

 Open access • Posted Content • DOI:10.1101/830067

## Frontoparietal action-oriented codes support novel instruction implementation

— [Source link](#) 

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

**Institutions:** Ghent University

**Published on:** 25 Nov 2020 - bioRxiv (Cold Spring Harbor Laboratory)

**Topics:** Cognitive flexibility

Related papers:

- [Frontoparietal action-oriented codes support novel instruction implementation.](#)
- [Frontoparietal action-oriented codes support novel task set implementation](#)
- [The effects of declaratively maintaining and proactively proceduralizing novel stimulus-response mappings.](#)
- [The effect of task demands on the neural patterns generated by novel instruction encoding](#)
- [Declarative and procedural working memory updating processes are mutually facilitative](#)

Share this paper:    

View more about this paper here: <https://typeset.io/papers/frontoparietal-action-oriented-codes-support-novel-30lbegjvz4>

1

2

3

4 **Frontoparietal action-oriented codes support novel task set**

5 **implementation**

6

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

8 Department of Experimental Psychology, Ghent University, Belgium

9

10 \*Corresponding author: Carlos González-García

11 (carlos.gonzalezgarcia@ugent.be)

12 **Abstract**

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

31

## 32 INTRODUCTION

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

45 Previous behavioral studies have consistently reported an intriguing signature of  
46 instruction processing, namely, a reflexive activation of responses on the basis of  
47 merely instructed stimulus-response (S-R) associations (defined as “intention-  
48 based reflexivity”, or IBR). IBR occurs even when instructions are task-irrelevant  
49 and have not been overtly executed before<sup>2-7</sup>, which suggests a rapid  
50 configuration of instructed content predominantly towards action. Instruction  
51 implementation also has a profound impact on brain activity, as shown by  
52 electroencephalography and fMRI studies. In particular, the intention to execute an  
53 instruction induces automatic motor activation<sup>8,9</sup>, engages different brain regions to

54 coordinate novel stimuli and responses<sup>10-14</sup>, and alters the neural code of the  
55 encoded instruction<sup>15,16</sup>.

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

68 However, evidence for such serial coding in control regions is lacking, primarily  
69 due to the fact that previous analytical approaches were unable to track  
70 representational formats of specific nature. Previous work thus identified some  
71 properties of the FPN during the implementation of novel instructions, such as  
72 enhanced decoding of stimulus category<sup>11,16</sup>, or altered similarity within to-be-  
73 implemented S-R associations<sup>13,15</sup>, but failed to determine the functional state  
74 underlying such representational effects. Therefore, currently, it cannot be  
75 discerned whether novel task setting is achieved through the proposed  
76 frontoparietal formatting. In fact, at least two alternatives to the serial coding

77 hypothesis could explain previous results. First, an *amplification hypothesis*  
78 disputes the notion of two independent representational states and proposes that  
79 the intention to implement rather induces deeper declarative processing of the  
80 initial semantic information conveyed by the instruction<sup>2</sup>. Under this proposal, the  
81 FPN would support instruction implementation through the preservation of relevant  
82 declarative signals rather than through a transformation of these signals into an  
83 action-oriented code. Last, an intermediate alternative concerns the possibility that  
84 implementation involves both the boost of an independent action-oriented signal  
85 and, additionally, the preservation of declarative representations. This *dual-coding*  
86 *hypothesis* thus predicts that novel task implementation is supported by non-  
87 overlapping declarative and procedural task representations in the FPN.

88 Here, we aimed at adjudicating between these three options. In the current study,  
89 participants performed a task in which 4 novel S-R associations were presented at  
90 the beginning of each trial (each S-R consisted of an image and a response finger;  
91 for instance, the picture of a cat and the word “index”). After the encoding screen, a  
92 retro-cue would select a subset of two S-Rs, prior to the onset of a target screen.  
93 Target screens displayed the image belonging to one of the selected mappings (for  
94 example, a picture of a cat), prompting participants to execute the associated  
95 response (Fig. 1). Based on recent experimental results<sup>7,21,22</sup> and theoretical  
96 models of working memory (WM)<sup>23</sup>, we assumed that retro-cues (i.e. cues that  
97 signal the relevance of one of the already encoded representations in WM) would  
98 prioritize relevant S-R associations into a behavior-optimized state, akin to  
99 implementation. As such, retro-cues served as a tool to locate in time the moment

100 after initial encoding in which implementation-specific signals should be magnified.  
101 Our primary goal was to capture which signals governed FPN activity during such  
102 implementation stage, prior to execution<sup>20</sup>. To discern the hypothesized procedural  
103 and declarative traces, we had participants perform two functional localizers that  
104 encouraged either a declarative or action-oriented maintenance of novel  
105 instructions. Using data from the localizers, we derived a canonical multivariate  
106 pattern of activity for each S-R in both declarative and procedural formats. We then  
107 assessed the extent to which these traces were independently activated in the  
108 main task, during the implementation stage.

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

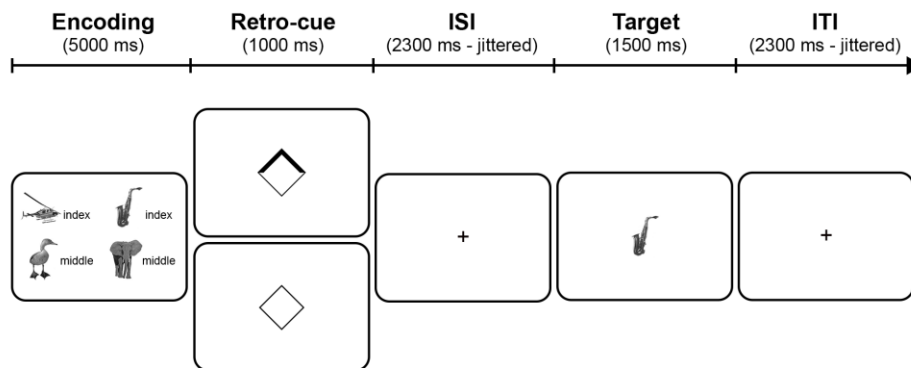
115

## 116 **RESULTS**

### 117 **Task set prioritization enhances instruction execution**

118 Twenty-nine healthy human participants (mean age = 23.28, 17 females; 3 more  
119 participants were excluded after data acquisition, see Methods) were shown 4  
120 novel S-R associations at the beginning of each trial. Importantly, even though  
121 specific S-R associations were presented only once throughout the experiment,

122 they could be grouped in categories depending on the specific combination of  
123 stimulus and response dimensions (for instance, “animate item and index finger  
124 response”; see Methods for a full description of S-R categories). Immediately after  
125 the encoding screen, a retro-cue signaled the relevance of two specific mappings  
126 (informative retro-cues in 75% of trials; in the remaining trials a neutral retro-cue  
127 did not select any mapping). The two selected mappings always belonged to the  
128 same S-R category, although the specific associations remained unique. Such  
129 grouping was crucial for analysis purposes since it allowed us to identify the  
130 *selected, unselected, and not presented* S-R categories on each trial. After the  
131 retro-cue, a target image prompted participants to provide the corresponding  
132 response (Fig. 1). To ensure that participants encoded all 4 S-R associations, ~6%  
133 of trials (regardless of the retro-cue validity) displayed a new, catch image,  
134 prompting participants to press all four available buttons simultaneously.



135

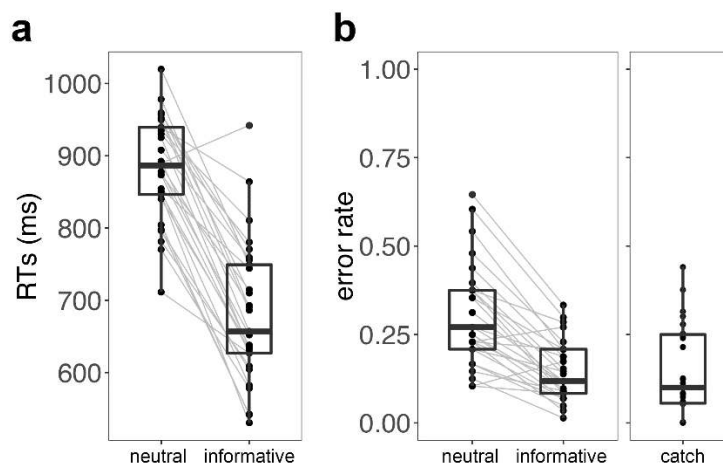
136 **Figure 1.** Behavioral paradigm. On each trial, participants first encoded four novel  
137 S-R mappings consisting in the association between an (animate or inanimate)



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

146

147 Analysis of participants’ behavioral performance revealed that retro-cues helped  
148 participants in prioritizing novel S-Rs. Specifically, participants were faster ( $t_{28,1} =$   
149 13.51,  $p < 0.001$ , Cohen’s  $d = 2.51$ ; Fig. 2a) and made less errors ( $t_{28,1} = 7.96$ ,  $p <$   
150 0.001, Cohen’s  $d = 1.47$ ; Fig. 2b, left panel) in trials with informative retro-cues,  
151 compared to neutral.



152

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

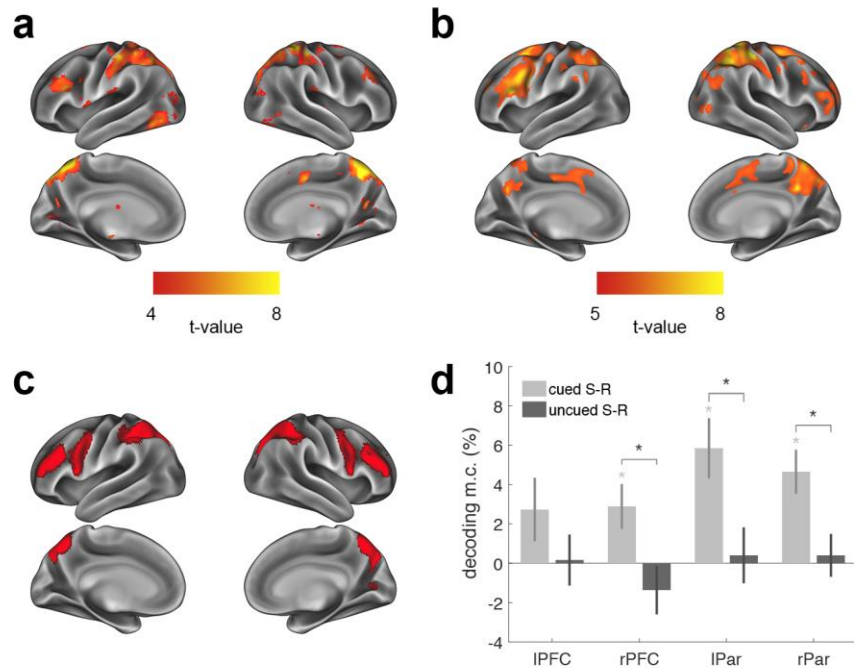
160

### 161 **Identifying task set prioritization activity**

162 As a first step, we investigated which brain regions were predominantly involved in  
163 instruction prioritization. Our intuition was that prioritization would boost  
164 implementation signals and, as such, we expected a frontoparietal network to be  
165 particularly crucial, as it is usually involved in the implementation of novel task  
166 sets<sup>11,14–17,24</sup>. We thus established a set of a priori candidate regions that  
167 encompassed frontal (inferior and middle frontal gyri) and (inferior and superior)  
168 parietal cortices (see Fig. 3c, and the Region-of-interest definition section in the  
169 Methods). We then performed two whole-brain analyses to find regions sensitive to  
170 task set prioritization (defined as informative vs. neutral retro-cues) in their overall  
171 activation magnitude or voxel-wise activity patterns, using a general linear model  
172 (GLM) and multivariate pattern analysis (MVPA), respectively. First, we found that  
173 informative retro-cues elicited significantly higher activity in regions of the FPN,  
174 including the inferior and middle frontal gyri, inferior and superior parietal cortices,  
175 as well as regions outside the FPN, such as the lateral occipital cortex (Fig. 3a,

176 primary voxel threshold [ $p < 0.001$  uncorrected] and cluster-defining threshold  
177 [FWE  $p < .05$ ]. Furthermore, a searchlight decoding analysis<sup>25</sup> revealed that the  
178 FPN contained information in its patterns of activity about the prioritization status  
179 (Fig. 3b, primary voxel threshold [ $p < 0.0001$  uncorrected] and cluster-defining  
180 threshold [FWE  $p < .05$ ]; see also Methods for details on how this analysis  
181 controlled for univariate differences in activity magnitude). Overall, the resulting  
182 statistical maps of these two analyses roughly overlap with the set of a priori  
183 defined regions of interest (ROIs; Fig. 3C), confirming the involvement of the FPN  
184 in task set prioritization.

185 To test our hypothesis that implementation would boost the representation of retro-  
186 cued S-R categories, we performed two similar decoding analyses in the 4 FPN  
187 ROIs. First, we tested if in the moment of the retro-cue the patterns of activity in  
188 these four regions carried information about the category of the cued S-R. We  
189 found significant category decoding in the right PFC and bilateral parietal ROIs  
190 (one-sample t-tests against chance level, all  $ps < 0.013$ , FDR-corrected for multiple  
191 comparisons), and close to significance decoding in the left PFC ( $t_{25,1} = 1.69$ ,  $p =$   
192  $0.052$ ). Next, we tested the extent to which the FPN also carried information about  
193 the encoded, but not cued category. In contrast with the previous results, decoding  
194 did not reach significance in any of the ROIs (all  $ps > 0.6$ ). Finally, we directly  
195 compared the decoding accuracies for the cued and uncued categories. This  
196 analysis revealed significantly stronger decoding of the cued category compared to  
197 the uncued one in right PFC and bilateral parietal cortices (paired t-tests, all  $ps <$   
198  $0.034$ , FDR-corrected; Fig. 3d).



199

200 **Figure 3.** Task set prioritization induced changes in frontoparietal neural activity.

201 (a) GLM contrast of informative > neutral retro-cue trials. Warm colors show

202 regions with significantly higher activity magnitude during informative compared to

203 neutral retro-cues (primary voxel threshold [ $p < 0.001$  uncorrected] and cluster-

204 defining threshold [FWE  $p < .05$ ]). (b) Searchlight decoding of prioritization

205 (informative vs. neutral retro-cue). Warm colors show regions with significant

206 decoding (primary voxel threshold [ $p < 0.0001$  uncorrected] and cluster-defining

207 threshold [FWE  $p < .05$ ]). (c) Set of regions-of-interest defined prior to analyses,

208 encompassing frontal (inferior and middle frontal gyri) and (inferior and superior)

209 parietal cortices. (d) Mean S-R category decoding (minus chance) within each

210 region of interest. Error bars denote between-participants s.e.m. Grey asterisks

211 denote significant decoding (chance level = 25%, one-sample t-test, FDR-

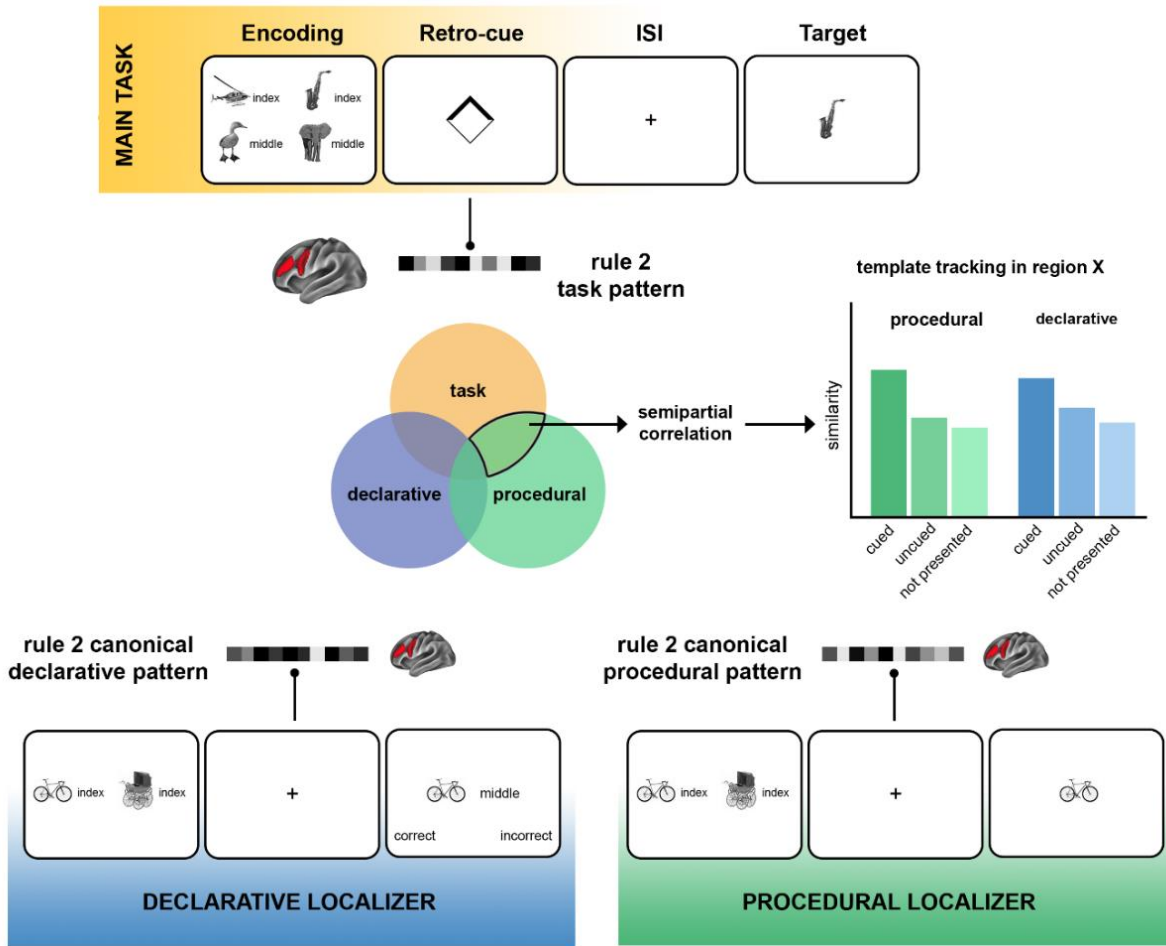
212 corrected). Black asterisks denote significantly higher decoding of cued compared  
213 to uncued S-R categories (paired t-test, FDR-corrected).

214

### 215 **Tracking format-unique task set patterns**

216 Altogether, these results show that instruction implementation has a profound  
217 impact on FPN activity, boosting the representation of prioritized task sets over  
218 encoded, but irrelevant ones. However, similarly to previous studies, they are  
219 agnostic regarding the nature of the signals underlying such effect. The main goal  
220 of our study was to test the extent to which, during this implementation stage,  
221 relevant task information was represented in a declarative and/or procedural  
222 format. In a first scenario (amplification hypothesis), implementation would merely  
223 preserve relevant declarative information. Alternatively, it could transform the initial  
224 representation of task information into a primarily action-oriented format (serial  
225 coding hypothesis). Last, action-oriented representations could coexist with  
226 preserved declarative representations (dual coding hypothesis). To adjudicate  
227 between these options, we implemented a canonical template tracking procedure  
228 that allowed us to estimate the degree of neural activation of specific S-R  
229 categories under the two functional formats of interest (see Figure 4, for a visual  
230 representation of the procedure). To do so, for each subject, we first obtained  
231 whole-brain templates of each S-R category in procedural and declarative formats,  
232 using data from two functional localizers. Subsequently, we estimated the extent to  
233 which these two traces governed the data of the main task, specifically during the  
234 presentation of informative retro-cues. We performed this step in an ROI-based

235 fashion. For each ROI and trial type, we extracted the pattern of activity during the  
236 retro-cue, keeping track of which S-R categories were either cued, uncued, or not  
237 presented in that trial. Then, we computed the semi-partial correlation between this  
238 pattern of activity and the declarative and procedural templates of each S-R  
239 category. Importantly, we used semi-partial correlations as they allowed us to  
240 estimate the amount of shared variance between task data and a given template  
241 (e.g. S-R category 1 in procedural state) that is not explained by the same  
242 template in the alternative state (e.g. S-R category 1 in declarative state).  
243 Therefore, processes common to both localizers (e.g. arousal, domain-general  
244 attention and/or task preparation) cannot inflate correlations, and any significant  
245 result rather reflects the activation of S-R information in a specific format during the  
246 main task.



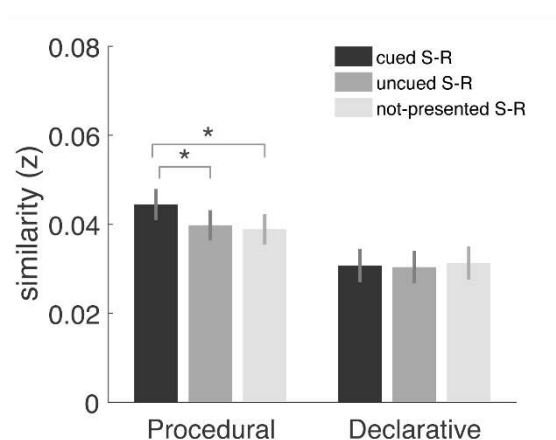
247

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

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

260

261 To validate this procedure outside the FPN, we created an ROI comprising the  
262 primary motor cortex, since predictions for this regions were straightforward: (1)  
263 boost of action-oriented information of the cued S-R category, compared to the  
264 uncued and not-presented ones; and (2) no boost of declarative information. The  
265 results obtained (Fig. 5) matched the predictions, revealing a specific  
266 enhancement of procedural information of the cued category compared to the  
267 uncued ( $t_{25,1} = 4.08$ ,  $p < 0.001$ , Cohen's  $d = 0.80$ ), and critically, to the empirical  
268 baseline defined by the not-presented categories ( $t_{25,1} = 5.45$ ,  $p < 0.001$ , Cohen's  $d$   
269  $= 1.07$ ). No reactivation of the uncued S-R category was found ( $t_{25,1} = 1.32$ ,  $p =$   
270  $0.2$ , Cohen's  $d = 0.26$ ). As predicted, no differences between cued, uncued and  
271 baseline categories were found in declarative signals (all  $t$ s  $< 1.53$ , all  $p$ s  $> 0.14$ ).



272

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



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

278

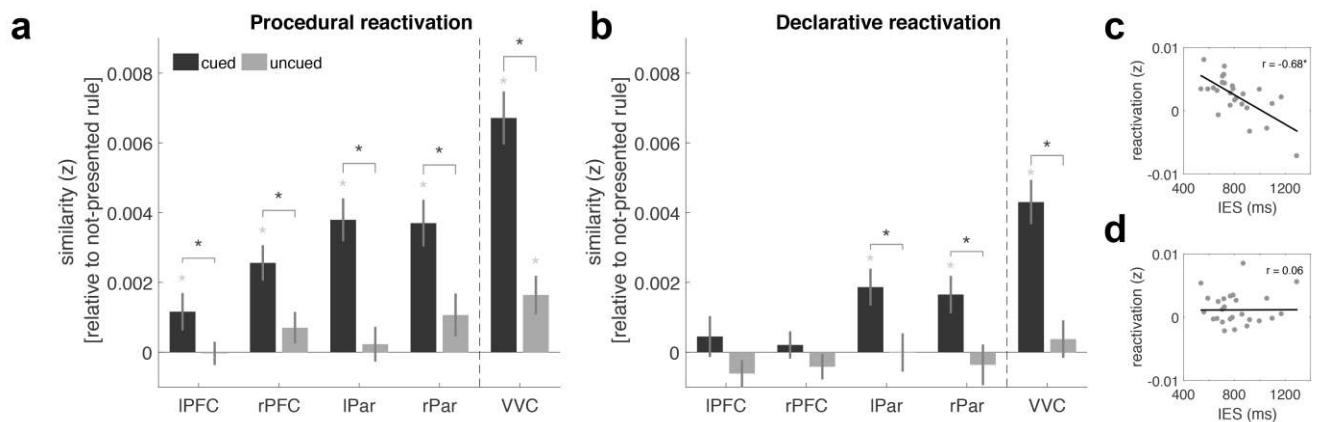
279 **Declarative and procedural representations in frontoparietal cortices (and**  
280 **beyond)**

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

286 This analysis (Fig. 6a) revealed that all FPN regions contain unique action-oriented  
287 information of relevant S-R categories during the presentation of the retro-cue  
288 (two-tail paired t-test against empirical baseline [not-presented rules], all  $t_s > 2.16$ ,  
289 all  $p_s < 0.04$ , all Cohen's  $d > 0.42$ ). Critically, procedural information of cued  
290 categories was significantly more activated than uncued categories (all  $t_s > 2.26$ ,  
291 all  $p_s < 0.04$ , all Cohen's  $d > 0.44$ ). Regarding declarative information (Fig. 6b),  
292 parietal nodes of the FPN showed a specific enhancement of declarative  
293 information of the cued S-R category, compared to the uncued one ( $t_s > 2.16$ , all  
294  $p_s < 0.02$ , all Cohen's  $d > 0.49$ ), whereas no significant differences were found in  
295 the right ( $t = 1.24$ ,  $p = 0.28$ ) and left ( $t = 2.05$ ,  $p = 0.051$ ) frontal nodes. To assess  
296 the reliability of these not significant findings, we performed Bayesian paired t-tests

297 with the same factors as before. The  $BF_{10}$  (evidence in favor of  $H_1$  against  
 298 evidence for  $H_0$ ) for the Cued – Not presented comparison was 0.27 and 0.24 for  
 299 the left and right frontal nodes, respectively. Similarly, the comparison Cued –  
 300 Uncued yielded a  $BF_{10} = 1.25$  in the left frontal node, and a  $BF_{10} = 0.41$  in the right  
 301 frontal node. Overall, this constitutes moderate evidence<sup>27</sup> for the null hypothesis  
 302 that declarative information of the cued category was not specifically enhanced in  
 303 frontal regions.

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



308

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

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

322

### 323 **Action-oriented codes support novel task setting**

324 What might be the behavioral relevance of declarative and procedural signals? We  
325 reasoned that if action-oriented representations are boosted during implementation  
326 in control-related regions, and implementation can be conceived as a behavior-  
327 optimized state, then the degree of action-oriented activation should predict the  
328 efficiency of instruction execution. To test this hypothesis, we first converted RTs  
329 and error rates of informative retro-cue trials into a single compound measure  
330 (Inverse Efficiency Scores; IES. IES were obtained by dividing each participant's  
331 mean RT by the percentage of accurate responses<sup>28</sup>). Then, we derived a  
332 template activation index by subtracting the degree of activation of cued categories  
333 to that of uncued categories for each region and format (procedural and  
334 declarative). Finally, we correlated individual IES with the activation indices on  
335 each region of the FPN. This analysis revealed significant negative correlations in  
336 all FPN regions between IES and procedural activation (all Pearson's  $r_s > -0.475$ ,

337 all  $ps < 0.02$ ). In contrast, IES did not correlate with declarative activation in any  
338 region (all  $rs < -0.34$ , all  $ps > 0.09$ ). When averaging activation indices across FPN  
339 regions, an identical pattern was found, namely, a significant correlation of IES with  
340 procedural ( $r = -0.679$ ,  $p < 0.001$ ) but not declarative ( $r = 0.06$ ,  $p = 0.77$ ) activation  
341 (Fig. 6c-d). Similar results were obtained when using RTs (procedural:  $r = -0.67$ ,  $p$   
342  $< 0.001$ ; declarative:  $r = 0.076$ ,  $p = .71$ ) and error rates (procedural:  $r = -0.54$ ,  $p =$   
343  $0.004$ ; declarative:  $r = -0.019$ ,  $p = 0.93$ ) as behavioral measures. Altogether, these  
344 results show that the more the FPN represented procedural information of relevant  
345 S-Rs, the faster and more accurate participants executed the instruction. In  
346 contrast, the strength of declarative signals of the same S-R association did not  
347 predict behavioral performance.

348

## 349 **DISCUSSION**

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

### 357 **Frontoparietal flexible coding of relevant task sets**

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

372 A first consideration concerns the exact nature of the reactivated signals. In the  
373 declarative localizer, participants had to remember specific S-R associations and  
374 match them to another S-R probe. In contrast, in the procedural localizer,  
375 participants' goal was to execute the correct response associated with a target  
376 stimulus. The different readout from WM thus encouraged different strategies, as  
377 suggested by previous studies<sup>3,7,16</sup>. Therefore, it is conceivable that templates will  
378 contain unique information: a persistent maintenance of the memoranda in the  
379 declarative localizer, and a proactive action-oriented representation, in the  
380 procedural localizer. However, templates likely share further information, for

381 instance, related to specific perceptual stimulation and general-domain processes,  
382 such as arousal or attention. We took several measures to reduce the influence of  
383 information not specifically related to declarative or procedural components. First,  
384 template reactivation was derived from semi-partial correlations between data from  
385 the main task and the localizers. Thus, our measure reflects unique shared  
386 variance between the task and the representation of an S-R category in a given  
387 localizer, partialling out the variance explained by the representation of the same  
388 S-R in the remaining localizer. Shared variance between both localizers and the  
389 main task could induce spurious similarity increases. For instance, domain-general  
390 selective attention is likely engaged towards selected mappings in the main task,  
391 as well as during the preparation interval of the localizers. Such a scenario would  
392 inflate the correlations between the templates of the cued S-R associations and the  
393 data from the main task, potentially leading to a significant difference from  
394 baseline. In contrast, semi-partial correlations ensured that procedural and  
395 declarative activation indices were derived from non-overlapping signals. Second,  
396 templates were built for S-R categories rather than unique mappings, and therefore  
397 a contribution of perceptual features to template reactivation seems unlikely.  
398 Moreover, semi-partial correlations were computed between data from the retro-  
399 cue screen (in the main task), and inter-stimulus interval (in the localizers), which  
400 reduces the likelihood of significant correlations due to perceptual similarity  
401 between templates and specific S-Rs. Therefore, we believe it is the most  
402 straightforward interpretation to consider that our procedure succeeded at tracking  
403 specific declarative and procedural signals, as also hinted by the validation results

404 in the motor cortex. From this standpoint, our results suggest that during task set  
405 implementation, FPN regions can maintain the declarative memoranda conveyed  
406 by the instruction and, simultaneously, an independent action-oriented S-R code  
407 that primarily drives task execution.

#### 408 **Heterogeneous task set coding within the FPN**

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

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

#### 449 **Implementation as a selective output gating process**



450 Remarkably, despite both signals coexisted in the FPN during implementation, only  
451 procedural representations predicted efficient behavior. The fact that  
452 implementation is signaled by retro-cues renders this effect relevant to current  
453 debates on information prioritization and WM architecture. In this regard, our  
454 results are consistent with the notion of an output gating mechanism. Similar to the  
455 idea of an input gate that limits what information enters WM, some computational  
456 models propose an additional gate that determines which pieces of this information  
457 will drive behavior<sup>41</sup>. Recent theoretical frameworks suggest a role of prioritization  
458 not only in selecting relevant content from WM but also in reformatting such  
459 content into a “behavior-guiding representational state”<sup>23</sup>, analogous to an output  
460 gating mechanism. Interestingly, these models propose that whereas other control-  
461 related regions might be involved in attention-driven representations of relevant  
462 content, frontal regions are thought to be especially important in transferring this  
463 content into a state that is optimal for behavior. In line with these ideas, we show  
464 that an action-oriented representation of task sets dominates activity in frontal  
465 cortices and that this representational format, and not a declarative one, is tightly  
466 linked to behavioral efficiency. Importantly, our results reveal, first, that the neural  
467 substrate of task set prioritization involves further brain regions, such as category-  
468 selective and parietal cortices. Second, action-oriented representations might  
469 coexist with declarative-like information in some of these regions. It should be  
470 noted, however, that fMRI data lacks the temporal resolution to discern whether  
471 these two signals fully overlap in time or whether action-oriented, behavior-  
472 optimized representations emerge after declarative information of relevant task

473 sets has been prioritized. Future studies should employ time-resolved techniques  
474 that can succeed at characterizing the dynamical contribution of different brain  
475 regions to separate control and WM processes<sup>42</sup>.

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

486

## 487 **METHODS**

488 Methods are reported, when applicable, in accordance with the Committee on Best  
489 Practices in Data Analysis and Sharing (COBIDAS) report<sup>43</sup>.

### 490 *Participants*

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

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

## 502 *Materials*

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

512 The response dimension was defined by the combination of a word ("index" or  
513 "middle") and the position of the mapping in the encoding screen. For instance, if  
514 an S-R pair containing the word "index" was displayed on the left-hand side of the  
515 screen, this informed participants that the correct response associated with that  
516 particular stimulus would be "*left index*". This allowed us to have 2 mappings on

517 screen that involved the same *response category* (e.g. index finger) but different  
518 effectors (e.g. *left* index finger vs *right* index finger).

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

#### 525 *Task and design specifications*

526 Each trial started with an encoding screen (5000 ms) that displayed 4 S-R  
527 associations. The two mappings on the upper half of the encoding screen  
528 belonged to one S-R category, and the other two belonged to another S-R  
529 category. Immediately after the encoding screen, a retro-cue appeared. Informative  
530 retro-cues (75% of trials) consisted of an arrow centered in the middle of the  
531 screen pointing either upwards or downwards. Therefore, informative retro-cues  
532 did not select a specific S-R mapping but rather two mappings belonging to the  
533 same S-R category (e.g. “animate - index finger”). Neutral retro-cues did not select  
534 any mapping. The retro-cue was displayed for 1000 ms and was followed by a  
535 fixation point (cue-target interval; CTI), which duration was jittered following a  
536 pseudo-logarithmic distribution (mean duration = 2266 ms, SD = 1276 ms, range =  
537 [600-5000]). Directly after the CTI, a target was on screen for 1500 ms. Target  
538 screens displayed the image belonging to one of the selected mappings, prompting  
539 participants to execute the associated response by pressing the corresponding

540 button in an MRI-compatible button box. In neutral trials, the target could be the  
541 stimulus of any of the 4 S-R encoded mappings. Additionally, in ~6% of trials, a  
542 catch target appeared. This consisted of a new image, different from any of the  
543 encoded stimuli, to which participants had to answer by pressing the 4 available  
544 buttons in the response box. Catch trials were included to ensure that participant  
545 encoded all four S-R associations. Last, after the target screen, a fixation point was  
546 shown between trials (inter-trial interval, ITI) for a jittered duration (following the  
547 same parameters as the CTI jitter). Each trial lasted on average 12 seconds.

548 The main task was divided into 4 runs. Each run contained 51 trials (48 regular and  
549 3 catch trials). Of the 48 regular trials, 75% contained an informative retro-cue, and  
550 the remaining trials displayed neutral retro-cues. The S-R categories selected and  
551 unselected by the retro-cue were fully counterbalanced, resulting in 36 trials per  
552 category across the entire experiment. For instance, there were 36 trials in which  
553 Category 1 mappings were selected by the retro-cue. Of these 36 trials, in one  
554 third, the unselected mappings (that is, mappings shown in the encoding screen  
555 but not selected by the retro-cue) belonged to Category 2, another third to  
556 Category 3, and the last third to Category 4. Each run lasted around 10 minutes,  
557 and the main task, containing 204 trials, lasted around 40 minutes in total. Prior to  
558 the main task, outside of the scanner, participants performed a practice session  
559 with trials following the same structure described above with the exception that  
560 feedback was included to help familiarization. The practice session was structured  
561 in blocks of 11 trials. Participants performed these blocks until they achieved at

562 least 9 correct responses. S-R mappings used during the practice were never used  
563 again.

564 After the main task, participants performed two localizer tasks aimed at obtaining a  
565 canonical representation of each S-R category in the two formats of interest  
566 (declarative and procedural). The structure of the task was almost identical in the  
567 two localizers and was designed to encourage either implementation or  
568 memorization strategies. In both localizers, trials started with an encoding screen  
569 (2000 ms) that contained two mappings of the same S-R category, followed by an  
570 inter-stimulus interval of jittered duration (same parameters as in the main task).  
571 Last, a target screen appeared (1500 ms) followed by a jittered ITI. The target  
572 screen differed in the two localizers and was inspired by previous studies  
573 investigating the dissociation of implementing vs. memorizing new instructions<sup>2,3,16</sup>.  
574 In the procedural localizer, the target was identical to the one in the main task. It  
575 consisted of a single image that prompted participants to execute the associated  
576 response. The declarative localizer, in contrast, displayed a memory probe  
577 consisting of one image and one response finger. Participants were trained to  
578 answer whether the displayed mapping was correct (same association as the  
579 encoded one) or incorrect (different association) by pressing both left-hand buttons  
580 (when “correct”) or both right-hand buttons (when “incorrect”). Therefore, in the  
581 memorization localizer, participants never had to prepare to execute the encoded  
582 mapping but rather just maintain its information. As in the main task, catch trials  
583 consisted of new images, to which participants had to respond by pressing all 4  
584 available buttons. Each trial lasted around 8 s on average, and each localizer

585 contained 66 trials (15 per rule + 6 catch trials), resulting in a total of 9 minutes per  
586 localizer.

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

### 591 *Data acquisition and preprocessing*

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

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



629 functional data space and results were later normalized and smoothed (in  
630 searchlight analyses) or pooled across participants (in region-of-interest analyses).

### 631 *General Linear Model (GLM) estimations*

632 Four GLMs were estimated for each participant in SPM. First, a GLM was used to  
633 assess changes in activation magnitude between informative and neutral retro-  
634 cues during the main task. A model was constructed including, for each run,  
635 regressors for the encoding screen (zero duration), informative/neutral retro-cues  
636 (with duration), informative/neutral CTI interval (with duration), probe (zero  
637 duration) and ITI interval (with duration). Trials with errors were included as a  
638 different regressor that encompassed the total duration of the trial. All regressors  
639 were convolved with a hemodynamic response function (HRF). At the population  
640 level, parameter estimates of each regressor were entered into a mixed-effects  
641 analysis. To correct for multiple comparisons, first we identified individual voxels  
642 that passed a 'height' threshold of  $p < 0.001$ , and then the minimum cluster size  
643 was set to the number of voxels corresponding to  $p < 0.05$ , FWE-corrected. This  
644 combination of thresholds has been shown to control appropriately for false-  
645 positives<sup>51</sup>. A second GLM was estimated on the non-normalized and unsmoothed  
646 main task data for all multivariate analyses. This GLM contained beta estimates  
647 that specified the cued/uncued S-R categories during informative retro-cues. For  
648 each participant and run, a model was built including the following regressors:  
649 encoding (zero duration), neutral retro-cues (with duration), probes (zero duration),  
650 CTI and ITI (with duration). For informative retro-cues, a regressor that  
651 encompassed the total duration of the retro-cue was created for each S-R category

652 combination (e.g. CuedCategory1\_UncuedCategory2), resulting in a total of 12  
653 regressors (3 per category). Errors were included as a different regressor  
654 encompassing the full duration of the trial. Last, a third and fourth GLMs were  
655 performed on the non-normalized and unsmoothed data from the two localizers.  
656 For each localizer, we built a model that contained regressors for the encoding  
657 screen (zero duration), encoding-probe interval (ISI, with duration) for each S-R  
658 category (total of 4 regressors), probe (zero duration), ITI (with duration), and  
659 errors (full trial). As in the previous GLM, these models were not used in a  
660 population-level GLM and were estimated for later use in the canonical template  
661 tracking procedure.

#### 662 *Multivariate pattern analysis (MVPA)*

663 MVPA was performed on the beta images of the second GLM using The Decoding  
664 Toolbox<sup>52</sup> (v3.99). First, to identify regions that contained information in their  
665 patterns of activity about the validity of the retro-cue (informative vs. neutral retro-  
666 cues), a whole-brain searchlight analysis was conducted using 3-voxel radius  
667 spheres and following a leave-one-run-out cross-validation scheme. In each fold,  
668 all beta images but two (one from each class) were used to train the classifier  
669 (linear support vector machine (SVM); regularization parameter = 1) which was  
670 then tested on the remaining two samples. To rule out the effect of univariate  
671 magnitude differences between classes, we z-scored the values of each condition  
672 across voxels before the analysis (therefore, each condition that entered the  
673 analysis had a mean activation of 0 and an s.d. of 1). The accuracy value was  
674 averaged across folds and assigned to the center voxel of each sphere. To assess

675 significance at the population level, accuracy maps were normalized to the atlas  
676 space and smoothed. The same analysis strategy as in the GLM analysis was  
677 used to threshold the statistical map (given the magnitude of the effect, a cluster-  
678 defining threshold of  $p < 0.0001$  instead of  $p < 0.001$  was used, and the minimum  
679 cluster size was set to the number of voxels corresponding to  $p < 0.05$ , FWE-  
680 corrected).

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

### 693 *Canonical template tracking procedure*

694 The main goal of the current study was to assess the extent to which procedural  
695 and declarative signals were activated during implementation. To do so, we  
696 followed a canonical template tracking procedure<sup>31</sup>. The main rationale of this  
697 analysis was (1) to obtain canonical representations of the different S-R categories

698 under the two different formats of interest (procedural and declarative), and later  
699 (2) estimate the extent of variance during implementation uniquely explained by  
700 each of these representations. The functional localizers performed after the main  
701 task allowed us to obtain a participant-specific canonical pattern of activation for  
702 each S-R category in declarative and procedural formats. All patterns were derived  
703 from beta weights of the GLMs described in the section General Linear Model  
704 estimations. Prior to analysis, betas were converted into t-maps and, to increase  
705 the reliability of our estimation, we performed multivariate noise normalization on  
706 each individual run of the main task and template separately<sup>53</sup>. To do so, we used  
707 the residuals of each participant's GLMs to estimate the noise covariance between  
708 voxels. These estimates, regularized by the optimal shrinkage factor<sup>54</sup>, were used  
709 to spatially pre-whiten the t-maps.

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

721 performed paired t-tests between the cued, uncued and not-presented S-R  
722 categories activation (FDR-corrected for multiple comparisons).

723 *Region-of-interest (ROI) definition*

724 Frontoparietal ROIs were obtained from a parcellated map of the multiple-demand  
725 network<sup>55</sup>. Specifically, frontal ROIs comprised the inferior and middle frontal gyrus  
726 regions of the map, and parietal ROIs comprised the inferior and superior parietal  
727 cortex regions. All ROIs were registered back to the native space of each subject  
728 using the inverse deformation fields obtained during segmentation.

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

734

735 **Data availability**

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

738

739 **References**

- 740 1. Cole, M. W., Laurent, P. & Stocco, A. Rapid instructed task learning: A new  
741 window into the human brain's unique capacity for flexible cognitive control.  
742 *Cogn. Affect. Behav. Neurosci.* **13**, 1–22 (2013).
- 743 2. Liefoghe, B. & De Houwer, J. Automatic effects of instructions do not  
744 require the intention to execute these instructions. *J. Cogn. Psychol.* 1–14  
745 (2018). doi:10.1080/20445911.2017.1365871
- 746 3. Liefoghe, B., Wenke, D. & De Houwer, J. Instruction-based task-rule  
747 congruency effects. *J. Exp. Psychol. Learn. Mem. Cogn.* **38**, 1325–1335  
748 (2012).
- 749 4. Liefoghe, B., Houwer, J. De & Wenke, D. Instruction-based response  
750 activation depends on task preparation. *Psychon. Bull. Rev.* **20**, 481–487  
751 (2013).
- 752 5. Meiran, N., Cole, M. W. & Braver, T. S. When planning results in loss of  
753 control: intention-based reflexivity and working-memory. *Front. Hum.*  
754 *Neurosci.* **6**, 104 (2012).
- 755 6. Meiran, N., Pereg, M., Kessler, Y., Cole, M. W. & Braver, T. S. The power of  
756 instructions: Proactive configuration of stimulus–response translation. *J. Exp.*  
757 *Psychol. Learn. Mem. Cogn.* **41**, 768–786 (2015).
- 758 7. González-García, C., Formica, S., Liefoghe, B. & Brass, M. Attentional  
759 prioritization reconfigures novel instructions into action-oriented task sets.

- 760            *Cognition* **194**, 104059 (2020).
- 761    8.    Everaert, T., Theeuwes, M., Liefvooghe, B. & De Houwer, J. Automatic motor  
762            activation by mere instruction. *Cogn. Affect. Behav. Neurosci.* **14**, 1300–  
763            1309 (2014).
- 764    9.    Meiran, N., Pereg, M., Kessler, Y., Cole, M. W. & Braver, T. S. Reflexive  
765            activation of newly instructed stimulus–response rules: evidence from  
766            lateralized readiness potentials in no-go trials. *Cogn. Affect. Behav.*  
767            *Neurosci.* **15**, 365–373 (2015).
- 768    10.   Demanet, J. *et al.* There is more into ‘doing’ than ‘knowing’: The function of  
769            the right inferior frontal sulcus is specific for implementing versus memorising  
770            verbal instructions. *Neuroimage* **141**, 350–356 (2016).
- 771    11.   González-García, C., Arco, J. E., Palenciano, A. F., Ramírez, J. & Ruz, M.  
772            Encoding, preparation and implementation of novel complex verbal  
773            instructions. *Neuroimage* **148**, 264–273 (2017).
- 774    12.   Hartstra, E., Kühn, S., Verguts, T. & Brass, M. The implementation of verbal  
775            instructions: An fMRI study. *Hum. Brain Mapp.* **32**, 1811–1824 (2011).
- 776    13.   Palenciano, A. F., González-García, C., Arco, J. E. & Ruz, M. Transient and  
777            Sustained Control Mechanisms Supporting Novel Instructed Behavior.  
778            *Cereb. Cortex* bhy273 (2018). doi:10.1093/cercor/bhy273
- 779    14.   Palenciano, A. F., González-García, C., Arco, J. E., Pessoa, L. & Ruz, M.  
780            Representational organization of novel task sets during proactive encoding.

- 781 *J. Neurosci.* 719–725 (2019). doi:10.1523/JNEUROSCI.0725-19.2019
- 782 15. Bourguignon, N. J., Braem, S., Hartstra, E., De Houwer, J. & Brass, M.  
783 Encoding of Novel Verbal Instructions for Prospective Action in the Lateral  
784 Prefrontal Cortex: Evidence from Univariate and Multivariate Functional  
785 Magnetic Resonance Imaging Analysis. *J. Cogn. Neurosci.* **30**, 1170–1184  
786 (2018).
- 787 16. Muhle-Karbe, P. S., Duncan, J., Baene, W. De, Mitchell, D. J. & Brass, M.  
788 Neural Coding for Instruction-Based Task Sets in Human Frontoparietal and  
789 Visual Cortex. *Cereb. Cortex* bhw032 (2016). doi:10.1093/cercor/bhw032
- 790 17. Woolgar, A., Afshar, S., Williams, M. A. & Rich, A. N. Flexible Coding of Task  
791 Rules in Frontoparietal Cortex: An Adaptive System for Flexible Cognitive  
792 Control. *J. Cogn. Neurosci.* **27**, 1895–1911 (2015).
- 793 18. Duncan, J., Emslie, H., Williams, P., Johnson, R. & Freer, C. Intelligence and  
794 the frontal lobe: the organization of goal-directed behavior. *Cogn. Psychol.*  
795 **30**, 257–303 (1996).
- 796 19. Bhandari, A. & Duncan, J. Goal neglect and knowledge chunking in the  
797 construction of novel behaviour. *Cognition* **130**, 11–30 (2014).
- 798 20. Brass, M., Liefoghe, B., Braem, S. & De Houwer, J. Following new task  
799 instructions: Evidence for a dissociation between knowing and doing.  
800 *Neurosci. Biobehav. Rev.* **81**, 16–28 (2017).
- 801 21. Yu, Q. & Postle, B. R. Different states of priority recruit different neural codes



- 802 in visual working memory. *bioRxiv* 334920 (2018). doi:10.1101/334920
- 803 22. Myers, N. E., Chekroud, S. R., Stokes, M. G. & Nobre, A. C. Benefits of  
804 flexible prioritization in working memory can arise without costs. *J. Exp.*  
805 *Psychol. Hum. Percept. Perform.* **44**, 398–411 (2018).
- 806 23. Myers, N. E., Stokes, M. G. & Nobre, A. C. Prioritizing Information during  
807 Working Memory: Beyond Sustained Internal Attention. *Trends Cogn. Sci.*  
808 **21**, 449–461 (2017).
- 809 24. Jackson, J. B. & Woolgar, A. Adaptive coding in the human brain: Distinct  
810 object features are encoded by overlapping voxels in frontoparietal cortex.  
811 *Cortex* **108**, 25–34 (2018).
- 812 25. Kriegeskorte, N., Goebel, R. & Bandettini, P. Information-based functional  
813 brain mapping. *Proc. Natl. Acad. Sci. U. S. A.* **103**, 3863–3868 (2006).
- 814 26. Morey, R. D. Confidence Intervals from Normalized Data: A correction to  
815 Cousineau (2005). *Tutor. Quant. Methods Psychol.* (2008).  
816 doi:10.20982/tqmp.04.2.p061
- 817 27. Jeffreys, H. *The theory of probability*. (OUP Oxford, 1998).
- 818 28. Townsend, J. & Ashby, F. G. *Stochastic modeling of elementary*  
819 *psychological processes*. (Cambridge: Cambridge University Press., 1983).
- 820 29. Ruge, H. & Wolfensteller, U. Rapid Formation of Pragmatic Rule  
821 Representations in the Human Brain during Instruction-Based Learning.  
822 *Cereb. Cortex* **20**, 1656–1667 (2010).

- 823 30. Bhandari, A., Gagne, C. & Badre, D. Just above Chance: Is It Harder to  
824 Decode Information from Human Prefrontal Cortex Blood Oxygenation Level-  
825 dependent Signals? *J. Cogn. Neurosci.* 1–26 (2018).  
826 doi:10.1162/jocn\_a\_01291
- 827 31. Wimber, M., Alink, A., Charest, I., Kriegeskorte, N. & Anderson, M. C.  
828 Retrieval induces adaptive forgetting of competing memories via cortical  
829 pattern suppression. *Nat. Neurosci.* **18**, 582–589 (2015).
- 830 32. Hartstra, E., Waszak, F. & Brass, M. The implementation of verbal  
831 instructions: Dissociating motor preparation from the formation of stimulus–  
832 response associations. *Neuroimage* **63**, 1143–1153 (2012).
- 833 33. De Baene, W., Albers, A. M. & Brass, M. The what and how components of  
834 cognitive control. *Neuroimage* **63**, 203–211 (2012).
- 835 34. Muhle-Karbe, P. S., Andres, M. & Brass, M. Transcranial Magnetic  
836 Stimulation Dissociates Prefrontal and Parietal Contributions to Task  
837 Preparation. *J. Neurosci.* **34**, 12481–12489 (2014).
- 838 35. Wisniewski, D., Reverberi, C., Tusche, A. & Haynes, J.-D. The Neural  
839 Representation of Voluntary Task-Set Selection in Dynamic Environments.  
840 *Cereb. Cortex* **25**, 4715–4726 (2015).
- 841 36. Nobre, A. C. *et al.* Orienting Attention to Locations in Perceptual Versus  
842 Mental Representations. *J. Cogn. Neurosci.* **16**, 363–373 (2004).
- 843 37. Tamber-Rosenau, B. J., Esterman, M., Chiu, Y.-C. & Yantis, S. Cortical

- 844 Mechanisms of Cognitive Control for Shifting Attention in Vision and Working  
845 Memory. *J. Cogn. Neurosci.* **23**, 2905–2919 (2011).
- 846 38. Lepsien, J. & Nobre, A. C. Attentional Modulation of Object Representations  
847 in Working Memory. *Cereb. Cortex* **17**, 2072–2083 (2007).
- 848 39. Kuo, B.-C., Stokes, M. G., Murray, A. M. & Nobre, A. C. Attention Biases  
849 Visual Activity in Visual Short-term Memory. *J. Cogn. Neurosci.* **26**, 1377–  
850 1389 (2014).
- 851 40. González-García, C., Mas-Herrero, E., de Diego-Balaguer, R. & Ruz, M.  
852 Task-specific preparatory neural activations in low-interference contexts.  
853 *Brain Struct. Funct.* (2015). doi:10.1007/s00429-015-1141-5
- 854 41. Chatham, C. H., Frank, M. J. & Badre, D. Corticostriatal Output Gating  
855 during Selection from Working Memory. *Neuron* **81**, 930–942 (2014).
- 856 42. Quentin, R. *et al.* Differential Brain Mechanisms of Selection and  
857 Maintenance of Information during Working Memory. *J. Neurosci.* **39**, 3728  
858 LP – 3740 (2019).
- 859 43. Nichols, T. E. *et al.* Best practices in data analysis and sharing in  
860 neuroimaging using MRI. *Nat. Neurosci.* **20**, 299–303 (2017).
- 861 44. Brady, T. F., Konkle, T., Alvarez, G. A. & Oliva, A. Visual long-term memory  
862 has a massive storage capacity for object details. *Proc. Natl. Acad. Sci.* **105**,  
863 14325–14329 (2008).
- 864 45. Brady, T. F., Konkle, T., Alvarez, G. A. & Oliva, A. Real-world objects are not

- 865 represented as bound units: Independent forgetting of different object details  
866 from visual memory. *J. Exp. Psychol. Gen.* **142**, 791 (2013).
- 867 46. Brodeur, M. B., Guérard, K. & Bouras, M. Bank of Standardized Stimuli  
868 (BOSS) phase ii: 930 new normative photos. *PLoS One* **9**, e106953 (2014).
- 869 47. Griffin, G., Holub, A. & Perona, P. *Caltech-256 object category dataset*.  
870 *Caltech Technical Report* (2006). doi:10.1021/jp953720e
- 871 48. Konkle, T., Brady, T. F., Alvarez, G. A. & Oliva, A. Conceptual  
872 distinctiveness supports detailed visual long-term memory for real-world  
873 objects. *J. Exp. Psychol. Gen.* **139**, 558 (2010).
- 874 49. Peirce, J. W. PsychoPy-Psychophysics software in Python. *J. Neurosci.*  
875 *Methods* (2007). doi:10.1016/j.jneumeth.2006.11.017
- 876 50. Gorgolewski, K. J. *et al.* BIDS apps: Improving ease of use, accessibility, and  
877 reproducibility of neuroimaging data analysis methods. *PLOS Comput. Biol.*  
878 **13**, e1005209 (2017).
- 879 51. Eklund, A., Nichols, T. E. & Knutsson, H. Cluster failure: Why fMRI  
880 inferences for spatial extent have inflated false-positive rates. *Proc. Natl.*  
881 *Acad. Sci.* **113**, 7900–7905 (2016).
- 882 52. Hebart, M. N., Görden, K. & Haynes, J.-D. The Decoding Toolbox (TDT): a  
883 versatile software package for multivariate analyses of functional imaging  
884 data. *Front. Neuroinform.* **8**, (2015).
- 885 53. Walther, A. *et al.* Reliability of dissimilarity measures for multi-voxel pattern

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

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

889 55. Fedorenko, E., Duncan, J. & Kanwisher, N. Broad domain generality in focal  
890 regions of frontal and parietal cortex. *Proc. Natl. Acad. Sci.* **110**, 16616–  
891 16621 (2013).

892

### 893 **Acknowledgements**

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

### 899 **Author contributions**

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

### 904 **Competing interests**

905 The authors declare no competing interests.