" and "task-negative" brain networks. Neuroimage 104, 231–240. http://dx.doi.org/10.1016/j.neuroimage.2014.10.021.

Lau, H.C., Rogers, R.D., Haggard, P., Passingham, R.E., 2004. Attention to intention. Science 303, 1208–1210. http://dx.doi.org/10.1126/science.1090973.

Leshikar, E.D., Duarte, A., Hertzog, C., 2012. Task-selective memory effects for successfully implemented encoding strategies. PLoS One 7, e38160. http://dx.doi.org/10.1371/journal.pone.0038160.

Liefooghe, B., Houwer, J. De, Wenke, D., 2013. Instruction-based response activation depends on task preparation. Psychon. Bull. Rev. 20, 481–487. http://dx.doi.org/10.3758/s13423-013-0374-7.

Liefooghe, B., Wenke, D., De Houwer, J., 2012. Instruction-based task-rule congruency effects. J. Exp. Psychol. Learn. Mem. Cogn. 38, 1325–1335. http://dx.doi.org/10.1037/a0028148.

Locke, E.A., Latham, G.P., 2002. Building a practically useful theory of goal setting and task motivation: a 35-year odyssey. Am. Psychol. 57, 705–717. http://dx.doi.org/10.1037/0003-066X.57.9.705.

Lundstrom, B.N., Ingvar, M., Petersson, K.M., 2005. The role of precuneus and left inferior frontal cortex during source memory episodic retrieval. Neuroimage 27, 824–834. http://dx.doi.org/10.1016/j.neuroimage.2005.05.008.

Madore, K.P., Gaesser, B., Schacter, D.L., 2014. Constructive episodic simulation: dissociable effects of a specificity induction on remembering, imagining, and describing in young and older adults. J. Exp. Psychol. Learn. Mem. Cogn. 40, 609–622. http://dx.doi.org/10.1037/a0034885.

McCandliss, B.D., Cohen, L., Dehaene, S., 2003. The visual word form area: expertise for reading in the fusiform gyrus. Trends Cogn. Sci. 7, 293–299.

Meiran, N., 1996. Reconfiguration of processing mode prior to task performance. J. Exp. Psychol. Learn. Mem. Cogn. 22, 1423–1442. http://dx.doi.org/10.1037/0278-7393.22.6.1423.

Meiran, N., Cole, M.W., Braver, T.S., 2012. When planning results in loss of control: intention-based reflexivity and working-memory. Front. Hum. Neurosci. 6, 104. http://dx.doi.org/10.3389/fnhum.2012.00104.

Momennejad, I., Haynes, J.-D., 2013. Encoding of prospective tasks in the human prefrontal cortex under varying task loads. J. Neurosci. 33, 17342–17349. http://dx.doi.org/10.1523/JNEUROSCI.0492-13.2013.

Momennejad, I., Haynes, J.-D., 2012. Human anterior prefrontal cortex encodes the "what" and "when" of future intentions. Neuroimage 61, 139–148. http://dx.doi.org/10.1016/j.neuroimage.2012.02.079.

Muhle-Karbe, P.S., Andres, M., Brass, M., 2014. Transcranial magnetic stimulation dissociates prefrontal and parietal contributions to task preparation. J. Neurosci. 34, 12481–12489. http://dx.doi.org/10.1523/JNEUROSCI.4931-13.2014.

Muhle-Karbe, P.S., Duncan, J., Baene, W. De, Mitchell, D.J., Brass, M., 2016. Neural Coding for Instruction-Based Task Sets in Human Frontoparietal and Visual Cortex. Cereb. Cortex bhw032. doi:http://dx.doi.org/10.1093/cercor/bhw032

Mumford, J.A., Turner, B.O., Ashby, F.G., Poldrack, R.A., 2012. Deconvolving BOLD activation in event-related designs for multivoxel pattern classification analyses. Neuroimage 59, 2636–2643. http://dx.doi.org/10.1016/j.neuroimage.2011.08.076.

Nichols, T., Brett, M., Andersson, J., Wager, T., Poline, J.-B., 2005. Valid conjunction inference with the minimum statistic. Neuroimage 25, 653–660. http://dx.doi.org/10.1016/j.neuroimage.2004.12.005.

Norman, D.A., Shallice, T., 1986. Attention to action: willed and automatic control of behaviour. In: Davidson, R.J., Schwartz, G.E., Shapiro, D. (Eds.), Consciousness and Self-Regulation: Advances in Research and Theory.. Plenum Press, New York (NY), 1–18.

Pereira, F., Mitchell, T., Botvinick, M., 2009. Machine learning classifiers and fMRI: a tutorial overview. Neuroimage 45, S199–S209. http://dx.doi.org/10.1016/j.neuroimage.2008.11.007.

Roberts, D.J., Woollams, A.M., Kim, E., Beeson, P.M., Rapcsak, S.Z., Lambon Ralph, M.A., 2013. Efficient visual object and word recognition relies on high spatial frequency coding in the left posterior fusiform gyrus: evidence from a case-series of patients with ventral occipito-temporal cortex damage. Cereb. Cortex 23, 2568–2580 . http://dx.doi.org/10.1093/cercor/bhs224.

Rogers, R.D., Monsell, S., 1995. Costs of a predictable switch between simple cognitive tasks. J. Exp. Psychol. Gen. 124, 207.

Rubinstein, J.S., Meyer, D.E., Evans, J.E., 2001. Executive control of cognitive processes in task switching. J. Exp. Psychol. Hum. Percept. Perform. 27, 763–797. http://dx.doi.org/10.1037/0096-1523.27.4.763.

Ruge, H., Wolfensteller, U., 2010. Rapid formation of pragmatic rule representations in the human brain during instruction-based learning. Cereb. Cortex 20, 1656–1667. http://dx.doi.org/10.1093/cercor/bhp228.

Sakai, K., 2008. Task set and prefrontal cortex. Annu. Rev. Neurosci. 31, 219–245. http://dx.doi.org/10.1146/annurev.neuro.31.060407.125642.

Sakai, K., Passingham, R.E., 2006. Prefrontal set activity predicts rule-specific neural processing during subsequent cognitive performance. J. Neurosci. 26, 1211–1218. http://dx.doi.org/10.1523/JNEUROSCI.3887-05.2006.

Sakai, K., Passingham, R.E., 2003. Prefrontal interactions reflect future task operations. Nat. Neurosci. 6, 75–81. http://dx.doi.org/10.1038/nn987.

Schacter, D.L., Addis, D.R., 2009. On the nature of medial temporal lobe contributions to the constructive simulation of future events. Philos. Trans. R. Soc. Lond. B Biol. Sci. 364, 1245–1253. http://dx.doi.org/10.1098/rstb.2008.0308.

Schlichting, M.L., Preston, A.R., 2015. Memory integration: neural mechanisms and implications for behavior. Curr. Opin. Behav. Sci. 1, 1–8. http://dx.doi.org/10.1016/j.cobeha.2014.07.005.

Stocco, A., Lebiere, C., O'Reilly, R.C., Anderson, J.R., 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. http://dx.doi.org/10.3758/s13415-012-0117-7.

Szpunar, K.K., Spreng, R.N., Schacter, D.L., 2014. A taxonomy of prospection: introducing an organizational framework for future-oriented cognition. Proc. Natl. Acad. Sci. 111, 18414–18421. http://dx.doi.org/10.1073/pnas.1417144111.

Turner, B.O., Mumford, J.A., Poldrack, R.A., Ashby, F.G., 2012. Spatiotemporal activity estimation for multivoxel pattern analysis with rapid event-related designs. Neuroimage 62, 1429–1438. http://dx.doi.org/10.1016/j.neuroimage.2012.05.057.

Vinckier, F., Dehaene, S., Jobert, A., Dubus, J.P., Sigman, M., Cohen, L., 2007. Hierarchical coding of letter strings in the ventral stream: dissecting the inner organization of the visual word-form system. Neuron 55, 143–156. http://dx.doi.org/10.1016/j.neuron.2007.05.031.

Wagner, A.D., Shannon, B.J., Kahn, I., Buckner, R.L., 2005. Parietal lobe contributions to episodic memory retrieval. Trends Cogn. Sci.. http://dx.doi.org/10.1016/j.tics.2005.07.001.

Wenke, D., Gaschler, R., Nattkemper, D., Frensch, P.A., 2009. Strategic influences on implementing instructions for future actions. Psychol. Res. 73, 587–601. http://dx.doi.org/10.1007/s00426-009-0239-x.

Ye, Z., Mestres-Missé, A., Rodriguez-Fornells, A., Münte, T.F., 2011. Two distinct neural networks support the mapping of meaning to a novel word. Hum. Brain Mapp. 32, 1081–1090. http://dx.doi.org/10.1002/hbm.21093.