Frontoparietal action-oriented codes support novel task set implementation

Carlos González-García*, Silvia Formica, David Wisniewski, and Marcel Brass

Department of Experimental Psychology, Ghent University, Belgium

*Corresponding author: Carlos González-García (carlos.gonzalezgarcia@ugent.be)

Abstract

A key aspect of human cognitive flexibility concerns the ability to rapidly convert complex symbolic instructions into novel behaviors. Previous research proposes that this fast configuration is supported by two differentiated neurocognitive states, namely, an initial declarative maintenance of task knowledge, and a progressive transformation into a pragmatic, action-oriented state necessary for optimal task execution. Furthermore, current models predict a crucial role of frontal and parietal brain regions in this transformation. However, direct evidence for such frontoparietal formatting of novel task representations is still lacking. Here, we report the results of an fMRI experiment in which participants had to execute novel instructed stimulus-response associations. We then used a multivariate pattern-tracking procedure to quantify the degree of neural activation of instructions in declarative and procedural representational formats. This analysis revealed, for the first time, format-unique representations of relevant task sets in frontoparietal areas, prior to execution. Critically, the degree of procedural (but not declarative) activation predicted subsequent behavioral performance. Our results shed light on current debates on the architecture of cognitive control and working memory systems, suggesting a contribution of frontoparietal regions to output gating mechanisms that drive behavior.

INTRODUCTION

Some of the most advanced collaborative human achievements rely on our ability to rapidly learn novel tasks. Instruction following constitutes a powerful instance of this ability as it combines the flexibility to specify complex abstract relationships with an efficiency far superior to other forms of task learning such as trial and error, or reinforcement learning. These unique characteristics make it a distinctive skill that separates humans from other species[1]. While recent years have witnessed substantial progress in our understanding of instruction following, the neural and cognitive mechanisms underlying this rapid transformation of complex symbolic information into effective behavior are still poorly understood. Specifically, a critical question that remains unresolved is whether a declarative representation of task information is sufficient or whether an additional representational state, closely linked to action, precedes optimal performance.

Previous behavioral studies have consistently reported an intriguing signature of instruction processing, namely, a reflexive activation of responses on the basis of merely instructed stimulus-response (S-R) associations (defined as “intention-based reflexivity”, or IBR). IBR occurs even when instructions are task-irrelevant and have not been overtly executed before[2–7], which suggests a rapid configuration of instructed content predominantly towards action. Instruction implementation also has a profound impact on brain activity, as shown by electroencephalography and fMRI studies. In particular, the intention to execute an instruction induces automatic motor activation[8,9], engages different brain regions to coordinate novel stimuli and responses[10–14], and alters the neural code of the encoded instruction[15,16].

These and other findings propose a crucial role of a frontoparietal network (FPN) in the instantiation of a highly efficient task readiness state[11–17]. Accordingly, evidence coming from frontal patients[18] and healthy participants[10,15,19], as well as prominent theoretical models[20] support a *serial coding hypothesis*, a two-step process in which the FPN first encodes instructed information into a primarily *declarative* representation, that is, a persistent representation of the memoranda conveyed by the instruction. Crucially, when this information becomes behaviorally relevant, FPN declarative representations are transformed into an independent state that is optimized for specific task demands[20]. This *procedural* state would entail a proactive binding of relevant perceptual and motor information into a compound representation that leads to the boost of relevant action codes related to behavioral routines[16].

However, evidence for such serial coding in control regions is lacking, primarily due to the fact that previous analytical approaches were unable to track representational formats of specific nature. Previous work thus identified some properties of the FPN during the implementation of novel instructions, such as enhanced decoding of stimulus category[11,16], or altered similarity within to-be-implemented S-R associations[13,15], but failed to determine the functional state underlying such representational effects. Therefore, currently, it cannot be discerned whether novel task setting is achieved through the proposed frontoparietal formatting. In fact, at least two alternatives to the serial coding hypothesis could explain previous results. First, an *amplification hypothesis* disputes the notion of two independent representational states and proposes that the intention to implement rather induces deeper declarative processing of the initial semantic information conveyed by the instruction[2]. Under this proposal, the FPN would support instruction implementation through the preservation of relevant declarative signals rather than through a transformation of these signals into an action-oriented code. Last, an intermediate alternative concerns the possibility that implementation involves both the boost of an independent action-oriented signal and, additionally, the preservation of declarative representations. This *dual-coding hypothesis* thus predicts that novel task implementation is supported by non-overlapping declarative and procedural task representations in the FPN.

Here, we aimed at adjudicating between these three options. In the current study, participants performed a task in which 4 novel S-R associations were presented at the beginning of each trial (each S-R consisted of an image and a response finger; for instance, the picture of a cat and the word "index"). After the encoding screen, a retro-cue would select a subset of two S-Rs, prior to the onset of a target screen. Target screens displayed the image belonging to one of the selected mappings (for example, a picture of a cat), prompting participants to execute the associated response (Fig. 1). Based on recent experimental results[7,21,22] and theoretical models of working memory (WM)[23], we assumed that retro-cues (i.e. cues that signal the relevance of one of the already encoded representations in WM) would prioritize relevant S-R associations into a behavior-optimized state, akin to implementation. As such, retro-cues served as a tool to locate in time the moment after initial encoding in which implementation-specific signals should be magnified. Our primary goal was to capture which signals governed FPN activity during such implementation stage, prior to execution[20]. To discern the hypothesized procedural and declarative traces, we had participants perform two functional localizers that encouraged either a declarative or action-oriented maintenance of novel instructions. Using data from the localizers, we derived a canonical multivariate pattern of activity for each S-R in both declarative and procedural formats. We then assessed the extent to which these traces were independently activated in the main task, during the implementation stage.

We first predicted that the intention to implement would boost the representation of retro-cued S-R associations in the FPN, compared to encoded but not cued S-Rs. We then tested whether this representational boost reflected the activation of the relevant S-R in two unique formats, namely, declarative and procedural. If so, this would indicate the extent to which multiple, non-overlapping representations of the same instructed content underlie novel task setting.

RESULTS

Task set prioritization enhances instruction execution

Twenty-nine healthy human participants (mean age = 23.28, 17 females; 3 more participants were excluded after data acquisition, see Methods) were shown 4 novel S-R associations at the beginning of each trial. Importantly, even though specific S-R associations were presented only once throughout the experiment, they could be grouped in categories depending on the specific combination of stimulus and response dimensions (for instance, “animate item and index finger response”; see Methods for a full description of S-R categories). Immediately after the encoding screen, a retro-cue signaled the relevance of two specific mappings (informative retro-cues in 75% of trials; in the remaining trials a neutral retro-cue did not select any mapping). The two selected mappings always belonged to the same S-R category, although the specific associations remained unique. Such grouping was crucial for analysis purposes since it allowed us to identify the *selected*, *unselected*, and *not presented* S-R categories on each trial. After the retro-cue, a target image prompted participants to provide the corresponding response (Fig. 1). To ensure that participants encoded all 4 S-R associations, ~6% of trials (regardless of the retro-cue validity) displayed a new, catch image, prompting participants to press all four available buttons simultaneously.

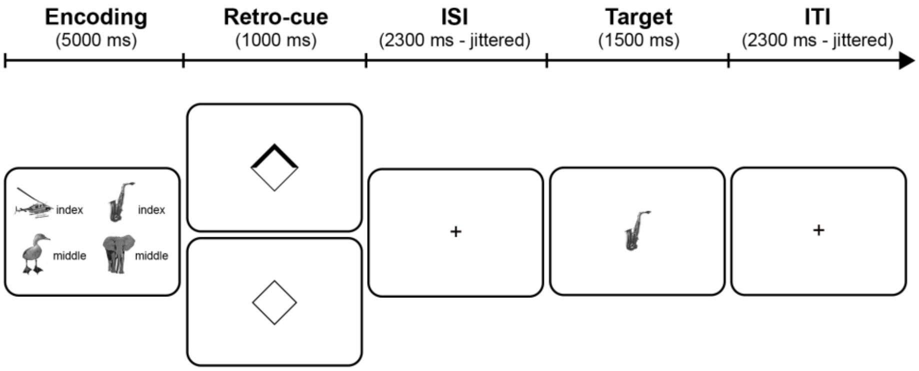

Figure 1. Behavioral paradigm. On each trial, participants first encoded four novel S-R mappings consisting in the association between an (animate or inanimate)

item and a response (index or middle fingers; response hand defined by the position of the mapping on the screen; e.g. “helicopter-index” on the left-hand side of the screen requested participants to press the *left* index if the target screen displayed a helicopter). After the encoding screen, an informative retro-cue (75% of the trials) signaled the relevance of two of the mappings. In the remaining 25% of trials, a neutral retro-cue appeared, and none of the mappings were cued. Last, after a jittered retro-cue-target interval, a target stimulus prompted participants to provide the associated response (in this example, “right index” finger press).

Analysis of participants’ behavioral performance revealed that retro-cues helped participants in prioritizing novel S-Rs. Specifically, participants were faster ($t_{28,1}$ = 13.51, p < 0.001, Cohen’s d = 2.51; Fig. 2a) and made less errors ($t_{28,1}$ = 7.96, p < 0.001, Cohen’s d = 1.47; Fig. 2b, left panel) in trials with informative retro-cues, compared to neutral.

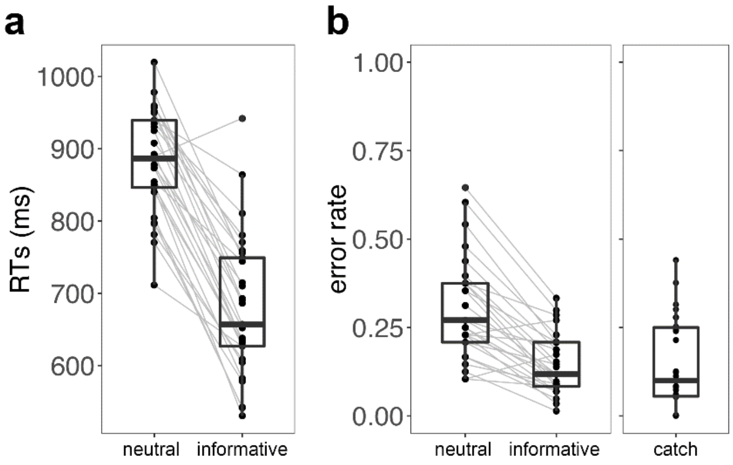

Figure 2. Behavioral results. (**a**) Reaction times in neutral and informative retro-cue trials. (**b**) Error rates in neutral, informative, and catch trials. The thick line inside box plots depicts the second quartile (median) of the distribution (n = 29). The bounds of the boxes depict the first and third quartiles of the distribution. Whiskers denote the 1.5 interquartile range of the lower and upper quartile. Dots represent individual subjects' scores. Grey lines connect dots corresponding to the same participant in two different experimental conditions.

Identifying task set prioritization activity

As a first step, we investigated which brain regions were predominantly involved in instruction prioritization. Our intuition was that prioritization would boost implementation signals and, as such, we expected a frontoparietal network to be particularly crucial, as it is usually involved in the implementation of novel task sets[11,14–17,24]. We thus established a set of a priori candidate regions that encompassed frontal (inferior and middle frontal gyri) and (inferior and superior) parietal cortices (see Fig. 3c, and the Region-of-interest definition section in the Methods). We then performed two whole-brain analyses to find regions sensitive to task set prioritization (defined as informative vs. neutral retro-cues) in their overall activation magnitude or voxel-wise activity patterns, using a general linear model (GLM) and multivariate pattern analysis (MVPA), respectively. First, we found that informative retro-cues elicited significantly higher activity in regions of the FPN, including the inferior and middle frontal gyri, inferior and superior parietal cortices, as well as regions outside the FPN, such as the lateral occipital cortex (Fig. 3a, primary voxel threshold [$p < 0.001$ uncorrected] and cluster-defining threshold [FWE $p < .05$]). Furthermore, a searchlight decoding analysis[25] revealed that the FPN contained information in its patterns of activity about the prioritization status (Fig. 3b, primary voxel threshold [$p < 0.0001$ uncorrected] and cluster-defining threshold [FWE $p < .05$]; see also Methods for details on how this analysis controlled for univariate differences in activity magnitude). Overall, the resulting statistical maps of these two analyses roughly overlap with the set of a priori defined regions of interest (ROIs; Fig. 3C), confirming the involvement of the FPN in task set prioritization.

To test our hypothesis that implementation would boost the representation of retro-cued S-R categories, we performed two similar decoding analyses in the 4 FPN ROIs. First, we tested if in the moment of the retro-cue the patterns of activity in these four regions carried information about the category of the cued S-R. We found significant category decoding in the right PFC and bilateral parietal ROIs (one-sample t-tests against chance level, all *p*s < 0.013, FDR-corrected for multiple comparisons), and close to significance decoding in the left PFC ($t_{25,1} = 1.69$, $p = 0.052$). Next, we tested the extent to which the FPN also carried information about the encoded, but not cued category. In contrast with the previous results, decoding did not reach significance in any of the ROIs (all *p*s > 0.6). Finally, we directly compared the decoding accuracies for the cued and uncued categories. This analysis revealed significantly stronger decoding of the cued category compared to the uncued one in right PFC and bilateral parietal cortices (paired t-tests, all *p*s < 0.034, FDR-corrected; Fig. 3d).

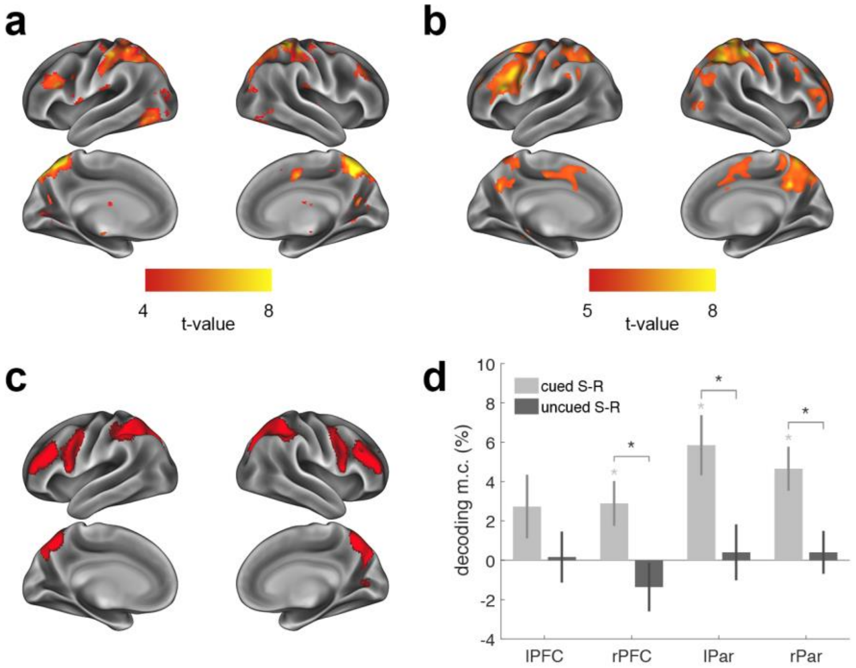

Figure 3. Task set prioritization induced changes in frontoparietal neural activity. (**a**) GLM contrast of informative > neutral retro-cue trials. Warm colors show regions with significantly higher activity magnitude during informative compared to neutral retro-cues (primary voxel threshold [*p* < 0.001 uncorrected] and cluster-defining threshold [FWE *p* < .05]). (**b**) Searchlight decoding of prioritization (informative vs. neutral retro-cue). Warm colors show regions with significant decoding (primary voxel threshold [*p* < 0.0001 uncorrected] and cluster-defining threshold [FWE *p* < .05]). (**c**) Set of regions-of-interest defined prior to analyses, encompassing frontal (inferior and middle frontal gyri) and (inferior and superior) parietal cortices. (**d**) Mean S-R category decoding (minus chance) within each region of interest. Error bars denote between-participants s.e.m. Grey asterisks denote significant decoding (chance level = 25%, one-sample t-test, FDR-

corrected). Black asterisks denote significantly higher decoding of cued compared to uncued S-R categories (paired t-test, FDR-corrected).

Tracking format-unique task set patterns

Altogether, these results show that instruction implementation has a profound impact on FPN activity, boosting the representation of prioritized task sets over encoded, but irrelevant ones. However, similarly to previous studies, they are agnostic regarding the nature of the signals underlying such effect. The main goal of our study was to test the extent to which, during this implementation stage, relevant task information was represented in a declarative and/or procedural format. In a first scenario (amplification hypothesis), implementation would merely preserve relevant declarative information. Alternatively, it could transform the initial representation of task information into a primarily action-oriented format (serial coding hypothesis). Last, action-oriented representations could coexist with preserved declarative representations (dual coding hypothesis). To adjudicate between these options, we implemented a canonical template tracking procedure that allowed us to estimate the degree of neural activation of specific S-R categories under the two functional formats of interest (see Figure 4, for a visual representation of the procedure). To do so, for each subject, we first obtained whole-brain templates of each S-R category in procedural and declarative formats, using data from two functional localizers. Subsequently, we estimated the extent to which these two traces governed the data of the main task, specifically during the presentation of informative retro-cues. We performed this step in an ROI-based fashion. For each ROI and trial type, we extracted the pattern of activity during the retro-cue, keeping track of which S-R categories were either cued, uncued, or not presented in that trial. Then, we computed the semi-partial correlation between this pattern of activity and the declarative and procedural templates of each S-R category. Importantly, we used semi-partial correlations as they allowed us to estimate the amount of shared variance between task data and a given template (e.g. S-R category 1 in procedural state) that is not explained by the same template in the alternative state (e.g. S-R category 1 in declarative state). Therefore, processes common to both localizers (e.g. arousal, domain-general attention and/or task preparation) cannot inflate correlations, and any significant result rather reflects the activation of S-R information in a specific format during the main task.

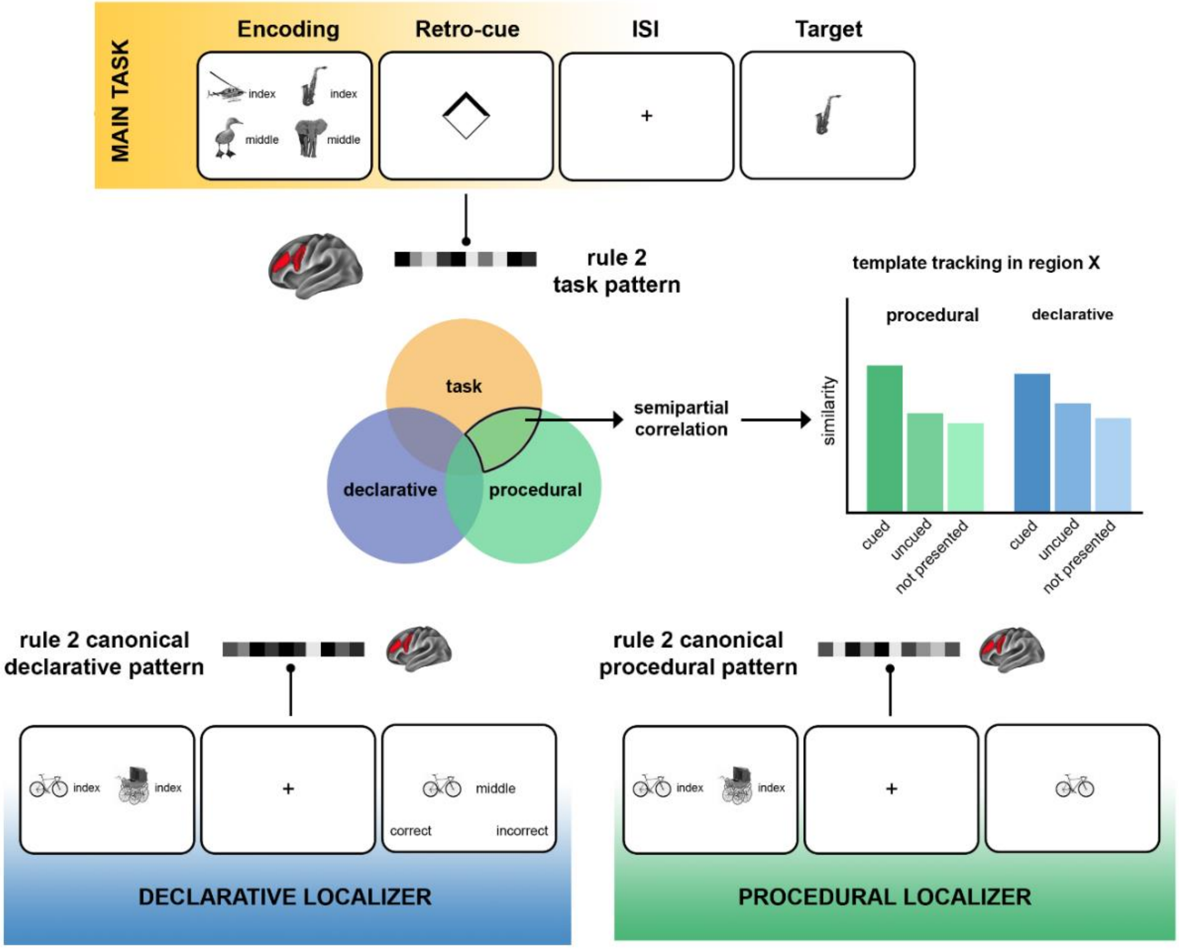

Figure 4. Schematic of the canonical template tracking procedure. For each region of interest, we extracted the pattern of activity of specific S-R categories during informative retro-cues (upper panel, in yellow) and computed similarity with canonical templates of such categories in declarative (bottom left, in blue) and procedural (bottom right, in green) formats, obtained in two separate localizers. Importantly, similarity was assessed via semi-partial correlations, obtaining the proportion of uniquely shared variance between task and template data (middle, Venn diagram) of the cued, uncued and not-presented S-R categories. Graphs represent a hypothetical set of results, in which implementation recruits non-overlapping procedural and declarative representations of cued S-R category. This

informational boost, relative to baseline (not-presented S-R categories), is superior to that of the uncued category.

To validate this procedure outside the FPN, we created an ROI comprising the primary motor cortex, since predictions for this regions were straightforward: (1) boost of action-oriented information of the cued S-R category, compared to the uncued and not-presented ones; and (2) no boost of declarative information. The results obtained (Fig. 5) matched the predictions, revealing a specific enhancement of procedural information of the cued category compared to the uncued ($t_{25,1}$ = 4.08, p < 0.001, Cohen's d = 0.80), and critically, to the empirical baseline defined by the not-presented categories ($t_{25,1}$ = 5.45, p < 0.001, Cohen's d = 1.07). No reactivation of the uncued S-R category was found ($t_{25,1}$ = 1.32, p = 0.2, Cohen's d = 0.26). As predicted, no differences between cued, uncued and baseline categories were found in declarative signals (all *t*s < 1.53, all *p*s > 0.14).

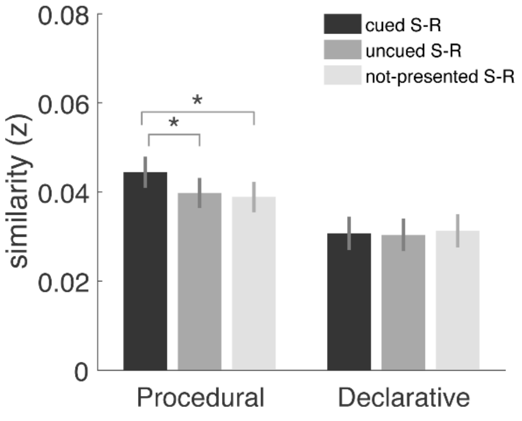

Figure 5. Template tracking procedure results in the primary motor cortex. Bars represent the normalized semi-partial correlation between task data and the

procedural and declarative templates of cued, uncued and not presented S-R categories. Error bars denote within-participants s.e.m[26]. Asterisks denote significant differences ($p < 0.05$, paired t-test).

Declarative and procedural representations in frontoparietal cortices (and beyond)

To elucidate which signals govern implementation in control-related regions, we carried out the template tracking procedure on each FPN region separately. Furthermore, we decided to include the ventral visual cortex (VVC) in this analysis to explore the effect of implementation in higher-order visual regions, since these have been consistently shown to be involved in instruction processing[11,13,14,16].

This analysis (Fig. 6a) revealed that all FPN regions contain unique action-oriented information of relevant S-R categories during the presentation of the retro-cue (two-tail paired t-test against empirical baseline [not-presented rules], all *t*s > 2.16, all *p*s < 0.04, all Cohen's *d* > 0.42). Critically, procedural information of cued categories was significantly more activated than uncued categories (all *t*s > 2.26, all *p*s < 0.04, all Cohen's *d* > 0.44). Regarding declarative information (Fig. 6b), parietal nodes of the FPN showed a specific enhancement of declarative information of the cued S-R category, compared to the uncued one (*t*s > 2.16, all *p*s < 0.02, all Cohen's *d* > 0.49), whereas no significant differences were found in the right ($t = 1.24$, $p = 0.28$) and left ($t = 2.05$, $p = 0.051$) frontal nodes. To assess the reliability of these not significant findings, we performed Bayesian paired t-tests with the same factors as before. The BF_{10} (evidence in favor of H_1 against evidence for H_0) for the Cued – Not presented comparison was 0.27 and 0.24 for the left and right frontal nodes, respectively. Similarly, the comparison Cued – Uncued yielded a $BF_{10} = 1.25$ in the left frontal node, and a $BF_{10} = 0.41$ in the right frontal node. Overall, this constitutes moderate evidence[27] for the null hypothesis that declarative information of the cued category was not specifically enhanced in frontal regions.

Last, higher-order visual regions showed a similar pattern to parietal nodes of the FPN, with significant enhancement of both procedural ($t = 6.19$, $p < 0.001$, Cohen's $d = 1.21$) and declarative ($t = 5.84$, $p < 0.001$, Cohen's $d = 1.15$) information of the cued S-R category, compared to the uncued one.

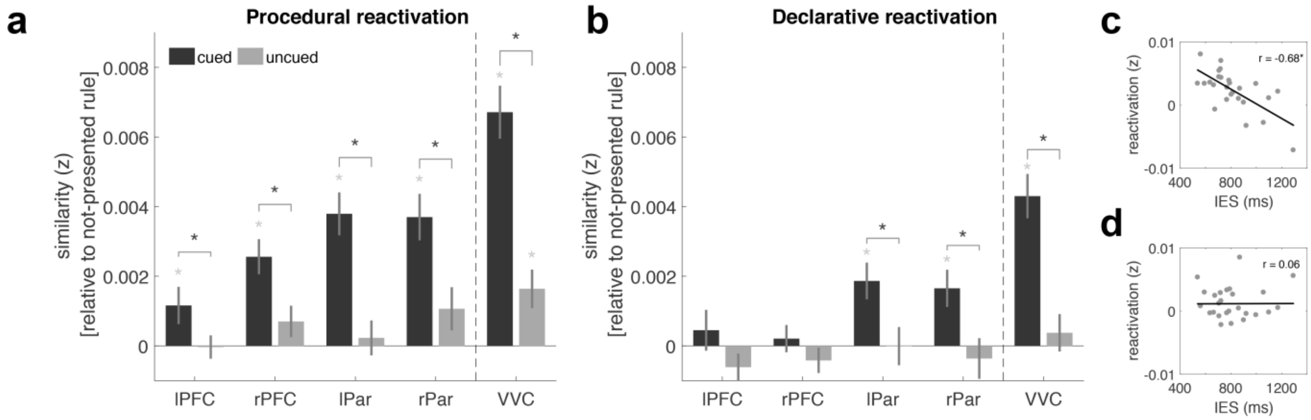

Figure 6. Canonical template tracking procedure results in frontoparietal cortices and ventral visual cortex. Bars represent the normalized semi-partial correlation between task data and (**a**) the procedural and (**b**) declarative templates of cued and uncued S-R categories, relative to empirical baseline (not-presented S-Rs). Error bars denote within-participants s.e.m. Gray asterisks denote a significant

increase from baseline ($p < 0.05$, paired t-test, FDR-corrected). Black asterisks denote significant differences between cued and uncued categories ($p < 0.05$, paired t-test, FDR-corrected). (**c**) Across-participant correlation of Inverse Efficiency Scores and procedural activation index in frontoparietal cortices. (**d**) Correlation of Inverse Efficiency Scores with declarative activation index in frontoparietal cortices. In **c** and **d**, dots represent individual participants, thick lines depict the linear regression fit, and asterisks denote significant Pearson's correlation ($p < 0.05$).

Action-oriented codes support novel task setting

What might be the behavioral relevance of declarative and procedural signals? We reasoned that if action-oriented representations are boosted during implementation in control-related regions, and implementation can be conceived as a behavior-optimized state, then the degree of action-oriented activation should predict the efficiency of instruction execution. To test this hypothesis, we first converted RTs and error rates of informative retro-cue trials into a single compound measure (Inverse Efficiency Scores; IES. IES were obtained by dividing each participant's mean RT by the percentage of accurate responses[28]). Then, we derived a template activation index by subtracting the degree of activation of cued categories to that of uncued categories for each region and format (procedural and declarative). Finally, we correlated individual IES with the activation indices on each region of the FPN. This analysis revealed significant negative correlations in all FPN regions between IES and procedural activation (all Pearson's *r*s > -0.475, all *p*s < 0.02). In contrast, IES did not correlate with declarative activation in any region (all *r*s < -0.34, all *p*s > 0.09). When averaging activation indices across FPN regions, an identical pattern was found, namely, a significant correlation of IES with procedural (*r* = -0.679, p < 0.001) but not declarative (*r* = 0.06, *p* = 0.77) activation (Fig. 6c-d). Similar results were obtained when using RTs (procedural: r = -0.67, p < 0.001; declarative: r = 0.076, p = .71) and error rates (procedural: r = -0.54, p = 0.004; declarative: r = -0.019, p = 0.93) as behavioral measures. Altogether, these results show that the more the FPN represented procedural information of relevant S-Rs, the faster and more accurate participants executed the instruction. In contrast, the strength of declarative signals of the same S-R association did not predict behavioral performance.

DISCUSSION

In the current study, we report a pervasive effect of novel task sets implementation across behavioral and neural data. Our results provide support for a frontoparietal dual coding of instructed task information. A canonical template tracking procedure revealed the boost of unique declarative and procedural representations in the FPN, prior to execution. This boost was specific to prioritized S-Rs and did not happen for irrelevant mappings. Critically, our results show that procedural (but not declarative) activation in the FPN predicted efficient execution of novel instructions.

Frontoparietal flexible coding of relevant task sets

Previous research has highlighted the important role of the FPN in the implementation of novel instructions[10–16,29]. Accordingly, our results show that FPN involvement during implementation reflects the boost of relevant S-R categories. However, these results remain agnostic regarding the nature of the signals underlying this effect. In principle, as proposed by the serial-coding hypothesis, they could reflect the emergence of procedural representations, in detriment of merely declarative signals[16,20]. However, the same pattern of results could be explained by a mere amplification of preserved declarative representations[2]. Last, the results could reflect both declarative preservation and procedural activation, as predicted by a dual-coding hypothesis. Using a canonical template tracking analysis we were able to adjudicate between these options and, for the first time, obtain evidence in favor of the dual coding hypothesis. As such, our results show that implementation engages independent procedural and declarative representations of relevant task information in the FPN.

A first consideration concerns the exact nature of the reactivated signals. In the declarative localizer, participants had to remember specific S-R associations and match them to another S-R probe. In contrast, in the procedural localizer, participants' goal was to execute the correct response associated with a target stimulus. The different readout from WM thus encouraged different strategies, as suggested by previous studies[3,7,16]. Therefore, it is conceivable that templates will contain unique information: a persistent maintenance of the memoranda in the declarative localizer, and a proactive action-oriented representation, in the procedural localizer. However, templates likely share further information, for instance, related to specific perceptual stimulation and general-domain processes, such as arousal or attention. We took several measures to reduce the influence of information not specifically related to declarative or procedural components. First, template reactivation was derived from semi-partial correlations between data from the main task and the localizers. Thus, our measure reflects unique shared variance between the task and the representation of an S-R category in a given localizer, partialling out the variance explained by the representation of the same S-R in the remaining localizer. Shared variance between both localizers and the main task could induce spurious similarity increases. For instance, domain-general selective attention is likely engaged towards selected mappings in the main task, as well as during the preparation interval of the localizers. Such a scenario would inflate the correlations between the templates of the cued S-R associations and the data from the main task, potentially leading to a significant difference from baseline. In contrast, semi-partial correlations ensured that procedural and declarative activation indices were derived from non-overlapping signals. Second, templates were built for S-R categories rather than unique mappings, and therefore a contribution of perceptual features to template reactivation seems unlikely. Moreover, semi-partial correlations were computed between data from the retro-cue screen (in the main task), and inter-stimulus interval (in the localizers), which reduces the likelihood of significant correlations due to perceptual similarity between templates and specific S-Rs. Therefore, we believe it is the most straightforward interpretation to consider that our procedure succeeded at tracking specific declarative and procedural signals, as also hinted by the validation results in the motor cortex. From this standpoint, our results suggest that during task set implementation, FPN regions can maintain the declarative memoranda conveyed by the instruction and, simultaneously, an independent action-oriented S-R code that primarily drives task execution.

Heterogeneous task set coding within the FPN

Although we did not have specific hypotheses for the role of individual FPN regions, a second important finding concerns the heterogeneity of results within this network. Whereas parietal nodes carried both procedural and declarative information in their patterns of activity, only action-oriented representations were found in frontal nodes. Given the overall low signal-to-noise ratio and pattern reliability in prefrontal cortices[30], one potential interpretation could be that slight differences inherent in the templates could affect the reactivation measures. For instance, it could be argued that signal quality of procedural templates in frontal nodes is intrinsically higher than that of declarative templates, which in turn might induce a lack of power to detect the reactivation of declarative templates in the same regions during the task. To rule out these concerns, and inspired by previous studies using similar canonical template tracking procedures[31], for each template and region of the FPN, we compared the signal-to-noise ratio (computed as mean t-value across voxels of the ROI divided by the standard deviation), informational content (computed as Shannon entropy) and correlationability of the templates (i.e. the degree to which individual templates correlated with other templates from the same localizer). This analysis revealed that procedural and declarative FPN templates did not differ in any of these measures (Supplementary Table 1).

Thus, our results suggest, first, that prefrontal representations carry action-oriented information during instruction following. This is line with previous studies that propose a crucial role of the frontolateral cortex in the integration of stimulus and response information into a task set based on verbal instructions[12,32,33], as well as in representing task rules[17,24] and goals[34]. In contrast, parietal cortices contained both declarative and procedural information of relevant S-Rs. Whereas the role of parietal regions in representing goals and task set information is widely acknowledged[11,13,16,17,24,34,35], it is unclear what drives such declarative activation. One possibility is that it reflects a category-specific top-down selection scheme, driven by increased attention towards the cued S-R[36,37]. The fact that a similar pattern was found in higher-order visual regions, which usually coordinate with parietal cortices to represent relevant task dimensions in anticipation of future demands[38–40], further supports this possibility. This tentative interpretation would be coherent with goal neglect effects reported in patients with frontal lobe damage[18]. These patients are capable of selecting, maintaining, and remembering task-relevant information, yet their ability to transform relevant information into goal-driven actions is impaired. Such dissociation goes at least partially in line with our results in that (1) prioritization of goal-oriented representations depends critically on prefrontal cortices (impaired in goal neglect patients), and (2) the involvement of other control-related regions, intact in these patients, boosts the declarative representation of specific task information, such as particular S-R categories, presumably in coordination with posterior category-selective regions.

Implementation as a selective output gating process

Remarkably, despite both signals coexisted in the FPN during implementation, only procedural representations predicted efficient behavior. The fact that implementation is signaled by retro-cues renders this effect relevant to current debates on information prioritization and WM architecture. In this regard, our results are consistent with the notion of an output gating mechanism. Similar to the idea of an input gate that limits what information enters WM, some computational models propose an additional gate that determines which pieces of this information will drive behavior[41]. Recent theoretical frameworks suggest a role of prioritization not only in selecting relevant content from WM but also in reformatting such content into a “behavior-guiding representational state”[23], analogous to an output gating mechanism. Interestingly, these models propose that whereas other control-related regions might be involved in attention-driven representations of relevant content, frontal regions are thought to be especially important in transferring this content into a state that is optimal for behavior. In line with these ideas, we show that an action-oriented representation of task sets dominates activity in frontal cortices and that this representational format, and not a declarative one, is tightly linked to behavioral efficiency. Importantly, our results reveal, first, that the neural substrate of task set prioritization involves further brain regions, such as category-selective and parietal cortices. Second, action-oriented representations might coexist with declarative-like information in some of these regions. It should be noted, however, that fMRI data lacks the temporal resolution to discern whether these two signals fully overlap in time or whether action-oriented, behavior-optimized representations emerge after declarative information of relevant task sets has been prioritized. Future studies should employ time-resolved techniques that can succeed at characterizing the dynamical contribution of different brain regions to separate control and WM processes[42].

In summary, the present study reveals the strong impact of novel task setting in frontoparietal regions. Following task prioritization, we observed a boost in information of the relevant S-R category in detriment of the irrelevant ones. This boost was accompanied by the activation of two non-overlapping neural codes in the FPN, one reflecting the declarative maintenance of task, and another, more pragmatic, action-oriented coding of the instruction. Importantly, only this procedural activation predicted behavioral performance. Altogether, our results support the idea that novel instructed content can be represented in multiple formats, and highlight the contribution of frontoparietal regions to output gating mechanisms that drive behavior.

METHODS

Methods are reported, when applicable, in accordance with the Committee on Best Practices in Data Analysis and Sharing (COBIDAS) report[43].

Participants

Thirty-two participants (mean age = 23.16, range = 19-33; 20 females) recruited from the participants' pool from Ghent University participated in exchange of 40 euros. They were all right-handed (confirmed by the Edinburgh handedness inventory), clinically healthy and MRI-safe. The study was approved by the UZ

Gent Ethics Committee and all participants provided informed consent before starting the experiment. Of the initial 32 participants, 3 were excluded after acquisition (1 participant performed at chance during the task; 1 participant had an error rate of 1 in catch trials (see below); 1 participant's within-run head movement exceeded voxel size), resulting in a final sample of 29 participants. Due to an incomplete orthogonalization of the cued and uncued S-R categories, the first three participants were excluded from multivariate analyses (n = 26).

Materials

S-R associations were created by combining images with words that indicated the response finger. Each S-R association was presented just once during the entire experiment to prevent the formation of long-term memory traces[6]. Given this prerequisite, images of animate (non-human animals) and inanimate (vehicles and instruments) items were compiled from different available databases[44–48], creating a pool of 1550 unique pictures (770 animate items, 780 inanimate). To increase perceptual similarity and facilitate recognition, the background was removed from all images, items were centered in the canvas, and images were converted to black and white.

The response dimension was defined by the combination of a word ("index" or "middle") and the position of the mapping in the encoding screen. For instance, if an S-R pair containing the word "index" was displayed on the left-hand side of the screen, this informed participants that the correct response associated with that particular stimulus would be "*left* index". This allowed us to have 2 mappings on screen that involved the same *response category* (e.g. index finger) but different effectors (e.g. *left* index finger vs *right* index finger).

The combination of the 2 stimulus dimensions (animate/inanimate items) and the 2 response dimensions (index/middle finger) lead to 4 *S-R categories*: Category 1 (animate-index), Category 2 (inanimate-index), Category 3 (animate-middle), and Category 4 (inanimate-middle). Although images were always unique and therefore the specific image-finger mapping changed on every trial, S-R associations were grouped into these 4 categories for analysis purposes.

Task and design specifications

Each trial started with an encoding screen (5000 ms) that displayed 4 S-R associations. The two mappings on the upper half of the encoding screen belonged to one S-R category, and the other two belonged to another S-R category. Immediately after the encoding screen, a retro-cue appeared. Informative retro-cues (75% of trials) consisted of an arrow centered in the middle of the screen pointing either upwards or downwards. Therefore, informative retro-cues did not select a specific S-R mapping but rather two mappings belonging to the same S-R category (e.g. “animate - index finger”). Neutral retro-cues did not select any mapping. The retro-cue was displayed for 1000 ms and was followed by a fixation point (cue-target interval; CTI), which duration was jittered following a pseudo-logarithmic distribution (mean duration = 2266 ms, SD = 1276 ms, range = [600-5000]). Directly after the CTI, a target was on screen for 1500 ms. Target screens displayed the image belonging to one of the selected mappings, prompting participants to execute the associated response by pressing the corresponding button in an MRI-compatible button box. In neutral trials, the target could be the stimulus of any of the 4 S-R encoded mappings. Additionally, in ~6% of trials, a catch target appeared. This consisted of a new image, different from any of the encoded stimuli, to which participants had to answer by pressing the 4 available buttons in the response box. Catch trials were included to ensure that participant encoded all four S-R associations. Last, after the target screen, a fixation point was shown between trials (inter-trial interval, ITI) for a jittered duration (following the same parameters as the CTI jitter). Each trial lasted on average 12 seconds.

The main task was divided into 4 runs. Each run contained 51 trials (48 regular and 3 catch trials). Of the 48 regular trials, 75% contained an informative retro-cue, and the remaining trials displayed neutral retro-cues. The S-R categories selected and unselected by the retro-cue were fully counterbalanced, resulting in 36 trials per category across the entire experiment. For instance, there were 36 trials in which Category 1 mappings were selected by the retro-cue. Of these 36 trials, in one third, the unselected mappings (that is, mappings shown in the encoding screen but not selected by the retro-cue) belonged to Category 2, another third to Category 3, and the last third to Category 4. Each run lasted around 10 minutes, and the main task, containing 204 trials, lasted around 40 minutes in total. Prior to the main task, outside of the scanner, participants performed a practice session with trials following the same structure described above with the exception that feedback was included to help familiarization. The practice session was structured in blocks of 11 trials. Participants performed these blocks until they achieved at least 9 correct responses. S-R mappings used during the practice were never used again.

After the main task, participants performed two localizer tasks aimed at obtaining a canonical representation of each S-R category in the two formats of interest (declarative and procedural). The structure of the task was almost identical in the two localizers and was designed to encourage either implementation or memorization strategies. In both localizers, trials started with an encoding screen (2000 ms) that contained two mappings of the same S-R category, followed by an inter-stimulus interval of jittered duration (same parameters as in the main task). Last, a target screen appeared (1500 ms) followed by a jittered ITI. The target screen differed in the two localizers and was inspired by previous studies investigating the dissociation of implementing vs. memorizing new instructions[2,3,16]. In the procedural localizer, the target was identical to the one in the main task. It consisted of a single image that prompted participants to execute the associated response. The declarative localizer, in contrast, displayed a memory probe consisting of one image and one response finger. Participants were trained to answer whether the displayed mapping was correct (same association as the encoded one) or incorrect (different association) by pressing both left-hand buttons (when “correct”) or both right-hand buttons (when “incorrect”). Therefore, in the memorization localizer, participants never had to prepare to execute the encoded mapping but rather just maintain its information. As in the main task, catch trials consisted of new images, to which participants had to respond by pressing all 4 available buttons. Each trial lasted around 8 s on average, and each localizer contained 66 trials (15 per rule + 6 catch trials), resulting in a total of 9 minutes per localizer.

All tasks were presented in PsychoPy 2[49] running on a Windows PC and back-projected onto a screen located behind the scanner. Participants responded using an MRI-compatible button box on each hand (each button box contained two buttons, on which participants placed their index and middle fingers).

Data acquisition and preprocessing

Imaging was performed on a 3T Magnetom Trio MRI scanner (Siemens Medical Systems, Erlangen, Germany), equipped with a 64-channel head coil. T1 weighted anatomical images were obtained using a magnetization-prepared rapid acquisition gradient echo (MP-RAGE) sequence (TR=2250 ms, TE=4.18 ms, TI=900 ms, acquisition matrix=256 × 256, FOV=256 mm, flip angle=9°, voxel size=1 × 1 × 1 mm). Moreover, 2 field map images (phase and magnitude) were acquired to correct for magnetic field inhomogeneities (TR=520 ms, TE1=4.92 ms, TE2=7.38 ms, image matrix=70 x 70, FOV=210 mm, flip angle=60°, slice thickness=3 mm, voxel size=3 x 3 x 2.5 mm, distance factor=0%, 50 slices). Whole-brain functional images were obtained using an echo planar imaging (EPI) sequence (TR=1730 ms, TE=30 ms, image matrix=84 × 84, FOV=210 mm, flip angle=66°, slice thickness=2.5 mm, voxel size=2.5 x 2.5 x 2.5 mm, distance factor=0%, 50 slices) with slice acceleration factor 2 (Simultaneous Multi-Slice acquisition). Slices were orientated along the AC-PC line for each subject.

For each run of the main task, 373 volumes were acquired, whereas 330 volumes were acquired during each localizer. In all cases, the first 8 volumes were discarded to allow for (1) signal stabilization, and (2) sufficient learning time for a noise cancellation algorithm (OptoACTIVE, Optoacoustics Ltd, Moshav Mazor, Israel). Before data preprocessing, DICOM images obtained from the scanner were converted into NIfTI files using HeuDiConv (https://github.com/nipy/heudiconv), in order to organize the dataset in accordance with the BIDS format[50]. Further data preprocessing was performed in SPM12 (v7487) running on Matlab R2016b. First, anatomical images were defaced to ensure anonymization. They were later segmented into gray matter, white matter and cerebro-spinal fluid components using SPM default parameters. In this step, we obtained inverse and forward deformation fields to later (1) normalize functional images to the atlas space (forward transformation) and (2) transform ROIs from the atlas on to the individual, native space of each participant (inverse transformation). Regarding functional images, preprocessing included the following steps in the following order: (1) Images were realigned and unwarped to correct for movement artifacts (using the first scan as reference slice) and magnetic field inhomogeneities (using fieldmaps); (2) slice timing correction; (3) coregistration with T1 (intra-subject registration): rigid-body transformation, normalized mutual information cost function; 4th degree B-spline interpolation; (4) registration to MNI space using forward deformation fields from segmentation: MNI 2mm template space, 4th degree B-spline interpolation; and (5) smoothing (8-mm FWHM kernel). Multivariate analyses were conducted on the unsmoothed, individual subject's functional data space and results were later normalized and smoothed (in searchlight analyses) or pooled across participants (in region-of-interest analyses).

General Linear Model (GLM) estimations

Four GLMs were estimated for each participant in SPM. First, a GLM was used to assess changes in activation magnitude between informative and neutral retro-cues during the main task. A model was constructed including, for each run, regressors for the encoding screen (zero duration), informative/neutral retro-cues (with duration), informative/neutral CTI interval (with duration), probe (zero duration) and ITI interval (with duration). Trials with errors were included as a different regressor that encompassed the total duration of the trial. All regressors were convolved with a hemodynamic response function (HRF). At the population level, parameter estimates of each regressor were entered into a mixed-effects analysis. To correct for multiple comparisons, first we identified individual voxels that passed a 'height' threshold of $p < 0.001$, and then the minimum cluster size was set to the number of voxels corresponding to $p < 0.05$, FWE-corrected. This combination of thresholds has been shown to control appropriately for false-positives[51]. A second GLM was estimated on the non-normalized and unsmoothed main task data for all multivariate analyses. This GLM contained beta estimates that specified the cued/uncued S-R categories during informative retro-cues. For each participant and run, a model was built including the following regressors: encoding (zero duration), neutral retro-cues (with duration), probes (zero duration), CTI and ITI (with duration). For informative retro-cues, a regressor that encompassed the total duration of the retro-cue was created for each S-R category combination (e.g. CuedCategory1_UncuedCategory2), resulting in a total of 12 regressors (3 per category). Errors were included as a different regressor encompassing the full duration of the trial. Last, a third and fourth GLMs were performed on the non-normalized and unsmoothed data from the two localizers. For each localizer, we built a model that contained regressors for the encoding screen (zero duration), encoding-probe interval (ISI, with duration) for each S-R category (total of 4 regressors), probe (zero duration), ITI (with duration), and errors (full trial). As in the previous GLM, these models were not used in a population-level GLM and were estimated for later use in the canonical template tracking procedure.

Multivariate pattern analysis (MVPA)

MVPA was performed on the beta images of the second GLM using The Decoding Toolbox[52] (v3.99). First, to identify regions that contained information in their patterns of activity about the validity of the retro-cue (informative vs. neutral retro-cues), a whole-brain searchlight analysis was conducted using 3-voxel radius spheres and following a leave-one-run-out cross-validation scheme. In each fold, all beta images but two (one from each class) were used to train the classifier (linear support vector machine (SVM); regularization parameter = 1) which was then tested on the remaining two samples. To rule out the effect of univariate magnitude differences between classes, we z-scored the values of each condition across voxels before the analysis (therefore, each condition that entered the analysis had a mean activation of 0 and an s.d. of 1). The accuracy value was averaged across folds and assigned to the center voxel of each sphere. To assess significance at the population level, accuracy maps were normalized to the atlas space and smoothed. The same analysis strategy as in the GLM analysis was used to threshold the statistical map (given the magnitude of the effect, a cluster-defining threshold of $p < 0.0001$ instead of $p < 0.001$ was used, and the minimum cluster size was set to the number of voxels corresponding to $p < 0.05$, FWE-corrected).

Furthermore, to assess the boost of cued S-R categories during implementation, we carried out ROI-based multiclass decoding of S-R categories. In each fold of the leave-one-run-out procedure, we trained a classifier on the identity of the *cued* S-R category using all informative retro-cue betas but four (one from each class). The classifier was then tested on the remaining samples. The accuracy was averaged across folds. Only one decoding was performed per ROI, using all voxels. To assess significance at the population level, for each ROI, we performed an across-participant one-sample t-test against chance level (25%). We then repeated the same procedure but now training and testing the classifier on the identity of the *uncued* S-R category. Finally, we compared the decoding accuracies of cued vs. uncued categories using across-participants paired t-tests. All statistical tests were FDR-corrected for multiple comparisons.

Canonical template tracking procedure

The main goal of the current study was to assess the extent to which procedural and declarative signals were activated during implementation. To do so, we followed a canonical template tracking procedure[31]. The main rationale of this analysis was (1) to obtain canonical representations of the different S-R categories under the two different formats of interest (procedural and declarative), and later (2) estimate the extent of variance during implementation uniquely explained by each of these representations. The functional localizers performed after the main task allowed us to obtain a participant-specific canonical pattern of activation for each S-R category in declarative and procedural formats. All patterns were derived from beta weights of the GLMs described in the section General Linear Model estimations. Prior to analysis, betas were converted into t-maps and, to increase the reliability of our estimation, we performed multivariate noise normalization on each individual run of the main task and template separately[53]. To do so, we used the residuals of each participant's GLMs to estimate the noise covariance between voxels. These estimates, regularized by the optimal shrinkage factor[54], were used to spatially pre-whiten the t-maps.

To measure the reactivation of the canonical patterns during the main task, for each region, we computed the semi-partial correlation between the pattern of activity during the retro-cue in the main task and the canonical template of each S-R category in the two formats. Since our GLM included different retro-cue regressors depending on the selected S-R category, we could obtain a specific reactivation value for cued, uncued and not-presented categories. Importantly, semi-partial correlations were used to obtain the amount of variance shared between the main task and a template of an S-R category (e.g. in procedural state) that is not explained by the template of that same category in the opposite state (e.g. declarative). To statistically test the boost of cued information, we first normalized the semi-correlation scores by using Fisher's z transformation and then performed paired t-tests between the cued, uncued and not-presented S-R categories activation (FDR-corrected for multiple comparisons).

Region-of-interest (ROI) definition

Frontoparietal ROIs were obtained from a parcellated map of the multiple-demand network[55]. Specifically, frontal ROIs comprised the inferior and middle frontal gyrus regions of the map, and parietal ROIs comprised the inferior and superior parietal cortex regions. All ROIs were registered back to the native space of each subject using the inverse deformation fields obtained during segmentation.

We obtained a ventral visual cortex ROI by extracting the following regions in the WFU pickatlas software (http://fmri.wfubmc.edu/software/PickAtlas): bilateral inferior occipital lobe, parahippocampal gyrus, fusiform gyrus, and lingual gyrus (all bilateral and based on AAL definitions). The primary motor cortex ROI was also obtained using WFU pickatlas by extracting the bilateral M1 region.

Data availability

The data that support the findings of this study are available from the corresponding author upon reasonable request.

References

1. Cole, M. W., Laurent, P. & Stocco, A. 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 (2013).

2. Liefooghe, B. & De Houwer, J. Automatic effects of instructions do not require the intention to execute these instructions. *J. Cogn. Psychol.* 1–14 (2018). doi:10.1080/20445911.2017.1365871

3. Liefooghe, B., Wenke, D. & De Houwer, J. Instruction-based task-rule congruency effects. *J. Exp. Psychol. Learn. Mem. Cogn.* **38**, 1325–1335 (2012).

4. Liefooghe, B., Houwer, J. De & Wenke, D. Instruction-based response activation depends on task preparation. *Psychon. Bull. Rev.* **20**, 481–487 (2013).

5. Meiran, N., Cole, M. W. & Braver, T. S. When planning results in loss of control: intention-based reflexivity and working-memory. *Front. Hum. Neurosci.* **6**, 104 (2012).

6. Meiran, N., Pereg, M., Kessler, Y., Cole, M. W. & Braver, T. S. The power of instructions: Proactive configuration of stimulus–response translation. *J. Exp. Psychol. Learn. Mem. Cogn.* **41**, 768–786 (2015).

7. González-García, C., Formica, S., Liefooghe, B. & Brass, M. Attentional prioritization reconfigures novel instructions into action-oriented task sets.

Cognition **194**, 104059 (2020).

8. Everaert, T., Theeuwes, M., Liefooghe, B. & De Houwer, J. Automatic motor activation by mere instruction. *Cogn. Affect. Behav. Neurosci.* **14**, 1300–1309 (2014).

9. Meiran, N., Pereg, M., Kessler, Y., Cole, M. W. & Braver, T. S. Reflexive activation of newly instructed stimulus–response rules: evidence from lateralized readiness potentials in no-go trials. *Cogn. Affect. Behav. Neurosci.* **15**, 365–373 (2015).

10. Demanet, J. *et al.* 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 (2016).

11. González-García, C., Arco, J. E., Palenciano, A. F., Ramírez, J. & Ruz, M. Encoding, preparation and implementation of novel complex verbal instructions. *Neuroimage* **148**, 264–273 (2017).

12. Hartstra, E., Kühn, S., Verguts, T. & Brass, M. The implementation of verbal instructions: An fMRI study. *Hum. Brain Mapp.* **32**, 1811–1824 (2011).

13. Palenciano, A. F., González-García, C., Arco, J. E. & Ruz, M. Transient and Sustained Control Mechanisms Supporting Novel Instructed Behavior. *Cereb. Cortex* bhy273 (2018). doi:10.1093/cercor/bhy273

14. Palenciano, A. F., González-García, C., Arco, J. E., Pessoa, L. & Ruz, M. Representational organization of novel task sets during proactive encoding.

J. Neurosci. 719–725 (2019). doi:10.1523/JNEUROSCI.0725-19.2019

15. Bourguignon, N. J., Braem, S., Hartstra, E., De Houwer, J. & Brass, M. Encoding of Novel Verbal Instructions for Prospective Action in the Lateral Prefrontal Cortex: Evidence from Univariate and Multivariate Functional Magnetic Resonance Imaging Analysis. *J. Cogn. Neurosci.* **30**, 1170–1184 (2018).

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

17. Woolgar, A., Afshar, S., Williams, M. A. & Rich, A. N. Flexible Coding of Task Rules in Frontoparietal Cortex: An Adaptive System for Flexible Cognitive Control. *J. Cogn. Neurosci.* **27**, 1895–1911 (2015).

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

19. Bhandari, A. & Duncan, J. Goal neglect and knowledge chunking in the construction of novel behaviour. *Cognition* **130**, 11–30 (2014).

20. Brass, M., Liefooghe, B., Braem, S. & De Houwer, J. Following new task instructions: Evidence for a dissociation between knowing and doing. *Neurosci. Biobehav. Rev.* **81**, 16–28 (2017).

21. Yu, Q. & Postle, B. R. Different states of priority recruit different neural codes

in visual working memory. *bioRxiv* 334920 (2018). doi:10.1101/334920

22. Myers, N. E., Chekroud, S. R., Stokes, M. G. & Nobre, A. C. Benefits of flexible prioritization in working memory can arise without costs. *J. Exp. Psychol. Hum. Percept. Perform.* **44**, 398–411 (2018).

23. Myers, N. E., Stokes, M. G. & Nobre, A. C. Prioritizing Information during Working Memory: Beyond Sustained Internal Attention. *Trends Cogn. Sci.* **21**, 449–461 (2017).

24. Jackson, J. B. & Woolgar, A. Adaptive coding in the human brain: Distinct object features are encoded by overlapping voxels in frontoparietal cortex. *Cortex* **108**, 25–34 (2018).

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

26. Morey, R. D. Confidence Intervals from Normalized Data: A correction to Cousineau (2005). *Tutor. Quant. Methods Psychol.* (2008). doi:10.20982/tqmp.04.2.p061

27. Jeffreys, H. *The theory of probability*. (OUP Oxford, 1998).

28. Townsend, J. & Ashby, F. G. *Stochastic modeling of elementary psychological processes*. (Cambridge: Cambridge University Press., 1983).

29. Ruge, H. & Wolfensteller, U. Rapid Formation of Pragmatic Rule Representations in the Human Brain during Instruction-Based Learning. *Cereb. Cortex* **20**, 1656–1667 (2010).

30. Bhandari, A., Gagne, C. & Badre, D. Just above Chance: Is It Harder to Decode Information from Human Prefrontal Cortex Blood Oxygenation Level-dependent Signals? *J. Cogn. Neurosci.* 1–26 (2018). doi:10.1162/jocn_a_01291

31. Wimber, M., Alink, A., Charest, I., Kriegeskorte, N. & Anderson, M. C. Retrieval induces adaptive forgetting of competing memories via cortical pattern suppression. *Nat. Neurosci.* **18**, 582–589 (2015).

32. Hartstra, E., Waszak, F. & Brass, M. The implementation of verbal instructions: Dissociating motor preparation from the formation of stimulus–response associations. *Neuroimage* **63**, 1143–1153 (2012).

33. De Baene, W., Albers, A. M. & Brass, M. The what and how components of cognitive control. *Neuroimage* **63**, 203–211 (2012).

34. Muhle-Karbe, P. S., Andres, M. & Brass, M. Transcranial Magnetic Stimulation Dissociates Prefrontal and Parietal Contributions to Task Preparation. *J. Neurosci.* **34**, 12481–12489 (2014).

35. Wisniewski, D., Reverberi, C., Tusche, A. & Haynes, J.-D. The Neural Representation of Voluntary Task-Set Selection in Dynamic Environments. *Cereb. Cortex* **25**, 4715–4726 (2015).

36. Nobre, A. C. *et al.* Orienting Attention to Locations in Perceptual Versus Mental Representations. *J. Cogn. Neurosci.* **16**, 363–373 (2004).

37. Tamber-Rosenau, B. J., Esterman, M., Chiu, Y.-C. & Yantis, S. Cortical

Mechanisms of Cognitive Control for Shifting Attention in Vision and Working Memory. *J. Cogn. Neurosci.* **23**, 2905–2919 (2011).

38. Lepsien, J. & Nobre, A. C. Attentional Modulation of Object Representations in Working Memory. *Cereb. Cortex* **17**, 2072–2083 (2007).

39. Kuo, B.-C., Stokes, M. G., Murray, A. M. & Nobre, A. C. Attention Biases Visual Activity in Visual Short-term Memory. *J. Cogn. Neurosci.* **26**, 1377–1389 (2014).

40. González-García, C., Mas-Herrero, E., de Diego-Balaguer, R. & Ruz, M. Task-specific preparatory neural activations in low-interference contexts. *Brain Struct. Funct.* (2015). doi:10.1007/s00429-015-1141-5

41. Chatham, C. H., Frank, M. J. & Badre, D. Corticostriatal Output Gating during Selection from Working Memory. *Neuron* **81**, 930–942 (2014).

42. Quentin, R. *et al.* Differential Brain Mechanisms of Selection and Maintenance of Information during Working Memory. *J. Neurosci.* **39**, 3728 LP – 3740 (2019).

43. Nichols, T. E. *et al.* Best practices in data analysis and sharing in neuroimaging using MRI. *Nat. Neurosci.* **20**, 299–303 (2017).

44. Brady, T. F., Konkle, T., Alvarez, G. A. & Oliva, A. Visual long-term memory has a massive storage capacity for object details. *Proc. Natl. Acad. Sci.* **105**, 14325–14329 (2008).

45. Brady, T. F., Konkle, T., Alvarez, G. A. & Oliva, A. Real-world objects are not

represented as bound units: Independent forgetting of different object details from visual memory. *J. Exp. Psychol. Gen.* **142**, 791 (2013).

46. Brodeur, M. B., Guérard, K. & Bouras, M. Bank of Standardized Stimuli (BOSS) phase ii: 930 new normative photos. *PLoS One* **9**, e106953 (2014).

47. Griffin, G., Holub, A. & Perona, P. *Caltech-256 object category dataset*. *Caltech Technical Report* (2006). doi:10.1021/jp953720e

48. Konkle, T., Brady, T. F., Alvarez, G. A. & Oliva, A. Conceptual distinctiveness supports detailed visual long-term memory for real-world objects. *J. Exp. Psychol. Gen.* **139**, 558 (2010).

49. Peirce, J. W. PsychoPy-Psychophysics software in Python. *J. Neurosci. Methods* (2007). doi:10.1016/j.jneumeth.2006.11.017

50. Gorgolewski, K. J. *et al.* BIDS apps: Improving ease of use, accessibility, and reproducibility of neuroimaging data analysis methods. *PLOS Comput. Biol.* **13**, e1005209 (2017).

51. Eklund, A., Nichols, T. E. & Knutsson, H. Cluster failure: Why fMRI inferences for spatial extent have inflated false-positive rates. *Proc. Natl. Acad. Sci.* **113**, 7900–7905 (2016).

52. Hebart, M. N., Görgen, K. & Haynes, J.-D. The Decoding Toolbox (TDT): a versatile software package for multivariate analyses of functional imaging data. *Front. Neuroinform.* **8**, (2015).

53. Walther, A. *et al.* Reliability of dissimilarity measures for multi-voxel pattern

analysis. *Neuroimage* **137**, 188–200 (2016).

54. Ledoit, O. & Wolf, M. A well-conditioned estimator for large-dimensional covariance matrices. *J. Multivar. Anal.* **88**, 365–411 (2004).

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

Acknowledgements

C.G.G. and S.F. were supported by the Special Research Fund of Ghent University BOF.GOA.2017.0002.03. D.W. was supported by FWO and the European Union’s Horizon 2020 Research and Innovation Program under the Marie Skłodowska-Curie grant agreement no. 665501. We thank Senne Braem for feedback on previous drafts of the manuscript.

Author contributions

All authors contributed to the design of the study. C.G.G and S.F. collected the data, which was analyzed by C.G.G. Data interpretation was done in conjunction with all other authors. C.G.G. wrote the manuscript and all authors were involved in revisions.

Competing interests

The authors declare no competing interests.