Neural representation of current and intended task sets during sequential judgements on human faces

Authors: Paloma Díaz-Gutiérrez[1], Sam J. Gilbert[2], Juan E. Arco[1], Alberto Sobrado[1] & María Ruz[1]

[1]Mind, Brain and Behavior Center, University of Granada

[2]Institute of Cognitive Neuroscience, University College London, UK

Word count: 10628

Contact information: Department of Experimental Psychology, University of Granada, 18071, Granada, Spain.

Telephone: +34 958 24 06 60. Facsimile number: +34 958 24 62 39

E-mail address: mruz@ugr.es (M. Ruz).

Abstract

Engaging in a demanding activity while holding in mind another task to be performed in the near future requires the maintenance of information about both the currently-active task set and the intended one. However, little is known about how the human brain implements such action plans. While some previous studies have examined the neural representation of current task sets and others have investigated delayed intentions, to date none has examined the representation of current and intended task sets within a single experimental paradigm. In this fMRI study, we examined the neural representation of current and intended task sets, employing sequential classification tasks on human faces. Multivariate decoding analyses showed that current task sets were represented in the orbitofrontal cortex (OFC) and fusiform gyrus (FG), while intended tasks could be decoded from lateral prefrontal cortex (lPFC). Importantly, a ventromedial region in PFC/OFC contained information about both current and delayed tasks, although cross-classification between the two types of information was not possible. These results help delineate the neural representations of current and intended task sets, and highlight the importance of ventromedial PFC/OFC for maintaining task-relevant information regardless of when it is needed.

Keywords: delayed intentions, dual-sequential task, PFC, fMRI, MVPA

1. **Introduction**

The selection and maintenance of relevant information is critical for our ability to pursue complex and hierarchically organized goals. In cases where we hold delayed intentions that need to be fulfilled later on (also known as prospective memory; Kliegel, McDaniel, & Einstein, 2008) or when we perform sequential tasks, it is necessary to represent the currently-active task and, in addition, the one to be performed later on. It is also important to switch flexibly from one task set to another (e.g. Monsell, 2003). Some studies have examined the neural representation of currently-active task sets in frontoparietal areas (e.g. Waskom, Kumaran, Gordon, Rissman, & Wagner, 2014; Woolgar, Thompson, Bor, & Duncan, 2011), while others have investigated the representations of delayed intentions suggesting a key role of medial prefrontal cortex (mPFC) in combination with more posterior areas (e.g. Gilbert, 2011; Haynes et al., 2007; Momennejad & Haynes, 2013). However, no previous study has examined the representation of current and intended task sets within a single experimental paradigm. This combination allows to investigate the extent to which currently-active and intended future task sets are represented in overlapping versus distinct brain networks, and also to contrast their activation patterns directly. Furthermore, previous studies have focused on representations of rather simple stimuli (i.e. geometric figures, objects, words, etc.; Crittenden, Mitchell, & Duncan, 2015; Waskom et al., 2014; Woolgar et al., 2011b), so it is not clear how well these findings generalize to more complex stimuli, such as human faces. In this study, we employed social categorization dual-sequential judgments on human faces to investigate the common and differential representation of current and delayed tasks.

The influence of maintaining an intended task-set on current task performance has previously been investigated with behavioural methods. These studies highlight how performance declines with an increment the number of tasks that need to be maintained,

showing that the representation of two tasks simultaneously is more demanding compared to one task only. For instance, Smith (2003) found that participants performed an ongoing task more slowly when they held in mind a pending intention, compared with performing the ongoing task alone. This behavioural effect is accompanied by changes in pupil dilation (Moyes, Sari-Sarraf, & Gilbert, 2019), which also serves as an indicator of task demands (see van der Wel & van Steenbergen, 2018). Further, dual-task costs have also been manifested in task switching paradigms, where participants must switch between two active task-sets (Monsell, 2003; Rogers & Monsell, 1995). Even when the same task is repeated from the previous trial, responses are slower and less accurate during mixed blocks (where more than one task is relevant) compared with pure blocks consisting of just one task (Marí-Beffa, Cooper, & Houghton, 2012).

Results at the neural level also indicate that the maintenance of two tasks compared with one alters activity in specific brain regions. Several studies have shown that a set of "task-positive" regions increase their activation during demanding tasks (also known as the Multiple Demand network, MD; Duncan, 2010). This network is also sensitive to cognitive load, increasing its sustained activation as task complexity is raised (Dumontheil, Thompson, & Duncan, 2011; Palenciano, González-García, Arco, & Ruz, 2019; but see Tschentscher, Mitchell, & Duncan, 2017). Among these areas, the lateral prefrontal (lPFC) and parietal cortices play a prominent role during dual-task performance. Both increase their activation during task-switching trials while anterior PFC shows sustained activation during task-switching blocks (Braver, Reynolds, & Donaldson, 2003). Similarly, others (Szameitat, Schubert, Müller, & Von Yves Cramon, 2002) have shown involvement of lPFC during dual-task blocks, proportionally to task difficulty, during simultaneous and interfering task processing. Further, some studies have employed multivoxel pattern analysis (MVPA) to show how these frontoparietal

(FP) regions code current task sets (Palenciano et al., *in press*; Qiao, Zhang, Chen, & Egner, 2017; Waskom et al., 2014; Woolgar, Hampshire, Thompson, & Duncan, 2011) and how the representation of task-relevant information in these areas increases with task demands (Woolgar et al., 2011b).

Traditionally, the role of FP regions has been opposed to "task-negative" areas, initially linked to decreased activity during effortful task performance (Fox et al., 2005), although recent studies suggest that it has a much broader role. This Default Mode Network (DMN) includes the ventro/dorsomedial PFC, orbitofrontal cortex (OFC), precuneus/posterior cingulate, inferior parietal lobe (IPL), lateral temporal cortex, and hippocampal formation (Buckner, Andrews-Hanna, & Schacter, 2008; Raichle, 2015). However, recent studies have qualified this view, showing that these regions also represent task-relevant information in different contexts (e.g. Crittenden et al., 2015; González-García et al., 2017; Palenciano et al., *in press*; Smith, Mitchell, & Duncan, 2018). Moreover, functional connectivity approaches have shown that the strength of connectivity among task-negative regions during a working memory task is associated with better performance (Hampson, Driesen, Skudlarski, Gore, & Constable, 2006). Similarly, Elton and Gao (2015) observed that the dynamics of connectivity among DMN regions during task performance were also related to behavioural efficiency. Altogether, the literature suggests a clear involvement of FP areas in the representation of current task-related information and highly demanding tasks. Conversely, the role of the DMN is less clear. Although it shows decreased activation during demanding tasks, its dynamics are also related to behaviour, and contain task information in different contexts. This suggests that these regions play a role in the representation of task-relevant knowledge.

Further, one of the main nodes of the DMN, the medial prefrontal cortex (mPFC) has an important role in the representation of intended behaviour during both task-free situations (Haynes et al., 2007) and delays concurrent with an ongoing task (Gilbert, 2011; Momennejad & Haynes, 2012; Momennejad & Haynes, 2013). This area also plays a role when holding decisions before they reach consciousness (Soon, Brass, Heinze, & Haynes, 2008). The evidence from studies of delayed intentions has led to suggested dissociations between the role of lateral and medial PFC (associated with the task-positive and task-negative networks, respectively). Momennejad & Haynes (2013) directly compared the representation of future intentions during delays with and without an ongoing task, and found that while the lPFC had a general role of encoding intentions regardless of whether there was or not an ongoing task during the delay, the mPFC was involved when the delay period was occupied by an ongoing task. Alternatively, Gilbert (2011) could not find encoding of delayed intentions in the lPFC but they did in the mPFC, suggesting that the former may play a content-free role in remembering delayed intentions while the latter would represent their specific content. However, these studies vary in the abstraction of the task rules employed. While Gilbert (2011) aimed to decode specific visual cues and responses, others focused on the anticipation of abstract task sets, such as arithmetic operations (addition vs. substraction; Haynes et al., 2007), or parity vs. magnitude judgements (Momennejad & Haynes, 2012; Momennejad & Haynes, 2013). This difference in abstraction could impact the brain region (lPFC vs. mPFC) maintaining information about future intentions (Momennejad & Haynes, 2013). Further, these studies also vary in whether the retrieval of the intended task was cued (Gilbert, 2011) or self-initiated (Momennejad & Haynes, 2012; Momennejad & Haynes, 2013). Therefore, although these studies have studied the representation of intentions in a variety of experimental settings, they have not directly addressed how the representation of a future

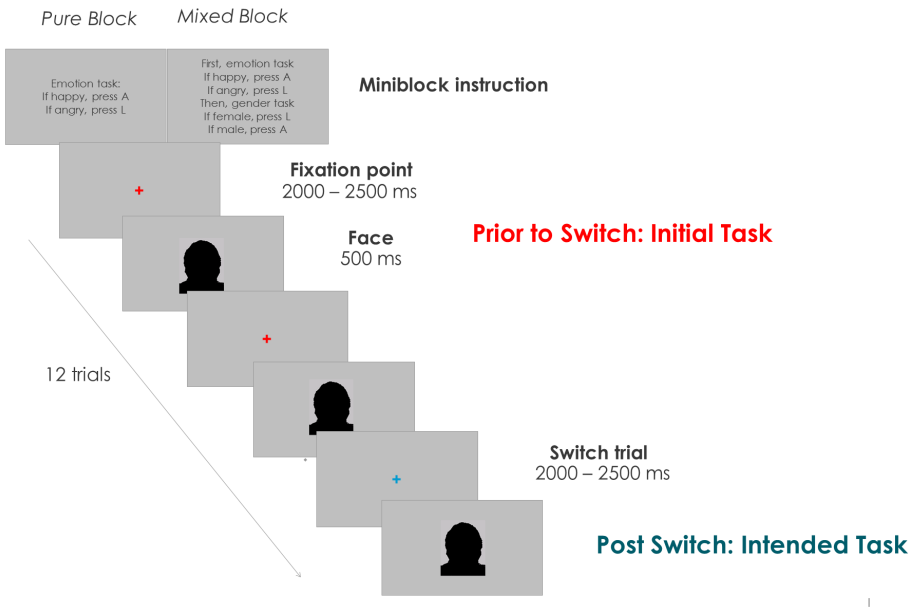
155 task set may differ from the representation of an ongoing task that is currently being
156 performed.
157
158 In addition, the studies so far have employed mainly non-social stimuli. In this context it
159 is worth noting that the DMN has also been related to processes relevant in the social
160 domain (Buckner & Carroll, 2007; Mars et al., 2012; Spreng, Mar, & Kim, 2008). For
161 instance, engagement of the DMN during rest is related to better memory for social
162 information (Meyer, Davachi, Ochsner, & Lieberman, 2018). Facial stimuli are an
163 important source of social knowledge, which is represented in a set of regions including
164 the fusiform gyri (FG; Haxby, Hoffman, & Gobbini, 2000; Kanwisher & Yovel, 2006).
165 This FG also shows different neural patterns distinguishing social categories (Kaul,
166 Ratner, & Van Bavel, 2014; Stolier & Freeman, 2017). Similarly, the representational
167 structure of social categories is altered by personal stereotypes both in the FG and in
168 higher-level areas such as the OFC (Stolier & Freeman, 2016), which is also linked to the
169 representation of social categories such as gender, race, or social status (Gilbert,
170 Swencionis, & Amodio, 2012; Kaul, Rees, & Ishai, 2011; Koski, Collins, Olson, &
171 Hospital, 2017) and the integration of contextual knowledge during face categorization
172 (Freeman et al., 2015). Likewise, during predictive face perception, the FG coactivates
173 with and receives top-down influences from dorsal and ventral mPFC (e.g. Summerfield
174 et al., 2006), which in turn have also been implicated on judgements about faces
175 (Mitchell, Macrae, & Banaji, 2006; Singer, Kiebel, Winston, Dolan, & Frith, 2004).
176 Therefore, given the special properties and influence of social information gathered from
177 faces, understanding how task-relevant current and delayed information may be
178 represented when it pertains to social information is important to extend and complement
179 previous findings.
180

In the current fMRI study, we employed a dual-sequential categorization task, where participants had to discriminate between features of three dimensions of facial stimuli and had to maintain for a period of time both the initial ongoing task and an intended one. In particular, we studied how demands (one vs. two sequential tasks) influence performance, and hypothesized that high demand would be associated with worse performance alongside with activation in frontoparietal regions, especially the lPFC. To examine the brain regions containing fine-grained information about both current and intended tasks we employed MVPA. Unlike traditional univariate methods, where the mean activation in a set of voxels is compared between conditions, MVPA focuses on the spatial distribution of activations. Here, a classifier is trained to distinguish response patterns associated with different experimental conditions (i.e. stimuli categories, cognitive states, etc.) in a certain brain region. If the trained classifier is able to predict the patterns of independent data, there is indication that the brain area under study represents specific information about those conditions. Thus, MVPA allows to examine finer-grained differences in how information is represented in the brain (for reviews see Haxby, Connolly, & Guntupalli, 2014; Haynes, 2015). In this work, we aimed to study how an intended task set might be represented differently from a currently-active ongoing task. For that reason, we focused on the initial pre-switch period, when the current task is being performed before switching to the intended task. Specifically, we performed separate analyses to decode: 1) the task currently being performed, regardless of the intended future task; 2) the task intended for the future, regardless of the current task (henceforth: "initial task" and "intended task", respectively). Given the extensive literature associating FP areas to the representation of task-relevant information (Qiao et al., 2017; Waskom et al., 2014; Woolgar et al., 2011a, 2011b), we expected to decode the initial relevant task in MD regions and the intended one in "task-negative" regions, especially the mPFC, in

206 line with previous studies showing its role in prospective memory (Gilbert, 2011; Haynes
207 et al., 2007; Momennejad & Haynes, 2012; Momennejad & Haynes, 2013) .
208
209 **2. Methods**
210 **2.1. Participants**
211 Thirty-two volunteers were recruited through adverts addressed to undergraduates and
212 postgraduate students of the University of Granada (range: 18-28, M = 22.5, SD = 2.84,
213 12 men). All of them were Caucasian, right-handed with normal or corrected-to-normal
214 vision and received economic remuneration (20-25 Euros, according to performance) in
215 exchange for their participation. Participants signed a consent form approved by the
216 Ethics Committee for Human Research of the University of Granada.
217
218 **2.2. Apparatus and stimuli**
219 We employed 24 face photographs (12 identities, 6 females, 6 black; 3 different identities
220 per sex and race) displaying happy or angry emotional expressions, extracted from the
221 NimStim dataset (Tottenham et al., 2009). E-Prime 2.0 software (Schneider, Eschman, &
222 Zuccolotto, 2002) was used to control and present the stimuli on a screen reflected on a
223 coil-mounted mirror inside the scanner.
224
225 **2.3. Design and procedure**
226 Participants had to perform a series of categorization tasks where they judged either the
227 emotion (happy vs. angry), the gender (female vs. male) or the race (black vs. white) of
228 series of facial displays. These tasks were arranged in miniblocks, which could each
229 contain one (Pure Miniblock; PM) or two sequential categorization tasks (Mixed
230 Miniblock; MM). At the beginning of each miniblock, participants received instructions
231 indicating the number of tasks to perform (1 vs. 2) and their order and nature (Emotion,

Gender and/or Race), as well as the key-response mappings. Thus, for PMs, the initial instruction indicated the one task that had to be performed during the whole miniblock. Conversely, for MMs, the instruction indicated two tasks, where the participant had to change from the first to the second task at a certain point of the miniblock. After the instruction, a coloured (blue or red) fixation point appeared on the screen, followed by a facial display (see Figure 1). Participants were told that, during MMs, they had to switch tasks when the fixation changed its colour (from blue to red or vice versa). Once it switched, they had to continue doing the second task until the end of the miniblock. To equate the perceptual conditions across blocks, the fixation colour also changed during PMs, although participants were told to ignore this change.

Hence, in each MM there was an initial task (first task to perform), an intended task (second task to perform) and an ignored task (non-relevant for that miniblock). Importantly, our main fMRI analyses focused on the period before the switch, while participants needed to represent both the initial task and the intended one. Task switches were evenly spaced across the miniblock, from trial 1 to 12. This allowed us to decorrelate brain activity associated with the pre-switch period, post-switch period, and the switch itself. In total, there were 9 different types of miniblocks: 3 pure (emotion[EE], gender[GG], race[RR]) and 6 mixed (emotion-gender[EG], emotion-race[ER], gender-emotion[GE], gender-race[GR], race-emotion[RE], race-gender[RG], see Figure 2). Across the experiment, pure miniblocks appeared 8 times each, while every type of mixed miniblock was repeated 12 times. The presentation order of the miniblocks and the assignment of response options (left or right index) were counterbalanced within each run. Additionally, to avoid response confounds in the analyses, response mappings changed between runs. Thus, for each participant odd and even runs had the opposite response mappings.

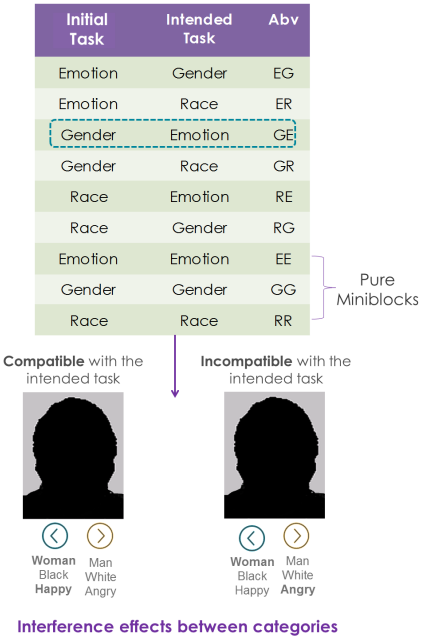


258

259 **Figure 1.** Display of the paradigm (face stimuli are obscured in the preprint version).

260 Example of a miniblock and sequence of trials. Inter-trial-interval (ITI) duration = 2-2.5 s.

261



262

263 **Figure 2.** Top: All possible combinations of miniblocks, depending on the initial and

264 intended tasks, and their abbreviation (Abv). Bottom: Example of interference between

265 initial and intended categories in a Gender-Emotion (GE) miniblock.

Participants performed a practice block to learn the different tasks and the response mappings. They were required to obtain a minimum of 80% accuracy at this practice block prior to entering the scanner. The sequence of each miniblock was as follows: First, the instruction slide presented the task/s to perform (Pure: 1, Mixed: 2), and the response mappings (right/left index), during 5 s. Then, a sequence of 12 trials appeared. In each of them, a fixation point (blue or red, counterbalanced) lasting 2-2.5 s (inter-trial-interval; ITI; in units of 0.25 s, randomly assigned to each fixation) was followed by a facial display of 0.5 s. The fixation for the switch trial lasted on average 2.24 s (SD = 0.022; all participants within a range of ± 2.5 standard deviations). The experiment consisted of 1152 trials, arranged in 96 miniblocks (72 mixed and 24 pure), distributed in 12 scanning runs. Hence, each run consisted of 8 miniblocks (6 mixed and 2 pure). Each type of miniblock was repeated 12 times, once per run, and each time the switch occurred on a different trial. Presentation order and switch point were counterbalanced through the experiment, to ensure that each switch occurred on every possible trial for each type of miniblock and that each identity was associated the same number of times with the switch. In total, the fMRI task lasted for 60.8 min.

In addition, we also studied the interference between tasks. Since in MMs the participant had to perform two tasks sequentially, the established stimulus-response association could be compatible or incompatible between the current and intended task, depending on the specific target face. For instance, the gender task could have a stimulus-response (S-R) association of female-left/male-right and the S-R in the emotion task might be happy-left/angry-right. Therefore, in a GE miniblock, during the gender task, participants could encounter happy female faces (both the initial gender and intended emotion tasks would require the same response: left) or angry female faces (the initial gender task would

lead to response with the left index and the emotion with the right one). Thus, the former would be an example of compatibility between initial and intended tasks, whereas the latter would entail incompatibility (see Figure 2).

2.4. Image acquisition and preprocessing

Volunteers were scanned with a 3T Siemens Magnetom Trio, located at the Mind, Brain and Behavior Research Center (CIMCYC) in Granada, Spain. Functional images were obtained with a T2*-weighted echo planar imaging (EPI) sequence, with a TR of 2.210 s. Forty descending slices with a thickness of 2.3 mm (20% gap) were extracted (TE = 23 ms, flip angle = 70 º, voxel size of 3x3x2.3 mm). The sequence was divided into 12 runs, consisting of 152 volumes each. Afterwards, an anatomical image for each participant was acquired using a T1-weighted sequence (TR = 2500 ms; TE = 3.69 ms; flip angle = 7º, voxel size of 1 mm^3). MRI images were preprocessed and analysed with SPM12 software (http://www.fil.ion.ucl.ac.uk/spm/software/spm12). The first 3 images of each run were discarded to allow the stabilization of the signal. The volumes were realigned and unwarped and slice-time corrected. Then, the realigned functional images were coregistered with the anatomical image and were normalized to 3 mm^3 voxels using the parameters from the segmentation of the anatomical image. Last, images were smoothed using an 8 mm Gaussian kernel, and a 128 high-pass filter was employed to remove low-frequency artefacts. Multivariate analyses used non-normalized and non-smoothed data (Bode & Haynes, 2009; Gilbert & Fung, 2018; Woolgar et al., 2011a, 2011b).

2.5. fMRI analyses

2.5.1. Univariate

First, we employed a univariate approach to examine the effect of context demands (one vs. two sequential tasks) and task switching. Our model contained, for each run, one

318 regressor for the instruction of each miniblock, four regressors corresponding to the two
319 types of miniblock (pure/mixed) with separate regressors for the pre-switch and post-
320 switch periods, one for the change in fixation colour during mixed miniblocks (indicating
321 a switch event), one for the change in fixation colour during pure miniblocks (serving as
322 a baseline for the switch events), and another one for the errors. Both instruction and
323 miniblock regressors were modelled as a boxcar function with the duration of the entire
324 pre/post switch period or instruction duration (5 s). Errors were modelled including the
325 duration of the face of that trial and the following fixation (2.5-3 s). Switch trials were
326 modelled as events, with stick functions with zero duration locked at the switch in colour
327 of the fixation point. This provided a model with a total of 8 regressors per run. At the
328 group level, t-tests were carried out for comparisons related to the effect of task demands
329 (one vs. two tasks) at the period prior to the switch, and also to compare switching cost
330 effects (switch trial in the mixed block > switch trial in pure blocks). We report clusters
331 surviving a family-wise error (FWE) cluster-level correction at $p < .05$ (from an initial
332 uncorrected threshold $p < .001$). Additionally, we also performed nonparametric
333 inference (see Supplementary Materials).
334
335 2.5.2. Multivariate analysis
336 We performed multivoxel pattern analyses (MVPA) to examine the brain areas
337 maintaining the representation of A) current-active initial tasks, and B) intended tasks.
338 These analyses examined brain activity during the pre-switch period only (although
339 additional, exploratory analyses were also performed on the post-switch period, see the
340 Supplementary Materials). Following a Least-Squares Separate Model approach (LSS;
341 Turner, 2010) we modelled each miniblock (EG, ER, GE, GR, RE, RG) during the period
342 prior to the switch separately. This method helps to reduce collinearity between regressors
343 (Abdulrahman & Henson, 2016), by fitting the standard hemodynamic response to two

regressors: one for the current event (a type of miniblock prior to the switch) and the second one for all the remaining events. As in the univariate approach, each miniblock regressor was modelled as a boxcar function with the duration of the entire pre-switch period duration.

The binary classification analyses were performed as follows. First, we classified a) between any two initial tasks while holding the intended task constant, then b) between any two intended tasks while holding the initial task constant. For instance, in case a) we contrasted initial gender task vs. initial race task when the intended task was emotion (GE vs. RE), and also intended gender vs. intended emotion task when the intended task was race (ER vs. GR), and intended emotion task vs. intended race task when the intended task was gender (EG vs. RG). We then averaged decoding accuracies across these analyses, which indicate whether a particular brain region shows different patterns of activity depending on what the initial, currently-active task set is, holding the intended task constant. Conversely, in case b), we compared intended gender vs. intended race when the initial, currently-active task was emotion (EG vs. ER), intended gender vs. intended emotion when the initial, currently-active task was race categorization (RE vs. RG) and intended emotion vs. intended race when the initial, currently-active task was gender (GE vs. GR). As above, we averaged across these analyses, which indicate whether a particular brain region shows different patterns of activity depending on what the intended task set is, holding the currently-performed task constant.

To carry out these analyses, we performed a whole brain searchlight (Kriegeskorte, Goebel, & Bandettini, 2006) on the realigned images employing the Decoding Toolbox (TDT; Hebart, Görgen, & Haynes, 2015) and custom-written MATLAB code. We created 4-voxel radius spheres and for each sphere, a linear support vector machine classifier (C

370 =1; Pereira, Mitchell, & Botvinick, 2009) was trained and tested using a leave-one-out
371 cross-validation. Due to the nature of the paradigm and the counterbalancing, once in
372 each block the switch took place at the first trial (here particcipants only performed the
373 intended task). Thus, there was an example of each type of miniblock before the switch
374 in only 11 runs, differently for each participant and miniblock (i.e. a participant could
375 lack miniblock EG in run 4 and miniblock RG in run 11). To avoid potential biases in the
376 classifier for having only one of the classes in a run, for each participant and comparison,
377 we performed the classification only in the 10 runs where there was an example of both
378 miniblocks. Resulting from this procedure, we employed the data from 10 scanning runs
379 (training was performed with data from 9 runs and tested on the remaining run, in an
380 iterative fashion). In the exceptional case (twice for each contrast in the total sample)
381 where the two miniblocks in the classification were absent in the same run, classification
382 was performed on the remaining 11 runs (training with data from 10 runs and testing on
383 the remaining run). In addition, we observed biases in the decoding estimates when the
384 switch trial from one of the conditions in the test set matched the opposite class in the
385 training set, which happened for every comparison in aproximately half of the cross-
386 validation steps. To avoid the biases resulting from this, we additionally removed those
387 runs where the switch position matched the test from the training set for that specific
388 cross-validation step.
389
390 Next, we averaged the accuracy maps for a) and b) to obtain a mean classification map
391 collapsing across initial and intended tasks. This allowed us to detect regions that
392 contained information about either initial or intended tasks (or both). It also allowed us to
393 define ROIs that could be used to compare decoding accuracies for initial versus intended
394 task-sets, in a manner that was unbiased between the two types of information. We

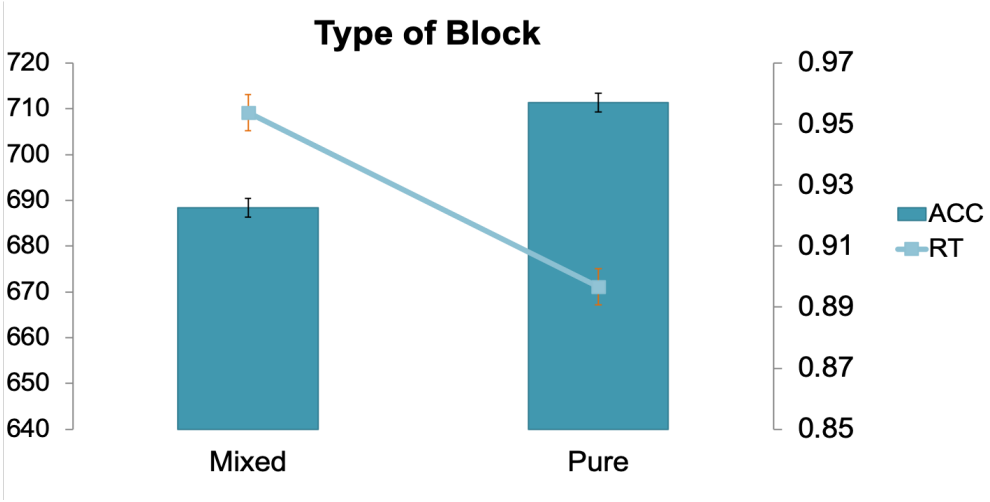
additionally conducted whole-brain analyses investigating decoding of the initial task only, decoding of the intended task only, and the comparisons between the two.

Afterwards, group analyses were performed by doing one-sample t-tests after normalising (same as for the univariate analyses) and smoothing the individual accuracy maps (4 mm Gaussian kernel, consistent with earlier MVPA studies such as Gilbert, 2011; Gilbert & Fung, 2018). Results were considered significant if they passed an FWE cluster-level correction at $p < .05$ (based on an uncorrected forming threshold of $p < .001$). This statistical approach is consistent with recent MVPA studies (Gilbert & Fung, 2018; Loose, Wisniewski, Rusconi, Goschke, & Haynes, 2017). We additionally carried out nonparametric inference (see Supplementary Materials).

3. Results

3.1. Behaviour

First, to study how the number of tasks influenced performance, we performed a paired t-test on both accuracy and reaction times (RTs), between the two types of Miniblock (Mixed/Pure), collapsing over pre- and post-switch periods (see Figure 3). Here, responses were more accurate for pure ($M = 95.7\%$, $SD = 4.3$), than for mixed miniblocks ($M = 92.3\%$, $SD = 3.6$), $t_{31} = 5.39$, $p<.001$, whereas they were faster for pure ($M = 671.12$ ms, $SD = 126.3$) than for mixed ($M = 709.18$ ms, $SD = 142.39$) miniblocks, $t_{31} = 4.83$, $p<.001$.

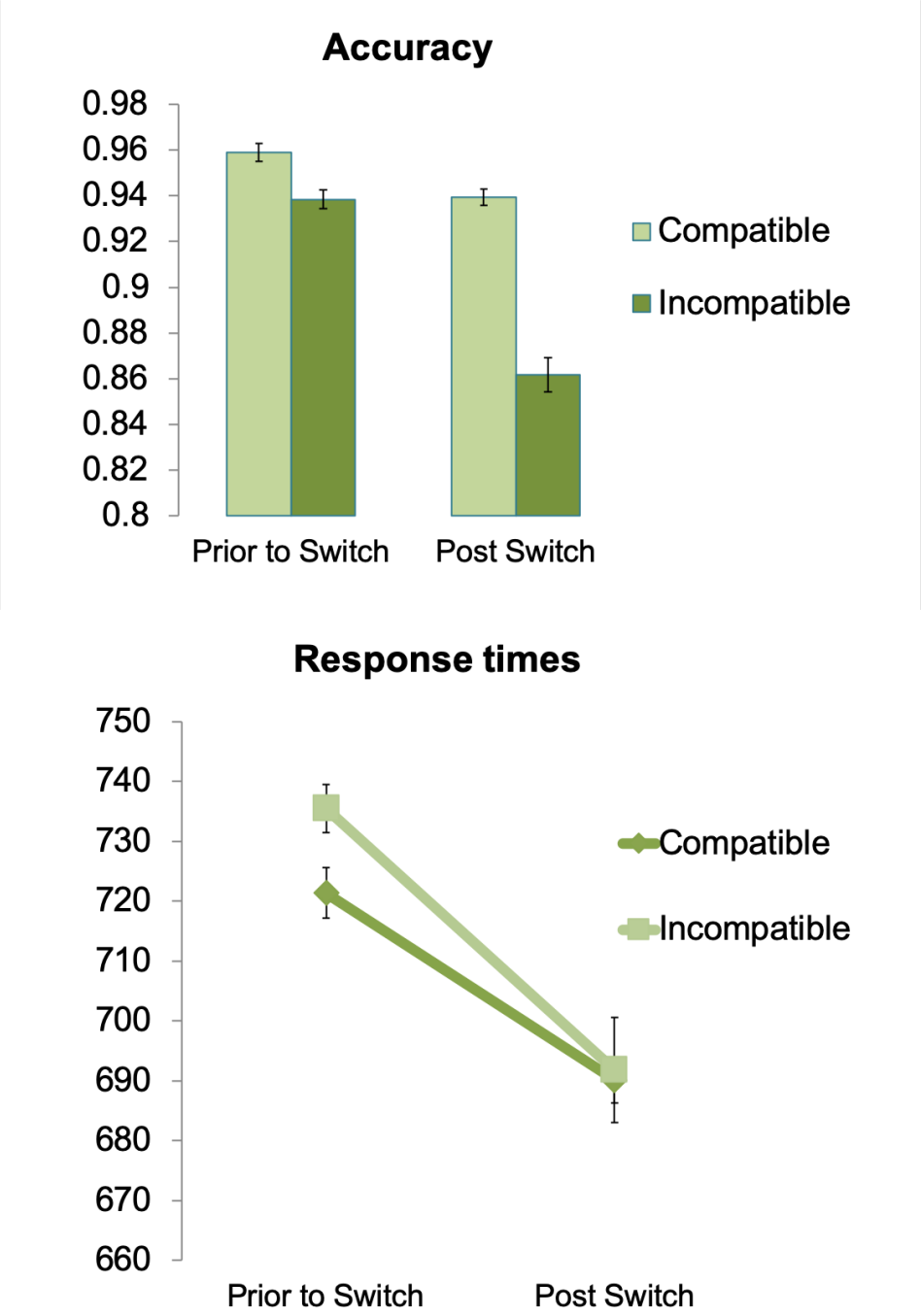


417

418 **Figure 3.** Influence of the type of block on performance. Error bars represent within-

419 subjects 95% confidence intervals (Cousineau, 2005).

420



421

422 **Figure 4.** Interference effects between initial and intended tasks before and after the
423 switch (MMs). Top: Accuracy rates. Bottom: Reaction times (ms). Error bars represent
424 within-subjects 95% confidence intervals (Cousineau, 2005).

425

426 In addition, we examined if the intended task influenced initial task performance, and
427 vice versa. For this, we selected only the mixed miniblocks and entered them into a
428 repeated measures (rm) ANOVA, with Task (Emotion/Gender/Race), Period of the

miniblock (Prior to/Post Switch), and Interference (Compatible/Incompatible) between initial and intended tasks. Note here that even if we did not have any specific hypothesis about the influence of the variable Task on performance, we included it as a factor in this second ANOVA to examine whether the Task modulated the effect of the other two variables of interest: Period of the block and Interference. Moreover, we refer to initial and intended as the tasks performed before and after the switch, respectively, to preserve consistency in the terminology throughout the entire manuscript. In addition, we refer as compatible trials when the correct response for the initial task-relevant dimension was associated with the same response for the intended dimension (see an example in Figure 1, right), and incompatible trials when the response associated with the initial task-relevant dimension interfered with the responses associated with the intended one. Similarly, after the switch, when the intended task was being performed, compatible trials referred to those where the correct responses for this task were associated with the same response for the previous initial task, and incompatible trials when the response associated with both dimensions differed. This way we could use interference effects as an indicator of the maintenance of the intended response dimensions during performance.

Accuracy did not show any main effect of Task ($F<1$, see Figure 4). However, we observed a main effect of Period of the miniblock, $F_{1,31} = 58.215$, $p < .001$, $\eta_p^2 = .653$, where participants responded more accurately before (M = 94.86%, SD = 3.6) than after the task switch (M = 90.05%, SD = 5.4). There was also a main effect of Interference, $F_{1,31} = 101.83$, $p < .001$, $\eta_p^2 = .767$, where accuracy was higher for compatible (M = 94.91%, SD = 4.76) than for incompatible trials (M = 90%, SD = 6.79). The interaction Task x Period was significant, $F_{2,62} = 3.831$, $p = .029$, $\eta_p^2 = .110$, showing that performance was better before than after the switch for all three tasks (all $Fs > 15$, $ps < .001$), but this difference was larger for the gender task ($\eta_p^2 = .679$). Similarly, the

455 interaction of Period x Interference was significant, $F_{1,31} = 52.244$, $p < .001$, $\eta_p^2 = .106$,
456 where accuracy was worse for incompatible compared to compatible trials (both Fs > 19,
457 ps < .001), but this pattern was more pronounced after ($F_{1,31} = 107.08$, p<.001) than before
458 the switch ($F_{1,31} = 19.21$, p<.001). No other interactions reached significance (p>.061).
459
460 RTs showed (see Figure 4) a main effect for Task ($F_{2,62} = 24.08$, $p < .001$, $\eta_p^2 = .437$),
461 were race was performed faster (M = 691.16, SD = 150.07), followed by gender (M =
462 709.29, SD = 147.11), and emotion (M = 728.59, SD = 144.14). In addition, we also
463 observed a main effect of Period ($F_{1,31} = 32.83$, $p < .001$, $\eta_p^2 = .514$), as participants were
464 faster after (M = 690.94, SD = 143.38) than prior to the switch (M = 728.42, SD = 155.72).
465 Further, we found a main effect of Interference, $F_{1,31} = 4.829$, p =.036), where participants
466 were faster for compatible (M = 705.51, SD = 147.79) than incompatible trials (M =
467 713.65, SD = 146.42). An interaction Period x Interference ($F_{1,32} = 24.08$, $p = .033$, $\eta_p^2 =$
468 .437) showed that this interference effect was significant before the switch ($F_{1,31} = 8.15$,
469 p = .008), but not after ($F_{1,31} = .178$, p>.67). None of the other interactions were significant
470 (p>.2).
471
472 During mixed blocks we observed higher accuracies and reaction times during the period
473 before the switch and the opposite pattern (low accuracies and faster responses) after it,
474 which could indicate a trade-off in the data. To address this possibility, we additionally
475 performed Pearson correlations between mean accuracy and reaction times during both
476 periods of the miniblock. Results show no association between the two measures, neither
477 before (r = .07; p > .35) or after (r = .17, p > .17) the switch.
478
479 **3.2. fMRI**
480 3.2.1. Univariate

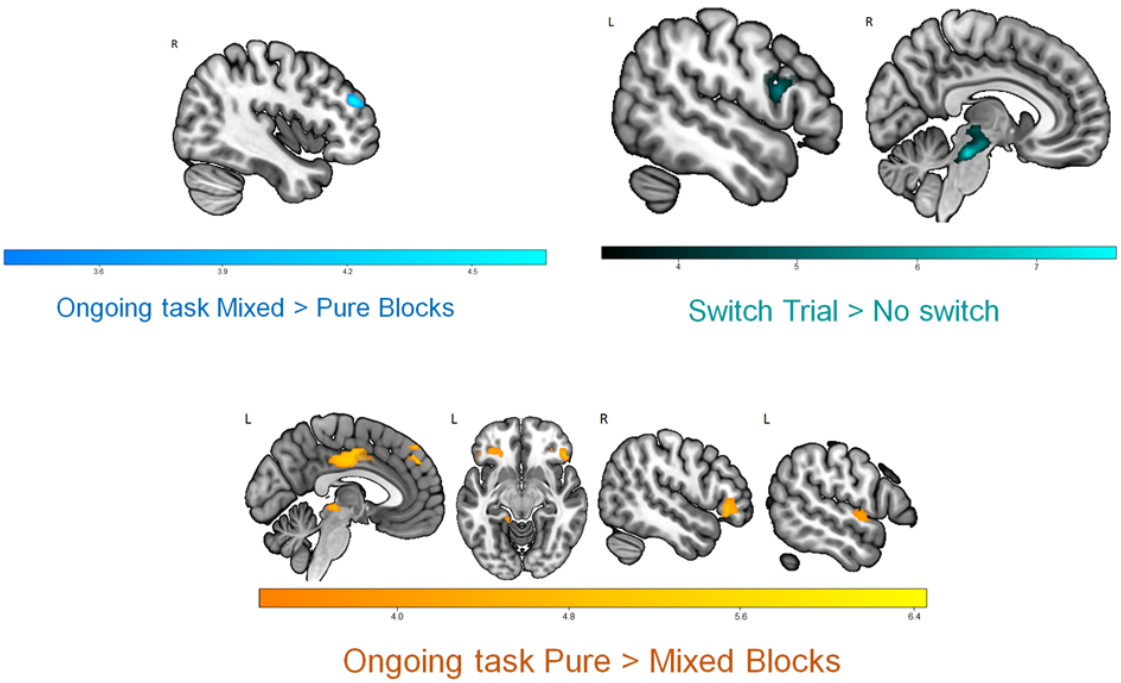
481 *3.2.1.1. Pure vs. Mixed blocks*

482 Before the switch, the right middle frontal gyrus (k = 56; MNI coordinates of peak voxel:
483 42, 44, 23) showed higher activation when participants had to maintain two tasks vs. one
484 (Mixed > Pure blocks). Conversely, in this scenario, we observed decreased activation
485 (Pure > Mixed blocks) in a set of regions. These included the bilateral middle cingulate
486 cortex (k = 251; -3, -10, 33), bilateral medial prefrontal cortex (mPFC; k = 80; -6, 44, 41),
487 left orbitofrontal cortex (OFC; k = 117; -33, 32, -16), right inferior frontal gyrus
488 (IFG)/OFC (k = 168; 54, 32, -1), left lingual and parahippocampal gyri (k = 148; -9, 52,
489 5) and left superior temporal gyrus (k = 62; -57, -1, -4).

490

491 *3.2.1.2. Switch vs. non-switch trials*

492 Transient activity during task switching was observed in the bilateral brainstem and
493 thalamus (k = 291; -3, -28, -22) as well as in a cluster including the left inferior/middle
494 frontal gyrus (IFG/MFG) and precentral gyrus (k = 160; -51, 11, 17).



495

Figure 5. Univariate results. Effect of task demands (one vs. two) and task switching. Scales reflect peaks of significant t-values ($p<.05$, FWE-corrected for multiple comparisons).

3.2.1.3. Interference effects

We assessed whether the interference effects observed on behaviour were matched at a neural level, by comparing incompatible vs. compatible trials before and after the switch. For this model, we included in each run one regressor for the instruction of each miniblock, one regressor corresponding to all Pure Miniblocks, four regressors corresponding to Compatible/Incompatible trials in Mixed Miniblocks, with separate regressors for the pre-switch and post-switch periods (onset at face presentation) and a last regressor for errors. Instructions and errors were modelled as described in section 2.5.1 Univariate analysis. Pure Miniblocks were modelled as a boxcar function with the duration of the entire pre/post switch period, while the four remaining regressors (Compatible Pre, Compatible Post, Incompatible Pre, Incompatible Post) were modelled as events with zero duration. This provided a model with a total of 7 regressors per run.

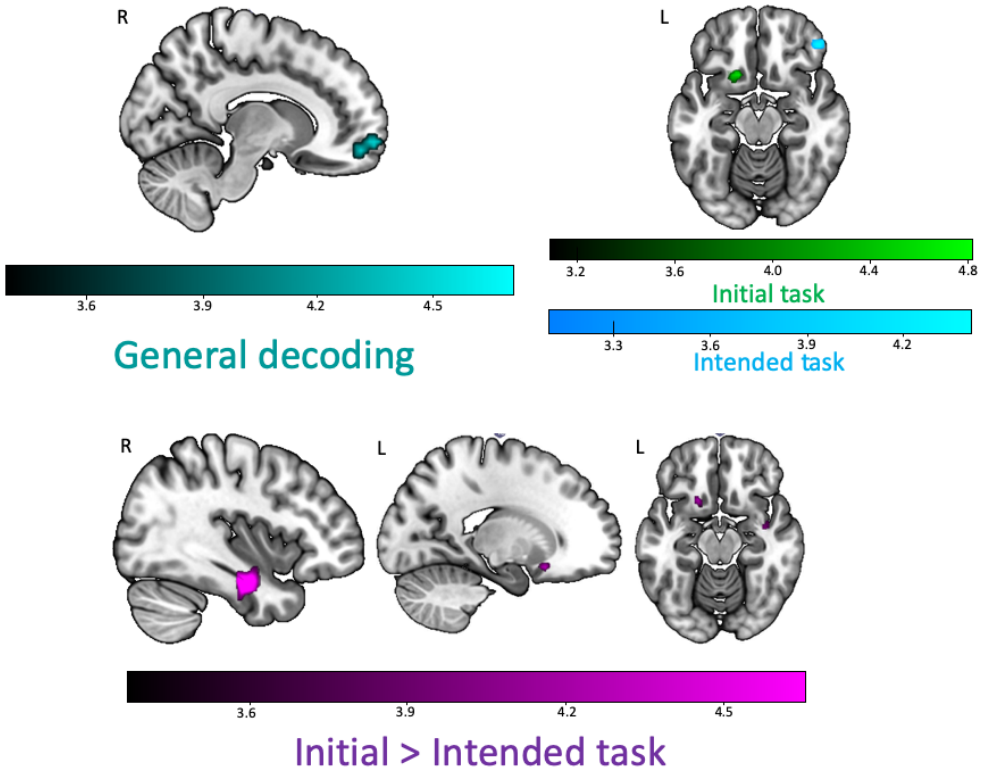
At the group level, t-tests were carried out for comparisons related to the effect of Interference before and after the switch. No effect survived multiple comparisons. However, at a more lenient threshold ($p<.001$, uncorrected), data showed higher activation in the left IFG ($k = 22$; -48, 35, 20) for incompatible > compatible trials. The opposite comparison (compatible > incompatible before the switch) or incompatible vs. compatible contrasts after the switch did not yield any significant results.

3.2.2. Decoding results

First, we averaged all individual classification maps to examine the regions sensitive to any kind of task (initial or intended) during the period prior to the switch. We found that

522 the rostromedial PFC/orbitofrontal cortex (OFC) presented significant accuracies above
523 chance (k = 68; 15, 56, -10). To further examine whether one of the tasks dominated the
524 classification, we extracted the decoding values from the initial and intended
525 classification from the general decoding ROI above. A paired t-test between the initial
526 and intended task decoding values showed no differences between decoding accuracy for
527 the two types of information, t_{31} = .846, p = .404. To further test this idea, we ran an ROI
528 analysis in this region to see whether decoding accuracy was significant for both the initial
529 and intended task. To avoid non-independency, we employed a Leave-One-Subject-Out
530 (LOSO) cross-validation approach (Esterman & Yantis, 2010) to select the ROI per
531 participant. That is, data from each participant was extracted from an ROI that was
532 defined based on the data from all the other participants, to avoid 'double dipping'
533 (Kriegeskorte, Simmons, Bellgowan, & Baker, 2009). After this, both the initial and
534 intended decoding values showed significant accuracy above chance (1-tailed, t_{31} = 3.018,
535 p=.0025 and t_{31} = 2.299, p = .0145, respectively). This suggests that the mPFC/OFC
536 region, prior to the switch, carries information about the relevant tasks to perform,
537 regardless of whether they are initial or intended.
538
539 To further characterise the information represented in the mPFC/OFC region, we
540 additionally examined whether we could cross-classify between the initial and intended
541 tasks, which speaks to the potential overlap between these representations. Therefore, we
542 performed ROI cross-classification analysis (Kaplan, Man, & Greening, 2015) in the
543 mPFC/OFC region from the general decoding. We followed the same classification
544 procedure as described in the methods section, but training the classifier on the initial task
545 and testing on the intended task, and vice versa. Note that in this scenario the cross-
546 classification was carried out in a totally independent and orthogonal analysis to that
547 employed to define the OFC region. However, results showed no significant cross-

548 classification in this region, $t_{31} = 1$, $p > .3$, which suggests that the initial vs. intended
549 nature of the tasks may change their representational format.
550
551 Moreover, we examined the initial and intended individual classification maps separately
552 to examine the regions sensitive to each type of task (initial or intended). With this, the
553 classification of the initial task alone showed a different cluster in the left OFC (k = 52, -
554 15, 17, -13). Conversely, we observed the right IFG (k = 53, 45, 41, -16) for the
555 classification of the intended task alone. Last, when comparing decoding accuracies
556 between the initial and intended tasks (subtracting initial – intended accuracy maps), we
557 observed significantly higher accuracies for the representation of the initial (vs. intended)
558 in the right FG and the hippocampus (k = 148; 39, -13, -25) and also in left OFC (k = 44;
559 -18, 17, -16). However, the opposite contrast (intended vs. initial) did not show any cluster
560 with significant differences.
561
562 Importantly, these MVPA results were unlikely to be due to differences in response times
563 (see Supplementary Materials *"RT analyses between miniblock pairs and correlation*
564 *with decoding accuracies"*).



565

566 **Figure 6.** Multivariate results during the period prior to the switch. Top left: General
567 decoding (cyan) of the region sensitive to any kind of task (initial or intended). Top right:
568 Decoding of the initial (green) and intended (blue) task separately. Bottom: Decoding of
569 initial > intended task (violet). Scales reflect peaks of significant t-values ($p<.05$, FWE-
570 corrected for multiple comparisons).

571

572 Last, although the main focus of this work is the period before the switch, we performed
573 exploratory classification analysis in the period after the switch to obtain further
574 understanding about how information is represented once the initial task is no longer
575 active. Following the same procedure as in section *2.5.2 Multivariate analysis*, we carried
576 out the decoding analysis with regressors of interest corresponding to the period after the
577 switch. As in the previous analyses, we first averaged all individual classification maps
578 to examine the regions sensitive to any kind of task during the period after the switch.
579 Here, we observed a cluster in left middle frontal/precental gyrus (k = 61; -36, 11, 35)

with significant accuracies above chance. Next, we averaged the classification maps separately for the currently-active post-switch task and the previously-performed initial task, to examine the regions sensitive to each type of task. The initial task could be decoded from bilateral IFG (left; k = 48; -54, 29, 26/right; k = 45; 54, 32, 5). Conversely, decoding of the currently-active post-switch task could only be observed at a more liberal threshold ($p < .01$, k = 100) in a cluster including the FG/cerebellum (k = 109; 15, -52, -19) and in left middle/inferior frontal gyrus (k = 122; -45, 8, 35).

4. Discussion

The present work aimed to examine the representation of A) currently-active, initial task sets, and B) intended task sets applied to social stimuli during a dual-sequential social categorization of faces. We found some regions from which we could decode only the currently-active or the intended task set, along with a region of vmPFC/OFC containing both types of information although cross-classification between the two was not possible.

The paradigm employed revealed behavioural costs due to the maintenance of intended task sets. Here, mixed blocks presented slower responses and lower accuracies, compared to pure ones, in line with previous studies (Los, 1996; Marí-Beffa et al., 2012). Response times were slower before than after the switch, when the intended task needed to be held while performing the initial one, in line with Smith (2003). These results reflect the higher demands associated with the maintenance of two relevant task sets. In addition, people need more time to respond when they hold the intention while performing the initial task, but not after, when they only need to focus on a single task. Further, we aimed to examine the activation of both initial and intended task sets before the switch by looking at possible interference between them. Behavioural results showed incompatibility effects in performance when participants needed to hold the intended task set while performing the

606 ongoing task. These results point to the maintenance of intended task settings, showing
607 that information about the pending set is maintained online during performance of the
608 ongoing task.
609
610 Turning to brain activation data, the behavioural costs we observed during the
611 maintenance of two task sets during mixed blocks were accompanied by increased
612 activation in the right MFG, while regions such as the left IFG and the thalamus increased
613 their activation during switching trials. Incompatible trials during this period were also
614 related to activation in left lPFC, but only at an uncorrected statistical threshold. These
615 results fit with previous work that associated lPFC with sustained control during dual
616 tasks (Braver et al., 2003; Szameitat et al., 2002) and the maintenance of delayed
617 intentions (Gilbert, 2011). Lateral PFC has also an important role in rule representation
618 and the selection of correct rules (Brass & Cramon, 2002; Crone, Wendelken, Donohue,
619 & Bunge, 2006), while the thalamus has also shown a role in cognitive flexibility during
620 task switching (Rikhye, Gilra, & Halassa, 2018). Overall, this pattern reflects the
621 cognitive demands of holding two tasks in mind, extending the results to the maintenance
622 of social categorization sets.
623
624 In addition, we observed a set of task-negative regions such as mPFC, middle cingulate,
625 OFC, IFG and temporal cortex which showed reduced signal when participants needed to
626 hold an intended task set. Similarly, Landsiedel & Gilbert (2015) carried out an intention-
627 offloading paradigm where participants had to remember a delayed intention, which they
628 had to fulfil after a brief filled delay. During the maintenance of such intention they found
629 a set of deactivated regions including mPFC, posterior cingulate cortex, infero-temporal
630 cortex, and temporo-parietal cortex. Interestingly, these authors showed that when
631 participants had the opportunity to offload intentions by setting an external reminder, this

632 decrease of activation was ameliorated. This reduction suggests that these task-negative
633 areas play a role in the representation of the delayed tasks. Pure blocks in our tasks are
634 similar to the offloading condition in Landsiedel & Gilbert (2015), as participants needed
635 to sustain only the initial task set to perform the correct social categorization. Therefore,
636 taking these results into account, it seems unlikely that task-negative regions reflect
637 simply a "default mode" (Fox et al., 2005), but rather that this deactivation is, to some
638 extent, playing a functional role during task performance (Spreng, 2012).
639
640 Currently, multivariate pattern analyses are one of the most powerful approaches to study
641 the information contained in different brain areas. In this work, we employed MVPA to
642 study the regions representing initial and intended task sets. Importantly, a novelty of our
643 approach was to do so by employing three different tasks instead of just two, unlike most
644 previous studies (e.g. Haynes et al., 2007; Momennejad & Haynes, 2013). We were able
645 to distinguish different regions representing initial and intended task sets. The currently-
646 active initial task was decoded from left OFC, while information about the intended one
647 was contained in the right IFG. Also, comparing between these two, a nearby OFC region
648 together with the FG showed higher fidelity of the representation for the initial vs.
649 intended task set. The left OFC region found for the initial task decoding has been
650 previously associated with representations of facial information related to social
651 categories (Freeman, Rule, Adams, & Ambady, 2010) and it also facilitates object
652 recognition in lower level areas such as the FG (Bar et al., 2006). In addition, the FG
653 showed greater fidelity of the representation of initial vs. intended task sets, which could
654 indicate the allocation of attentional resources to process current relevant information in
655 earlier perceptual regions. Besides having a prominent role in processing faces (Haxby et
656 al., 2000), previous studies have been able to decode task-relevant information related to
657 social categories (e.g. female vs. male, black vs. white) in perceptual regions such as the

FG (Gilbert et al., 2012; Kaul et al., 2014, 2011; Stolier & Freeman, 2017). On top of this, previous work using facial stimuli in the field of social perception and prejudice has shown how the FG is influenced by stereotypical associations and evaluations of social categories associated with the OFC (Gilbert et al., 2012; Stolier & Freeman, 2016). Our data fit and extend this work, suggesting that both OFC and FG contain information about the appropriate task set to perform the initial classification, pointing out the role of both high and lower-level perceptual regions in the representation of ongoing task sets performed on facial stimuli.

Conversely, decoding of the intended task was possible from the IFG. Previous work has related lPFC with the representation of intentions (Haynes et al., 2007; Momennejad & Haynes, 2012; Momennejad & Haynes, 2013; Soon et al., 2008) and it has also been linked to task-set preparation (e.g. González-García, Arco, Palenciano, Ramírez, & Ruz, 2017; González-García, Mas, de Diego-Balaguer & Ruz, 2016; Sakai & Passingham, 2003). This result, however, contradicts the proposal of Gilbert (2011) in which lPFC would serve as a general store for delayed intentions without information about their content. Nonetheless, while Gilbert (2011) decoded simpler stimulus-response mappings, in our study we classified abstract task-set information, as previous studies that also decoded intention from lPFC did (e.g. Haynes et al., 2007; Momennejad & Haynes, 2013). Therefore, our work agrees with Momennejad & Haynes (2013), suggesting that lPFC may represent delayed intentions when their content is abstract enough.

Interestingly, we found a region in vmPFC/OFC that contained information about both initial and intended task sets. We examined if this vmPFC/OFC region contained overlapping representations of both initial and intended task sets by performing a cross-classification analysis, and observed a lack of generalization from the two sets of

684 representations. Our cluster is close to the mPFC region found by previous intention
685 studies (e.g. Gilbert, 2011), but located in a more ventral area. Notably, in our paradigm
686 both the initial and intended tasks that participant performed were equally relevant and
687 were based on the same set of stimuli. Altogether, this suggests that mPFC may be
688 recruited when relevant task sets need to be held for a period of time, irrespective if this
689 information is being using at the moment or later. This result also fits with a recent
690 proposal (Schuck, Cai, Wilson, & Niv, 2016) characterising this region as a repositoiry
691 for cognitive maps of task states. Specifically, the vmPFC/OFC would represent task-
692 relevant states that are hard to discriminate based on sensory information alone (Schuck
693 et al., 2016; Wilson, Takahashi, Schoenbaum, & Niv, 2014), similar to our study. Overall,
694 this suggests a role for this region in the maintenance of task-relevant information,
695 regardless of when it is needed. Given that cross-classification between initial and
696 intended task sets was not possible in this region, this would be compatible with
697 multiplexed representations of the two types of information (i.e. representations using
698 orthogonal representational codes). It could also be compatible with nonoverlapping
699 populations of cells representing initial and intended tasks within the voxels in this region.
700 Further, it could be argued that the lack of cross-classification between initial and
701 intended tasks could be explained by differential processes underlying the representation
702 of these tasks. That is, the initial one could be represented simply by the activation of the
703 S-R associations, while the intended one would rely on the maintenance of a general task
704 set. Nonetheless, an explanation relying solely on this distinction seems unlikely, since
705 we observed interference effects. Even though the interference effect for accuracy was
706 smaller before the switch, for reaction times we observed interference only during this
707 period. This indicates that the S-R association for both tasks is active prior the switch,
708 although there could be some distinctions in the way these response representations are

maintained. However, the lack of significant cross-classification could also reflect simply a lack of statistical power.

Despite our decoding results, we did not find the dissociation pattern that we predicted for initial and intended task sets. Although we could decode intended task sets from mPFC, consistent with previous studies (Gilbert, 2011; Haynes et al., 2007; Momennejad & Haynes, 2012; Momennejad & Haynes, 2013), we did not find a frontoparietal representation of initial task sets as previous studies have (Qiao et al., 2017; Waskom et al., 2014; Woolgar et al., 2011b). This discrepancy could be explained by differences in the stimuli and task employed. Previous studies have decoded task information in the form of classification between different stimulus-responses mappings or different types of stimuli. In contrast, in our task we employed the same stimuli for all tasks, and the differences between the information to decode relied in the representation of the social category needed for the specific part of the miniblock, rather than perceptual features (e.g. the colour or shape of target stimuli), which may be easier to decode (Bhandari, Gagne, & Badre, 2018).

To conclude, in the present work we examined the representation of A) currently-active initial task sets, and B) intended task sets during a social categorization dual-sequential task. Crucially, we directly examined the common and differential representation of initial and delayed task sets, extending previous work studying these mechanisms separately. Moreover, we employed faces as target stimuli to complement prior research. Apart from replicating previous findings in dual tasks with social stimuli, we show how task set information was contained in different regions, depending on when it was needed. Thus, currently-active initial tasks were represented in specialized regions related to face processing and social categorization such as the OFC and FG, while intended ones were

735 represented in lPFC. On top of that, we showed a common brain region in vmPFC/OFC
736 maintaining a general representation of task-relevant information, irrespective of when
737 the task is performed, albeit it is not clear whether overlapping patterns of activation
738 represent both types of information. Moreover, the results from the classification after the
739 switch suggest that the representation of the two relevant tasks varies once the switch
740 takes place and the initial task is no longer active. Future research should. further
741 characterize the representational format of relevant task information depending on when
742 it is needed and examine the structure of the task set representation within these regions,
743 for instance using Representational Similarity Analysis (RSA, Kriegeskorte, Mur, &
744 Bandettini, 2008). Also, studying the interaction of sustained task sets with specific
745 conditions of each trial would extend our knowledge of the representational dynamics of
746 current and intended task-relevant information. Last, due to the social significance of
747 faces, one step forward would be to examine how their representation may vary in social
748 scenarios, contexts in which faces and other social stimuli are particularly relevant to
749 guide behaviour (Díaz-Gutiérrez, Alguacil, & Ruz, 2017).
750

751 **Funding**

752 This work was supported by the Spanish Ministry of Science and Innovation (PSI2016-
753 78236-P to M.R.) and the Spanish Ministry of Education, Culture and Sports
754 (FPU2014/04272 to P.D.G.).
755

756 **Acknowledgments**

757 This research is part of P.D.G's activities for the Psychology Graduate Program of the
758 University of Granada.
759
760

761 **References**

762 Abdulrahman, H., & Henson, R. N. (2016). Effect of trial-to-trial variability on optimal
763 event-related fMRI design: Implications for Beta-series correlation and multi-voxel
764 pattern analysis. *NeuroImage*, *125*, 756–766.
765 https://doi.org/10.1016/j.neuroimage.2015.11.009
766 Bar, A. M., Kassam, K. S., Ghuman, A. S., Boshyan, J., Schmid, A. M., Dale, A. M.,
767 Hämäläinen, M.S., Marinkovic, K., Schacter, D.L., Rosen, B.R., & Halgren, E.
768 (2006). Top-down facilitation of visual recognition. *Proceedings of the National*
769 *Academy of Sciences*, *103*(2), 449–454. https://doi.org/10.1073/pnas.0507062103
770 Bhandari, A., Gagne, C., & Badre, D. (2018). Just above Chance: Is It Harder to Decode
771 Information from Prefrontal Cortex Hemodynamic Activity Patterns? *Journal of*
772 *Cognitive Neuroscience*, *30*(10), 1473–1498. https://doi.org/10.1162/jocn
773 Bode, S., Haynes, J. D. (2009). Decoding sequential stages of task preparation in the
774 human brain. *Neuroimage, 45*(2), 606–613.
775 http://dx.doi.org/10.1016/j.neuroimage.2008.11.031.
776 Brass, M., & Cramon, D. Y. Von. (2002). The Role of the Frontal Cortex in Task
777 Preparation. *Cerebral Cortex*, *12*(9), 908–914.
778 https://doi.org/10.1093/cercor/12.9.908
779 Braver, T. S., Reynolds, J. R., & Donaldson, D. I. (2003). Neural Mechanisms of
780 Transient and Sustained Cognitive Control during Task Switching. *Neuron*, *39*(4),
781 713–726. https://doi.org/10.1016/S0896-6273(03)00466-5
782 Buckner, R. L., Andrews-Hanna, J. R., & Schacter, D. L. (2008). The Brain's Default
783 Network. *Annals of the New York Academy of Sciences*, *1124*(1), 1–38.
784 https://doi.org/10.1196/annals.1440.011
785 Buckner, R. L., & Carroll, D. C. (2007). Self-projection and the brain. *Trends in*
786 *Cognitive Sciences*, *11*(2), 49–57. https://doi.org/10.1016/j.tics.2006.11.004

787 Cousineau, D. (2005). Confidence intervals in within-subject designs: A simpler
788 solution to Loftus and Masson's method. *Tutorials in Quantitative Methods for*
789 *Psychology*, *1(1)*, 42-45. https://doi.org/10.20982/tqmp.01.1.p042
790 Crittenden, B. M., Mitchell, D. J., & Duncan, J. (2015). Recruitment of the default
791 mode network during a demanding act of executive control. *ELife*, *4:e06481*.
792 https://doi.org/10.7554/eLife.06481
793 Crone, E. A., Wendelken, C., Donohue, S. E., & Bunge, S. A. (2006). Neural evidence
794 for dissociable components of task-switching. *Cerebral Cortex*, *16*(4), 475–486.
795 https://doi.org/10.1093/cercor/bhi127
796 Díaz-Gutiérrez, P., Alguacil, S., & Ruz, M. (2017). Bias and control in social decision-
797 making. In A. Ibáñez, L. Sedeño, & A. Gacría (Eds.), *Neuroscience and Social*
798 *Science: The Missing Link* (pp. 47–68). Cham, Switzerland: Springer.
799 https://doi.org/10.1007/978-3-319-68421-5
800 Dumontheil, I., Thompson, R., & Duncan, J. (2011). Assembly and Use of New Task
801 Rules in Fronto-parietal Cortex. *Journal of Cognitive Neuroscience*, *23*(1), 168–
802 182. https://doi.org/10.1162/jocn.2010.21439
803 Duncan, J. (2010). The multiple-demand (MD) system of the primate brain: mental
804 programs for intelligent behaviour. *Trends in Cognitive Sciences*, *14*(4), 172–179.
805 https://doi.org/10.1016/j.tics.2010.01.004
806 Elton, A., & Gao, W. (2015). Task-positive Functional Connectivity of the Default
807 Mode Network Transcends Task Domain. *Journal of Cognitive Neuroscience*,
808 *27*(12), 2369–2381. https://doi.org/10.1162/jocn
809 Esterman, M., & Yantis, S. (2010). Perceptual expectation evokes category-selective
810 cortical activity. *Cerebral Cortex*, *20*(5), 1245–1253.
811 https://doi.org/10.1093/cercor/bhp188
812 Fox, M. D., Snyder, A. Z., Vincent, J. L., Corbetta, M., Essen, D. C. Van, & Raichle,

M. E. (2005). The human brain is intrinsically organized into dynamic, anticorrelated functional networks. *Proceedings of the National Academy of Sciences*, *102*(27), 9673–9678. https://doi.org/10.1073/pnas.0504136102

Freeman, J. B., Ma, Y., Barth, M., Young, S. G., Han, S., & Ambady, N. (2015). The neural basis of contextual influences on face categorization. *Cerebral Cortex*, *25*(2), 415–422. https://doi.org/10.1093/cercor/bht238

Freeman, J. B., Rule, N. O., Adams, R. B., & Ambady, N. (2010). The neural basis of categorical face perception: Graded representations of face gender in fusiform and orbitofrontal cortices. *Cerebral Cortex*, *20*(6), 1314–1322. https://doi.org/10.1093/cercor/bhp195

Gilbert, S. J. (2011). Decoding the Content of Delayed Intentions. *Journal of Neuroscience*, *31*(8), 2888–2894. https://doi.org/10.1523/JNEUROSCI.5336-10.2011

Gilbert, S. J., & Fung, H. (2018). Decoding intentions of self and others from fMRI activity patterns. *NeuroImage*, *172*, 278–290. https://doi.org/10.1016/j.neuroimage.2017.12.090

Gilbert, S. J., Swencionis, J. K., & Amodio, D. M. (2012). Evaluative vs. trait representation in intergroup social judgments: Distinct roles of anterior temporal lobe and prefrontal cortex. *Neuropsychologia*, *50*(14), 3600–3611. https://doi.org/10.1016/j.neuropsychologia.2012.09.002

González-García, C., Arco, J. E., Palenciano, A. F., Ramírez, J., & Ruz, M. (2017). Encoding, preparation and implementation of novel complex verbal instructions. *NeuroImage*, *148*, 264–273. https://doi.org/10.1016/j.neuroimage.2017.01.037

González-García, C., Mas-Herrero, E., de Diego-Balaguer, R., & Ruz, M. (2016). Task-specific preparatory neural activations in low-interference contexts. *Brain Structure and Function*, *221*(8), 3997–4006. https://doi.org/10.1007/s00429-015-

1141-5

Hampson, M., Driesen, N. R., Skudlarski, P., Gore, J. C., & Constable, R. T. (2006). Brain Connectivity Related to Working Memory Performance. *Journal of Neuroscience*, *26*(51), 13338–13343. https://doi.org/10.1523/JNEUROSCI.3408-06.2006

Haxby, J. V., Hoffman, E. A., & Gobbini, M. I. (2000). The Distributed Human Neural System for Face Perception. *Trends in Cognitive Sciences, 4*(6), 223–233. https://doi.org/10.1016/S1364-6613(00)01482-0

Haxby, J. V., Connolly, A.C. & Guntupalli, J.S. (2014). Decoding Neural Representational Spaces Using Multivariate Pattern Analysis. *Annual Review of Neuroscience, 35*, 435–456. https://doi.org/ 10.1146/annurev-neuro-062012-170325

Haynes, J. D. (2015). A Primer on Pattern-Based Approaches to fMRI: Principles, Pitfalls, and Perspectives. *Neuron, 87*, 257-270. https://dx.doi.org/10.1016/j.neuron.2015.05.025.

Haynes, J. D., Sakai, K., Rees, G., Gilbert, S., Frith, C., & Passingham, R. E. (2007). Reading Hidden Intentions in the Human Brain. *Current Biology*, *17*(4), 323–328. https://doi.org/10.1016/j.cub.2006.11.072

Hebart, M. N., Görgen, K., & Haynes, J. D. (2015). The Decoding Toolbox (TDT): a versatile software package for multivariate analyses of functional imaging data. *Frontiers in Neuroinformatics*, *8*, 88. https://doi.org/10.3389/fninf.2014.00088

Kanwisher, N., & Yovel, G. (2006). The fusiform face area: A cortical region specialized for the perception of faces. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *361*(1476), 2109–2128. https://doi.org/10.1098/rstb.2006.1934

Kaplan, J. T., Man, K., & Greening, S. G. (2015). Multivariate cross-classification:

applying machine learning techniques to characterize abstraction in neural representations. *Frontiers in Human Neuroscience*, *9*:151. https://doi.org/10.3389/fnhum.2015.00151

Kaul, C., Ratner, K. G., & Van Bavel, J. J. (2014). Dynamic representations of race: Processing goals shape race decoding in the Fusiform gyri. *Social Cognitive and Affective Neuroscience*, *9*(3), 326–332. https://doi.org/10.1093/scan/nss138

Kaul, C., Rees, G., & Ishai, A. (2011). The Gender of Face Stimuli is Represented in Multiple Regions in the Human Brain. *Frontiers in Human Neuroscience*, 4:238. https://doi.org/10.3389/fnhum.2010.00238

Kliegel, M., McDaniel, M., & Einstein, G. (2008). *Prospective memory: cognitive, neuroscience, developmental, and applied perspectives.* Mahwah, NJ: Erlbaum.

Koski, J. E., Collins, J. A., Olson, I. R., & Hospital, G. (2017). The neural representation of social status in the extended face- processing network. *European Journal of Neuroscience*, *46*(12), 2795–2806. https://doi.org/10.1111/ejn.13770.The

Kriegeskorte, N., Goebel, R., & Bandettini, P. (2006). Information-based functional brain mapping. *Proceedings of the National Academy of Sciences*, *103*, 3863–3868. https://doi.org/10.1073/pnas.0600244103

Kriegeskorte, N., Mur, M., & Bandettini, P. (2008). Representational similarity analysis - connecting the branches of systems neuroscience. *Frontiers in Systems Neuroscience*, *2: 4*. https://doi.org/10.3389/neuro.06.004.2008

Kriegeskorte, N., Simmons, W. K., Bellgowan, P. S., & Baker, C. I. (2009). Circular analysis in systems neuroscience: The dangers of double dipping. *Nature Neuroscience*, *12*(5), 535–540. https://doi.org/10.1038/nn.2303

Landsiedel, J., & Gilbert, S. J. (2015). Creating external reminders for delayed intentions: Dissociable influence on “task-positive” and “task-negative” brain

networks. *NeuroImage*, *104*, 231–240. https://doi.org/10.1016/j.neuroimage.2014.10.021

Loose, L. S., Wisniewski, D., Rusconi, M., Goschke, T., & Haynes, J.-D. (2017). Switch-Independent Task Representations in Frontal and Parietal Cortex. *The Journal of Neuroscience*, *37*(33), 8033–8042. https://doi.org/10.1523/jneurosci.3656-16.2017

Los, S. A. (1996). On the origin of mixing costs: Exploring information processing in pure and mixed blocks of trials. *Acta Psychologica*, *94*(2), 145–188. https://doi.org/10.1016/0001-6918(95)00050-X

Marí-Beffa, P., Cooper, S., & Houghton, G. (2012). Unmixing the mixing cost: Contributions from dimensional relevance and stimulus-response suppression. *Journal of Experimental Psychology: Human Perception and Performance*, *38*(2), 478–488. https://doi.org/10.1037/a0025979

Mars, R. B., Neubert, F. X., Noonan, M. P., Sallet, J., Toni, I., & Rushworth, M. F. (2012). On the relationship between the "default mode network" and the "social brain." *Frontiers in Human Neuroscience*, *6*, 1–9. https://doi.org/10.3389/fnhum.2012.00189

Meyer, M. L., Davachi, L., Ochsner, K. N., & Lieberman, M. D. (2018). Evidence That Default Network Connectivity During Rest Consolidates Social Information. *Cerebral Cortex*, 29(5), 1–11. https://doi.org/10.1093/cercor/bhy071

Mitchell, J. P., Macrae, C. N., & Banaji, M. R. (2006). Dissociable Medial Prefrontal Contributions to Judgments of Similar and Dissimilar Others. *Neuron*, *50*(4), 655–663. https://doi.org/10.1016/j.neuron.2006.03.040

Momennejad, I., & Haynes, J. D. (2012). Human anterior prefrontal cortex encodes the "what" and "when" of future intentions. *NeuroImage*, *61*(1), 139–148. https://doi.org/10.1016/j.neuroimage.2012.02.079

917 Momennejad, I., & Haynes, J. D. (2013). Encoding of Prospective Tasks in the Human
918 Prefrontal Cortex under Varying Task Loads. *Journal of Neuroscience*, *33*(44),
919 17342–17349. https://doi.org/10.1523/JNEUROSCI.0492-13.2013
920 Monsell, S. (2003). Task Switching. *Trends in Cognitive Sciences*, *7*(3), 134–140.
921 https://doi.org/doi:10.1016/S1364-6613(03)00028-7
922 Moyes, J., Sari-Sarraf, N., & Gilbert, S. J. (2019). Characterising monitoring processes
923 in event-based prospective memory: Evidence from pupillometry. *Cognition*, *184*,
924 83–95. https://doi.org/10.1016/j.cognition.2018.12.007
925 Palenciano, A. F., González-García, C., Arco, J. E., & Ruz, M. (2019). Transient and
926 Sustained Control Mechanisms Supporting Novel Instructed Behavior. *Cerebral*
927 *Cortex*, 29, 3948–3960. https://doi.org/10.1093/cercor/bhy273
928 Palenciano, A. F., González-García, C., Arco, J. E., Pessoa, L. & Ruz, M. (2019).
929 Representational organization of novel task sets during proactive encoding.
930 *Journal of Neuroscience,39,* 0725-19. https://doi.org/10.1523/JNEUROSCI.0725-
931 19.2019
932 Pereira, F., Mitchell, T., & Botvinick, M. (2009). Machine learning classifiers and
933 fMRI: A tutorial overview. *NeuroImage*, *45*(1), S199–S209.
934 https://doi.org/10.1016/j.neuroimage.2008.11.007
935 Qiao, L., Zhang, L., Chen, A., & Egner, T. (2017). Dynamic Trial-by-Trial Recoding of
936 Task-Set Representations in the Frontoparietal Cortex Mediates Behavioral
937 Flexibility. *The Journal of Neuroscience*, *37*(45), 11037–11050.
938 https://doi.org/10.1523/jneurosci.0935-17.2017
939 Raichle, M. (2015). The Brain's Default Network. *Annals of the New York Academy of*
940 *Sciences*, *38*, 433–447. https://doi.org/10.1196/annals.1440.011
941 Rikhye, R. V., Gilra, A., & Halassa, M. M. (2018). Thalamic regulation of switching
942 between cortical representations enables cognitive flexibility. *Nature*

943 *Neuroscience*, *21*(12), 1753–1763. https://doi.org/10.1038/s41593-018-0269-z

944 Rogers, R. D., & Monsell, S. (1995). Costs of a predictible switch between simple

945 cognitive tasks. *Journal of Experimental Psychology: General*, *124*(2), 207–231.

946 https://doi.org/http://dx.doi.org/10.1037/0096-3445.124.2.207

947 Sakai, K., & Passingham, R. E. (2003). Prefrontal interactions reflect future task

948 operations. *Nature Neuroscience*, *6*(1), 75–81. https://doi.org/10.1038/nn987

949 Schneider, W., Eschman, A., & Zuccolotto, A. (2002). E-Prime user's guide. Pittsburgh:

950 Psychology Software Tools Inc.

951 Schuck, N. W., Cai, M. B., Wilson, R. C., & Niv, Y. (2016). Human Orbitofrontal

952 Cortex Represents a Cognitive Map of State Space. *Neuron*, *91*(6), 1402–1412.

953 https://doi.org/10.1016/j.neuron.2016.08.019

954 Singer, T., Kiebel, S. J., Winston, J. S., Dolan, R. J., & Frith, C. D. (2004). Brain

955 Responses to the Acquired Moral Status of Faces. *Neuron*, *41*(4), 653–662.

956 https://doi.org/10.1016/S0896-6273(04)00014-5

957 Smith, R. E. (2003). The Cost of Remembering to Remember in Event-Based

958 Prospective Memory: Investigating the Capacity Demands of Delayed Intention

959 Performance. *Journal of Experimental Psychology: Learning Memory and*

960 *Cognition*, *29*(3), 347–361. https://doi.org/10.1037/0278-7393.29.3.347

961 Smith, V., Mitchell, D. J., & Duncan, J. (2018). Role of the Default Mode Network in

962 Cognitive Transitions, *Cerebral Cortex, 28*(10), 3685–3696.

963 https://doi.org/10.1093/cercor/bhy167

964 Soon, C. S., Brass, M., Heinze, H. J., & Haynes, J. D. (2008). Unconscious

965 determinants of free decisions in the human brain. *Nature Neuroscience*, *11*(5),

966 543–545. https://doi.org/10.1038/nn.2112

967 Spreng, N. R. (2012). The fallacy of a "task-negative" network. *Frontiers in*

968 *Psychology*, *3*, 1–5. https://doi.org/10.3389/fpsyg.2012.00145

969 Spreng, R. N., Mar, R. A., & Kim, A. S. N. (2008). The Common Neural Basis of
970 Autobiographical Memory, Prospection, Navigation, Theory of Mind, and the
971 Default Mode: A Quantitative Meta-analysis. *Journal of Cognitive Neuroscience*,
972 *21*(3), 489–510. https://doi.org/10.1162/jocn.2008.21029
973 Stolier, R. M., & Freeman, J. B. (2016). Neural pattern similarity reveals the inherent
974 intersection of social categories. *Nature Neuroscience*, *19*(6), 795–797.
975 https://doi.org/10.1038/nn.4296
976 Stolier, R. M., & Freeman, J. B. (2017). A Neural Mechanism of Social Categorization.
977 *The Journal of Neuroscience*, *37*(23), 5711–5721.
978 https://doi.org/10.1523/JNEUROSCI.3334-16.2017
979 Summerfield, C., Egner, T., Greene, M., Koechlin, E., Mangels, J., & Hirsch, J. (2006).
980 Predictive Codes for Forthcoming Perception in the Frontal Cortex. *Science*,
981 *314*(5803), 1311–1314. https://doi.org/10.1126/science.1132028
982 Szameitat, A. J., Schubert, T., Müller, K., & Von Yves Cramon, D. (2002). Localization
983 of executive functions in dual-task performance with fMRI. *Journal of Cognitive*
984 *Neuroscience*, *14*(8), 1184–1199. https://doi.org/10.1162/089892902760807195
985 Tottenham, N., Tanaka, J. W., Leon, A. C., McCarry, T., Nurse, M., Hare, T. A., &
986 Nelson, C. (2009). The NimStim set of facial expressions: Judgments from
987 untrained research participants. *Psychiatry Research*, *168*(3), 242–249.
988 https://doi.org/http://doi.org/10.1016/j.psychres.2008.05.006
989 Tschentscher, N., Mitchell, D., & Duncan, J. (2017). Fluid Intelligence Predicts Novel
990 Rule Implementation in a Distributed Frontoparietal Control Network. *The Journal*
991 *of Neuroscience*, *37*(18), 4841–4847. https://doi.org/10.1523/jneurosci.2478-
992 16.2017
993 Turner, B. (2010). Comparison of methods for the use of pattern classificaion on rapid
994 event-related fMRI data. In: Annual Meeting of the Society for Neuroscience, San

995 Diego, CA.

996 van der Wel, P. & van Steenbergen, H. (2018). Pupil dilation as an index of effort in
997 cognitive control tasks: A review. *Psychonomic Bulletin and Review, 25*, 2005–
998 2015. https://doi.org/10.3758/s13423-018-1432-y

999 Waskom, M. L., Kumaran, D., Gordon, A. M., Rissman, J., & Wagner, A. D. (2014).
1000 Frontoparietal Representations of Task Context Support the Flexible Control of
1001 Goal-Directed Cognition. *Journal of Neuroscience*, *34*(32), 10743–10755.
1002 https://doi.org/10.1523/JNEUROSCI.5282-13.2014

1003 Wilson, R. C., Takahashi, Y. K., Schoenbaum, G., & Niv, Y. (2014). Orbitofrontal
1004 cortex as a cognitive map of task space. *Neuron*, *81*(2), 267–279.
1005 https://doi.org/10.1016/j.neuron.2013.11.005

1006 Woolgar, A., Hampshire, A., Thompson, R., & Duncan, J. (2011a). Adaptive Coding of
1007 Task-Relevant Information in Human Frontoparietal Cortex. *Journal of*
1008 *Neuroscience*, *31*(41), 14592–14599. https://doi.org/10.1523/JNEUROSCI.2616-
1009 11.2011

1010 Woolgar, A., Thompson, R., Bor, D., & Duncan, J. (2011b). Multi-voxel coding of
1011 stimuli, rules, and responses in human frontoparietal cortex. *NeuroImage*, *56*(2),
1012 744–752. https://doi.org/10.1016/j.neuroimage.2010.04.035