Individual-subject functional localization increases univariate activation but not multivariate pattern discriminability in the 'multiple-demand' frontoparietal network

*Sneha Shashidhara[1]

*Floortje S. Spronkers[1]

Yaara Erez[1]

* Equal contribution

1 MRC Cognition and Brain Sciences Unit, 15 Chaucer Road, Cambridge CB2 7EF, UK

Abbreviated title: Localizing the frontoparietal network

Correspondence: Sneha Shashidhara, MRC Cognition and Brain Sciences Unit, 15 Chaucer

Road, Cambridge CB2 7EF, UK. Email: sneha.shashidhara@mrc-cbu.cam.ac.uk

Abstract

The frontoparietal 'multiple-demand' (MD) control network plays a key role in goal-directed behavior. Recent developments of multivoxel pattern analysis (MVPA) for fMRI data allows for more fine-grained investigations into the functionality and properties of brain systems. In particular, MVPA in the MD network was used to gain better understanding of control processes such as attentional effects, adaptive coding, and representation of multiple task-relevant features, but its use for this network has been limited by overall low decoding levels. A common practice of applying MVPA is by investigating pattern discriminability within a region-of-interest (ROI) using a template mask, thus ensuring that the same brain areas are studied in all participants. This approach offers high sensitivity, but does not take into account differences between individuals in the spatial organization of brain regions. An alternative approach uses independent localizer data for each subject to select the most responsive voxels and define individual ROIs within the boundaries of a group template. Such an approach allows for a refined and targeted localization based on the unique pattern of activity of individual subjects while ensuring that functionally similar brain regions are studied for all subjects. In the current study we tested whether using individual ROIs leads to changes in decodability of task-related neural representations as well as univariate activity across the MD network compared to when using a group template. We used three localizer tasks to separately define subject-specific ROIs: spatial working memory, verbal working memory, and a Stroop task. We then systematically assessed univariate and multivariate results in a separate rule-based criterion task. All the localizer tasks robustly recruited the MD network and evoked highly reliable activity patterns in individual subjects. Consistent with previous studies, we found a clear benefit of the subject-specific ROIs for univariate results from the criterion task, with increased activity in the individual ROIs based on the localizers' data, compared to the activity observed when using the group template. In contrast, there was no benefit of the subject-specific ROIs, as well as no cost, for multivariate results. Both univariate and multivariate results were similar in the individual ROIs defined by each of the three localizers. Our results provide an important empirical evidence for researchers in the field of cognitive control for the use of individual ROIs in the frontoparietal network

for both univariate and multivariate analysis of fMRI data, and serve as another step towards standardization and increased comparability across studies.

1. Introduction

Multiple studies have provided consistent evidence for the involvement of a large distributed network of frontal and parietal regions in cognitive control and flexible goal-directed behavior (Desimone & Duncan, 1995; J Duncan & Owen, 2000; John Duncan, 2006, 2010, 2013; E. Fedorenko, Duncan, & Kanwisher, 2013; Stiers, Mennes, & Sunaert, 2010). This network has been termed the 'multiple-demand' (MD) network (Duncan, 2006), and it closely resembles other networks that have been associated with control processes such as the cognitive control network (e.g. Cole and Schneider, 2007), task-activation ensemble (Seeley et al., 2007), and task-positive network (Fox et al., 2005). The MD network includes the intraparietal sulcus (IPS), the anterior-posterior axis of the middle frontal gyrus (MFG), the anterior insula and adjacent frontal operculum (AI/FO), the pre-supplementary motor area (pre-SMA) and the dorsal anterior cingulate (ACC) (John Duncan, 2010, 2013). A primary characteristic of this network is an increase in activity with increased demand, especially seen through functional magnetic resonance imaging (fMRI) blood oxygenation level dependent (BOLD) activity, across a variety of cognitive domains such as working memory, task switching, inhibition, math, and problem solving (Cole & Schneider, 2007; Dove, Pollmann, Schubert, Wiggins, & Yves von Cramon, 2000; E. Fedorenko et al., 2013; Shashidhara, Mitchell, Erez, & Duncan, 2019).

The development of multivoxel pattern analysis (MVPA) methods (Haxby et al., 2001; Haynes & Rees, 2006; N. Kriegeskorte, Goebel, & Bandettini, 2006) has further led to a variety of findings related to the representation of multiple aspects of cognitive control across the MD network. These include attentional effects, adaptive coding, and coding of target features and task rules (Erez & Duncan, 2015; Etzel, Cole, Zacks, Kay, & Braver, 2016; Nee & Brown, 2012; Nelissen, Stokes, Nobre, & Rushworth, 2013; Wisniewski, Goschke, & Haynes, 2016; Woolgar, Thompson, Bor, & Duncan, 2011). MVPA allows for a fine-grained investigation of distributed patterns of activity and the information that is conveyed in these patterns related to different experimental conditions and their respective cognitive constructs. However, its use in the frontoparietal MD network has been limited by overall low decoding levels, or discriminability between conditions, compared to other brain systems (Bhandari, Gagne, & Badre, 2018). This could be, at least in part, due to differences between individuals in respect

to the spatial organization of the network. Indeed, it has been demonstrated that across the fronto-parietal lobes, several close by regions have been fractionated into different networks in resting state studies, with different boundaries in individuals (Glasser et al., 2016; Schaefer et al., 2018; Yeo et al., 2011). Nevertheless, many MVPA studies of the MD network used a group template to define regions of interest (ROIs) for all subjects to investigate task-related representations. Such a group template is robust and easily comparable across studies. However, it provides high sensitivity at the cost of specificity to individual differences, as it might not accurately identify regions in individual subjects due to both anatomical and functional differences (Brett, Johnsrude, & Owen, 2002; E. Fedorenko et al., 2013; Nieto-Castañón & Fedorenko, 2012).

An alternative to the group template is using an independent functional localizer task to establish subject-specific ROIs. In this approach, participants perform a short task in addition to the main task in the scanning session, and the data from this task is used to localize regions-of-interest to be tested with data from the main task. This method is commonly used in vision research (Erez & Yovel, 2014; Lafer-Sousa, Conway, & Kanwisher, 2016; Reddy & Kanwisher, 2007; Weiner et al., 2018). For example, specific tasks are used to identify regions in individuals that are recruited for face processing (Berman et al., 2010; Kanwisher, McDermott, & Chun, 1997) and object processing (Malach et al., 1995). Task contrasts such as faces versus scrambled faces or objects versus scrambled objects are applied, and ROIs can be identified by the experimenter as clusters of activity in individual subjects. However, this alternative of manually defining ROIs by the experimenter using a functional localizer is subjective and therefore may be prone to biases, inaccuracies, and reduced reproducibility (Garrison et al., 2015; Krishnan, Slavin, Tran, Doraiswamy, & Petrella, 2006).

To overcome the limitations of both the group template and the individual manually-defined regions, a hybrid group-constrained subject-specific approach has been proposed for use in the language system (E. Fedorenko, Hsieh, Nieto-Castañón, Whitfield-Gabrieli, & Kanwisher, 2010; Evelina Fedorenko, Nieto-Castañon, & Kanwisher, 2012), and later expanded to the ventral pathway of the visual system (Julian, Fedorenko, Webster, & Kanwisher, 2012) and theory of mind regions (Paunov, Blank, & Fedorenko, 2019). In this approach, independent subject-specific localizer data are collected in addition to the main experimental task, then the thresholded contrast data from this task are masked

4

with a group template of regions and only the voxels that were responsive to the localizer task within the group template are used for further analysis in the main experimental task. The advantage of this approach is the use of a group template that ensures targeting of similar areas for all participants, as well as refining this localization by subject-specific activations within these areas. It therefore offers an objective experimenter-independent definition of subject-specific regions that does not require manual region definition. Importantly, it supports comparability across different studies because selected voxels in individual subjects are constrained to a group template. Using this group-constrained subject-specific ROIs approach has been shown to increase the detected univariate BOLD response associated with contrasts of interest in language-related areas compared to when a group template was used as ROIs (E. Fedorenko et al., 2010). More generally, the benefit of using individually defined ROIs for univariate results has been shown for the visual system (Saxe, Brett, & Kanwisher, 2006) and has been modelled and demonstrated using simulation data (Nieto-Castañón & Fedorenko, 2012).

The hybrid group-constrained subject-specific approach has been subsequently used for the MD network in studies that used both univariate (Blank & Fedorenko, 2017; Blank, Kanwisher, & Fedorenko, 2014; Mineroff, Blank, Mahowald, & Fedorenko, 2018; Paunov et al., 2019) and multivariate measures (Erez & Duncan, 2015; Shashidhara & Erez, 2019) related to control processes. A spatial working memory task that has been previously demonstrated to robustly recruit the MD network, was used as a localizer. In this localizer task, a highly demanding condition is contrasted with an easier version of the same task to identify the network in individual subjects, then constrained by an anatomical or group-average functional template to define the subject-specific ROIs. However, it remains an open question whether using these refined ROIs at the single subject level has any benefit for multivariate results in the MD network, similarly to the benefits that were previously reported for univariate results (E. Fedorenko et al., 2010; Nieto-Castañón & Fedorenko, 2012; Saxe et al., 2006). Using the MD group template for MVPA means that many voxels outside the individually-defined functional MD regions are included in the analysis, which may not express the domain-general characteristics of the MD network. In fact, they may be part of other nearby brain systems (Glasser et al., 2016; Schaefer et al., 2018; Yeo et al., 2011), and their inclusion in the multivariate analysis may potentially mask out pattern-based differences between the experimental conditions of interest. If this

5

is indeed the case, then the identification of the MD network in individual subjects has the potential to critically improve our ability to detect the neural signature of control processes as measured by multivariate methods, thus substantially increasing the benefit of using MVPA in cognitive control research. On the other hand, it has been previously demonstrated in the visual system and using simulations that even voxels outside the regions of increased univariate activity contribute to multivoxel discrimination (Haxby et al., 2001; N. Kriegeskorte et al., 2006). This implies that using more finely-defined subject-specific ROIs instead of the large and broad group template may potentially reduce discrimination levels. A related point concerns the size of ROIs. Increased decoding levels have been previously linked with increased size of ROI, at least in the visual system (Eger, Ashburner, Haynes, Dolan, & Rees, 2008; Said, Moore, Engell, Todorov, & Haxby, 2010; Walther, Caddigan, Fei-Fei, & Beck, 2009), implicating the importance of controlling for the ROI size for MVPA. Using the group-constrained subject-specific ROIs allows for such control by selecting a fixed number of voxels from each ROI with the largest localizer contrast values. Such control for ROI size enables the comparison of decoding levels between the different regions within the network, which vary in size, as well as with regions outside this network. This, however, should not come at the expense of reduced decodability, if indeed the use of smaller ROIs reduces pattern discriminability. Overall, existing data provide only limited evidence regarding the link between the use of subject-specific ROIs and multivoxel pattern measures in the MD network.

In the current study, we build on the previously reported findings and ask whether using functionally-defined subject-specific ROIs affects multivoxel pattern results in the MD network. Because the recruitment of the MD network at the group level is observed across a range of cognitive domains (E. Fedorenko et al., 2013), different tasks can be potentially used as localizers. We therefore also ask whether different localizers may have different effects on multivariate results. To address these questions, we use three localizer tasks and an independent rule-based criterion task. The localizer tasks are spatial working memory, verbal working memory, and a Stroop-like task, which have all been previously shown to consistently recruit MD regions (E. Fedorenko et al., 2013). We first assess the reliability and variability of the level of recruitment of the MD network by the localizers. We then assess the benefit of the subject-specific ROIs for univariate results in the independent criterion task, aiming

6

to replicate and generalize previous findings. Finally, we systematically test for the effect of using the subject-specific ROIs defined by each of the three localizers on multivariate results in the independent criterion task. We provide an important empirical evidence for the effect of using subject-specific ROIs on the ability to detect task-related representations in the MD network as reflected in distributed patterns of activity.

2. Methods

2.1. Participants

A total of 25 healthy participants (18 female, mean age 23.8 years) took part in the study. Three participants were excluded because of movements larger than 5 mm during at least one of the scanning runs, and one participant was excluded due to slice by slice variance larger than 300 after slice time correction in more than two runs. In addition, two participants were excluded due to technical problems with the task scripts. Lastly, one participant was excluded to maintain the balance of the order of localizers across participants. This participant was chosen randomly out of the participants who had the same order of localizers and prior to any data analysis beyond pre-processing. Overall, 18 participants were included in the analysis. All participants were right-handed and had normal or corrected-to-normal vision. Participants were either native English speakers or had learnt English at a young age and received their education in English. Participants gave written informed consent prior to participation and received a monetary reimbursement at the end of the experiment. Ethical approval was obtained from the Cambridge Psychology Research Ethics Committee.

2.2. Experimental paradigm

The study consisted of three localizer tasks and one rule-based similarity judgement task. The localizers were: a spatial working memory (WM) task, a verbal working memory task and a Stroop task, variations of which have previously been shown to recruit the MD network (E. Fedorenko et al., 2013). The rule-based task was used as a criterion task to test for univariate effects and rule decoding using MVPA, using both subject-specific ROIs based on activation data from the localizers and the group template. Participants practiced all tasks before the start of the scanning session. During scanning, participants performed two runs of each localizer followed by four runs of the rule-based task. The two

7

runs of the same localizer always followed each other, and the order of the three localizers was balanced across participants. The average total scanning session duration was 105 minutes.

2.3. Localizer tasks and criterion task

The spatial WM and verbal WM localizer tasks were adapted from Fedorenko et al. (2013) and the Stroop task was adapted from Hampshire et al. (2012). The localizers were chosen based on their consistent recruitment of the MD network as has been shown by Fedorenko et al. (2013). The localizers all followed a blocked design. Each run contained 10 blocks, alternating between Easy (5 blocks) and Hard (5 blocks) task conditions. There were no indications for the start or end of each block. The localizer tasks were designed to be used with a contrast of Hard versus Easy conditions, and in order to keep them as short as possible they did not include fixation blocks. The first run always started with an Easy block, and the second with a Hard block. All blocks lasted for 32 seconds, leading to a total run duration of 5 minutes and 20 seconds.

All tasks were coded and presented using Psychtoolbox (Brainard, 1997) for MatLab (The MathWorks, Inc.). Stimuli were projected on a 1920 x 1080 screen inside the scanner, and participants used a button box, with one finger from each hand to respond.

2.3.1. Localizer 1: spatial WM task

In the spatial WM task (**Figure 1A**), each trial started with an initial fixation dot (0.5 s), followed by a 3x4 grid with either one (Easy condition) or two (Hard condition) highlighted cells. The highlighted cells were displayed over four seconds, with different cells highlighted every one second, leading to an overall four (Easy condition) or eight (Hard condition) highlighted cells in each grid. In a subsequent two-forced choice display, two grids with highlighted cells were presented on the right and left sides of the screen. Participants pressed a button (left or right) to indicate which grid matched the previously highlighted cells. After a response window of 3.25 s, a feedback was presented for 0.25 s. The correct grid appeared an equal number of times on the right and left. Overall, each trial was 8 s long, and each task block contained four trials.

2.3.2. Localizer 2: verbal WM task

The verbal WM task (**Figure 1B**) followed a similar design to the spatial WM task. Following fixation (0.5 s), participants were presented with four consecutive screens containing one (Easy condition) or two (Hard condition) written digits. In a following two choice display, participants indicated the correct sequence of digits by pressing a button. The two answer options were displayed at the center of the screen, one above the other for ease of reading. The left button was used to choose the sequence on top, while the right button was used for the sequence at the bottom. The correct sequence appeared an equal number of times on the top and bottom. Following a response window of 3.25 s, participants were given feedback at the end of each trial (0.25 s). Each trial was 8 s long, and each task block contained four trials.

2.3.3. Localizer 3: Stroop task

The third localizer was a variation of the Stroop task (**Figure 1C**). On each trial, following a fixation dot (0.5 s), participants were presented with a test word, which was the name of a color, written in color at the top of the screen. In the Easy condition, the ink color was the same as the color name (congruent) (e.g. the word 'green' written in green ink), and in the Hard condition the ink color and the color name were different (incongruent) (e.g. the word 'red' written in green ink). Participants had to indicate the ink color (rather than the written color name) by choosing one of two answer options at the bottom of the screen, displayed at the same time. The answer options were color name words, and their ink color was different from its name. Participants had to choose the word, i.e. the written color name (regardless of the ink color), that matched the ink color of the test word at the top. Therefore, participants had to switch between attending to the ink color of the test word (ignoring the written color name) to detecting the matching written color name (and ignoring the ink color) out of the two answer options at the bottom. We used a total of six colors throughout the task. In the congruent condition, the ink color of the answer options was chosen randomly, excluding the color name (and ink color) of the test word (e.g. the options for the above congruent test word example could be the word 'blue' written in brown ink and 'green' written in purple ink, with the latter being the correct answer). In the incongruent trials, the ink color of one of the answer options matched the color name (and not the ink color) of the test

9

word. On half of the incongruent trials, it was the correct answer that had the same ink color as the test word color name (for example, if the test word is 'red' written in green, then a correct answer could be 'green' written in red). On the other half, it was the incorrect option with ink color the same as the test word color name (e.g. if the test word is 'red' written in green, then the incorrect option could be 'purple' written in red, see example in Figure 1C). This was done to further increase the conflict between stimuli and thus the difficulty level of the hard blocks while ensuring that the ink color of the test word cannot be used when choosing an answer. A total of six colors were used in this task (red, green, blue, orange, purple, and brown). Participants had 1.25 s to view the stimuli and respond, after which they received feedback for 0.25 s. Each trial was 2 s long, and blocks consisted of 16 trials each.
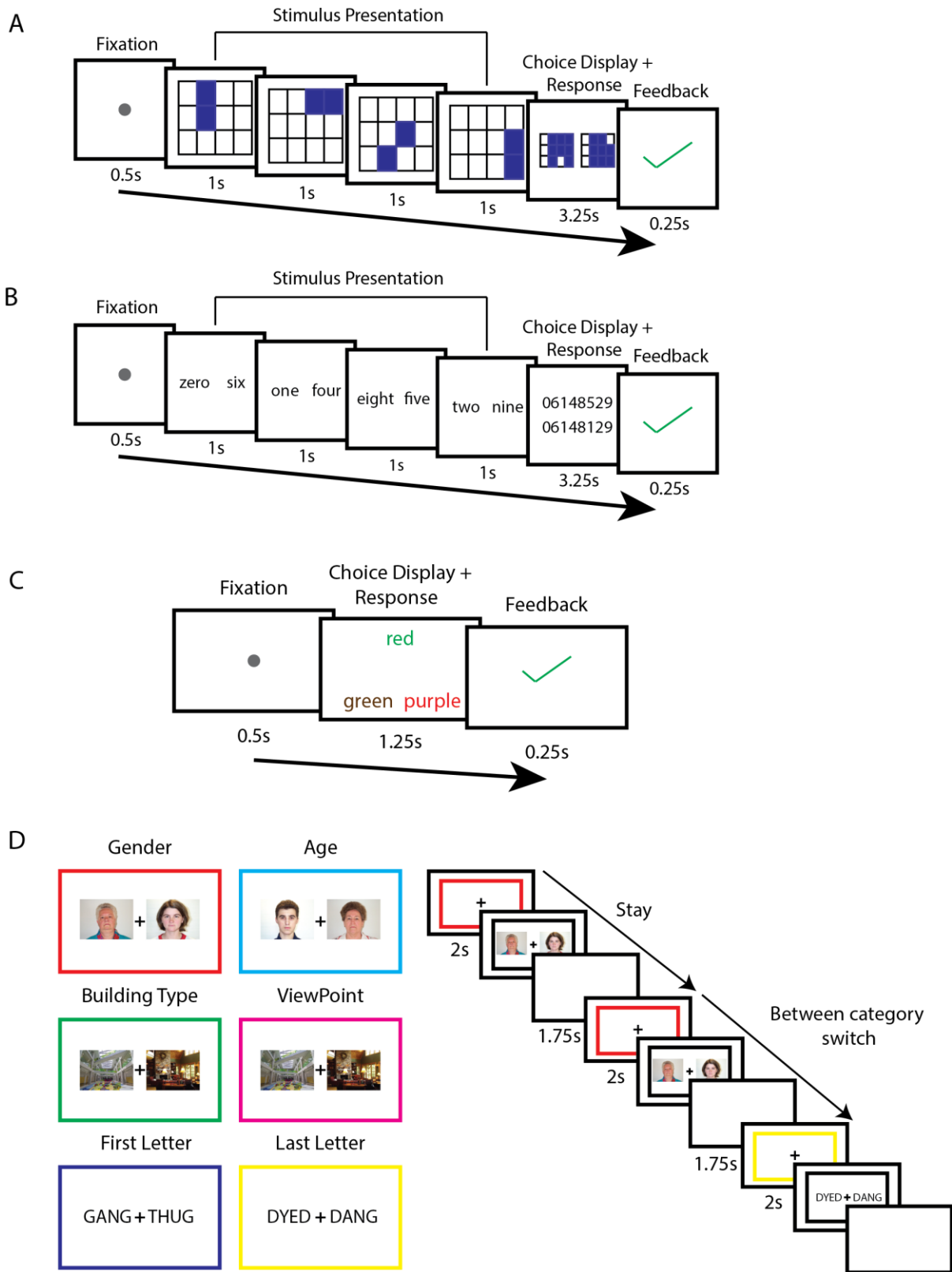
Figure 1. Schematic overview of the Hard condition of the three localizer tasks and the rule-based similarity judgement task. **A: Spatial working memory task.** Participants were presented with highlighted cells in a 3x4 grid, on four consecutive screens. In the Easy condition, one cell was highlighted at a time, and in the

11

Hard condition two cells were highlighted in each screen. They selected the grid with the correctly highlighted cells in a subsequent two-forced choice display. They received feedback after each trial. Positive feedback was indicated by a green tick, and negative by a red cross. **B: Verbal working memory task.** A design similar to the spatial working memory task was used, with written digits instead of a grid. **C: Stroop task.** In each trial, three color names were presented. Participants selected the answer option (out of two options at the bottom) that described the ink color of the test word on top. In the Easy condition, the ink color of the test word matched its color name (congruent), and in the Hard condition they were different (incongruent). In this example of a Hard trial, the correct answer is the word 'green' written in brown, and the ink color of the distractor (the word 'purple' written in red ink) matches the color name of the test word, thus increasing the difficulty level. See the text (*2.3.3*) for more details. For all localizer tasks, feedback was presented at the end of each trial. **D: Rule-based similarity judgement task.** The six rules with the corresponding colored frames, paired by the category domain that they should be applied on (left). In each trial, a colored frame indicated the rule, followed by two images. Participants indicated whether the images are the 'same' or 'different' based on the rule. Transitions between trials could be either the same rule and category domain ('Stay'), a switch of rule within the same category domain ('within-category Switch'), or a switch of rule between category domains ('between-category Switch'). In this example (right), participants indicated whether the two faces have the same gender or not in the first trial, followed by a Stay trial and a between-category Switch trial.

2.3.4. Criterion rule-based similarity judgement task

The rule-based similarity judgement task was a variation of a task previously used by Crittenden and colleagues (2016, 2015) and Smith et al. (2018). This was chosen as the criterion task because it allowed for testing both univariate and multivariate effects. Univariate effects were addressed using a task switching aspect of the task. Multivariate effects were addressed using rule decoding, which has been previously observed across the MD network using this task. Additionally, this task enabled a more detailed investigation of the potential effect of using individual ROIs on the difference between two types of rule discriminations.

Prior to the start of the scanning session, participants learned to associate colored frames with six rules (**Figure 1D**). In each trial, participants indicated whether two displayed images were the same or different based on the given rule. The six rules were applied on stimuli from three different category domains (faces, buildings, words), with two rules per category. The rules and category domains were: (1) Gender (male/female, red frame) and (2) Age (old/young, light blue frame) applied on faces; (3) Building type (cottage/skyscraper, green frame) and (4) Viewpoint (seen from the outside/inside, magenta frame) applied on buildings; (5) First letter (dark blue frame) and (6) Last letter (yellow frame) applied on words and pseudo-words.

Each trial began with a colored frame (2 s) that indicated the rule to be applied. This was followed by two stimuli presented to the left and right of a fixation cross (**Figure 1D**) and the cue replaced by a black frame. Participants had to respond 'same' or 'different' based on the rule by pressing the left or right button, and response mapping was counterbalanced across subjects. The task was self-faced and the stimuli were on the screen until a response was made. Each trial was followed by an inter-trial interval (ITI) of 1.75 s. The colored frame was 14.96° visual angle along the width and 11.60° along the height. The choice stimuli were displayed at 3.68° eccentricity from the center and were 6.0° and 4.5° along the width and height.

We used an event-related design, with 12 trials per rule in each run. Out of these, half (6) of the trials had 'same' as correct response and half (6 trials) had 'different' as correct response. Out of the 6 trials with 'same' as correct response, half (3 trials) also had 'same' as correct response if the other rule for the category was applied, therefore identical responses using either rule for this category. The other half (3 trials) had 'different' as correct response using the other rule in this category, therefore different responses for the two rules. A similar split was used for the 6 trials with 'different' as the correct response. To decorrelate the cue and stimulus presentation phases, 4 out of 12 trials of each rule were chosen randomly to be catch trials, in which the colored frame indicating the rule was shown, but was not followed by the stimuli.

The task included switches between rules that were used for the univariate analysis. Trial switches were defined based on the rules in two consecutive trials. In 'Stay' trials, the previous trial had the same rule and category domain. For 'within-category Switch' trials, the previous trial had the same

13

category domain (faces, buildings, or letters) but a different rule (e.g. age vs. gender for faces). For 'between-category Switch' trials, the previous trial had a different category domain and therefore necessarily a different rule. There was an equal number of Stay, within-category Switch and between-category Switch trials (24). The order of the trials was determined pseudo-randomly while balancing the types of switches. Each run included a total of 73 trials, with an extra trial at the beginning of the run that was required for the balancing of the switch types. This trial was assigned a random rule and was excluded from the analysis.

To address multivariate effects, we used decoding of rule pairs in the task. Each pair of rules out of the six could then be referred to as either 'within-category', i.e., applied on the same category domain, or 'between-category', i.e., applied on different category domain. The idea being, while all rules may be decoded across the frontoparietal network, the 'between-category' rules might be more distinct (i.e. higher decoding levels) than 'within-category' rules (Crittenden et al., 2016). To avoid confounding the rule representation with visual information in the task used by Crittenden et al. 2016, we used a variant of this task, with separate cue and stimulus presentation phases in each trial (Smith et al., 2018).

The participants practiced the task prior to the scanning session until they learned the rules. The practice consisted of two parts. During the first part, trials included feedback, while during the second part feedback was omitted. There was no feedback during the scanning session runs of this task. During the scanning session, after completing the localizer tasks, the participants were asked to state the six rules to make sure they remembered the rules before starting the rule-based task. They were also shown the rules again if they requested it. The task was self-paced, with an average duration of each run of 5 minutes and 52 seconds (SD = 29.1 seconds).

2.4. fMRI data acquisition

Participants were scanned in a Siemens 3T Prisma MRI scanner with a 32-channel head coil. A T2*-weighted 2D multiband Echo-Planar Imaging, with a multi-band factor of 3, was used to acquire 2 mm isotropic voxels (Feinberg et al., 2010). Other acquisition parameters were: 48 slices, no slice gap, TR = 1.1 s, TE = 30 ms, flip angle = 62^0, field of view (FOV) = 205 mm, in plane resolution: 2 x

14

2 mm. In addition to functional images, T1-weighted 3D multi-echo MPRAGE (van der Kouwe et al., 2008) structural images were obtained (voxel size 1 mm isotropic, TR = 2530 ms, TE = 1.64, 3.5, 5.36, and 7.22 ms, FOV = 256 mm x 256 mm x 192 mm). A single structural image was computed per subject by taking the voxelwise root mean square across the four MPRAGE images that are generated in this sequence.

2.5. Data Analysis

2.5.1. Pre-processing

Pre-processing was performed using the automatic analysis (aa) pipelines (Cusack et al., 2014) and SPM12 (Wellcome Trust Centre for Neuroimaging, University College London, London) for MatLab. The data were first motion corrected by spatially realigning the EPI images. The images were then unwarped using the fieldmaps, slice time corrected and co-registered to the structural T1-weighted image. The structural data were normalized to the Montreal Neurological Institute (MNI) template using nonlinear deformation, after which the transformation matrix was used to normalize the EPI images. For the univariate ROI analyses, the localizer data were pre-processed without smoothing. For the whole-brain random-effects analysis and voxel selection for subject-specific ROIs, the localizer data were smoothed with a Gaussian kernel of 5 mm full-width at half-maximum (FWHM). No smoothing was applied to the data of the rule-based criterion task.

2.5.2. General linear model (GLM)

A general linear model (GLM) was estimated per participant for each localizer task. Regressors were created for the Easy and Hard task blocks and were convolved with a canonical hemodynamic response function (HRF). Run means and movement parameters were used as covariates of no interest. The resulting β-estimates were used to construct the contrast of interest between the Hard and Easy conditions, and the difference in β-estimates (Δbeta) was used to estimate the activation evoked by each localizer.

For the rule-based criterion task, two different GLMs were used to address univariate and multivariate effects separately. The GLM for the univariate analysis was based on types of switches

15

between trials. A GLM was estimated for each participant using cue phase regressors for the three switch types: Stay, within-category Switches and between-category Switches. Additional stimulus phase regressors were used separately for each switch type, from stimulus onset to response, but these were not analyzed further. A similar GLM was used for multivariate analysis, based on the six rule types instead of switch types. The duration of the regressors was 2 s for all conditions in both models. The regressors were convolved with the canonical HRF and the six movement parameters and run means were included as covariates of no interest.

2.5.3. Localizers activity patterns

We used both random-effects whole-brain and ROI analyses to assess and compare the activation levels and recruitment of the MD network by the localizers. All measures used voxel data of the Hard versus Easy contrast and the resulting difference in beta estimates (Δbeta), computed across the two runs of each localizer as well as separately for each run when required and as detailed below for the specific analyses. Additional Hard versus Easy contrasts were computed using combinations of different localizers.

For ROI analysis, we used a template for the MD network, derived from an independent task-based fMRI dataset (Fedorenko et al., 2013, http://imaging.mrc-cbu.cam.ac.uk/imaging/MDsystem). The template is bilateral with an equal number of voxels in each hemisphere for each ROI. ROIs and their respective number of voxels per hemisphere include the anterior insula (AI, 992 voxels), posterior/dorsal lateral frontal cortex (pdLFC, 1132 voxels), intraparietal sulcus (IPS, 4260 voxels), the anterior, middle and posterior middle frontal gyrus (MFG) (621, 712 and 1269 voxels, respectively) and the pre-supplementary motor area (preSMA, 1247 voxels). The group-level ROI analyses were computed using the MarsBaR SPM toolbox (Brett, Anton, Valabregue, & Poline, 2002).

2.5.4. Measures to compare activation patterns of the localizers

We conducted several analyses to examine and compare the spread and similarity of activation patterns of the localizers.

2.5.4.1 Whole-brain spatial spread of activity patterns of localizers across subjects

We conducted a whole-brain analysis to examine the spread of activation patterns across individual participant data. For each voxel we computed the number of participants with significant activations by applying FDR ($p < 0.05$) across all voxels and all participants. This yielded a whole-brain map in which voxel data represents the number of participants with significant activation.

2.5.4.2 Correlation measures to compare activation patterns of the localizers

To quantify and compare activity patterns of the three localizers, we used Fisher-transformed Pearson correlation of Δbeta estimates (Hard − Easy) across voxels. Contrast data were estimated for each run separately and correlations were computed across all voxels in each MD ROI. First, we computed the reliability of activation patterns. For each subject and localizer we correlated the Δbeta estimates of the two runs, separately for MD ROIs and then averaged. Second, to compare activity patterns between subjects, we then estimated the similarity in activity patterns between subjects. For each subject, each localizer and each MD ROI we computed the correlation of Δbeta estimates between the first run of that subject and the second run of another subject of the same localizer. For each subject, this was computed 17 times, using all other subjects and averaged across them to get between-subject correlation of activity, separately for each of the three localizers. Lastly, to estimate the similarity in activity patterns between the different localizers, we computed for each subject the correlation between the first runs of each pair of localizers.

2.5.5. Individual MD localization using the group-constrained subject-specific approach

Individual subject ROIs were defined using the group-constrained subject-specific approach. For each localizer, we used the Hard versus Easy contrast data across the two runs to obtain Δbeta estimates. Then, for each ROI, the 200 voxels with the largest Δbeta estimates were selected. The number of voxels that were selected was defined prior to any data analysis.

The selected voxels were then used to test for the effects of subject-specific ROIs and choice of localizer on both the univariate and multivariate activity as measured in the rule-based criterion task.

We compared measures of activity using both subject-specific ROIs based on the different localizers and the group template (i.e., using all voxels within each ROI), as well as between localizers.

2.5.6 Univariate analysis of the criterion task

For the univariate activity in the criterion task, we used two contrasts with varying cognitive demand to test for the effect of subject-specific ROIs and choice of localizer across the MD network. For each subject, the GLM cue regressors across all four runs were used to compute two univariate contrasts: within-category Switch versus Stay trials, and between-category Switch versus Stay trials. The results were then averaged across hemispheres and subjects. Similarly to the Hard vs. Easy contrast in the localizer tasks, activity across the MD network is expected to increase with increased demand.

2.5.7. Multivoxel pattern analysis (MVPA) of the criterion task

We used rule decoding in the criterion task to test for the effect of subject-specific ROIs and choice of localizer on multivariate activity. The decoding analysis focused on the task rules during the cue phase, when only colored frames appear on the screen, therefore avoiding any confounds of the subsequent stimuli. Classification accuracy was computed using a support vector machine classifier (LIBSVM library for MATLAB, c=1) implemented in the Decoding Toolbox (Hebart, Gorgen, & Haynes, 2015). A leave-one-run-out cross-validation was employed to compute pairwise classifications for all task rule combinations (15 in total), and classification accuracy was averaged across all folds for each pair of rules. The average accuracy of all rule pairs, as well as separately for within- and between-category rule pairs, were computed for each subject and ROI. For each subject, classification accuracies were computed using subject-specific ROIs based on the three independent localizers' data separately, as well as using all voxels within each ROI (i.e. group template). Decoding accuracies above chance (50%) were then averaged across hemispheres and subjects and were tested against zero using one-tailed t-tests.

Since all localizers were chosen because their robust and consistent recruitment of the MD network, combinations of localizers could also be used to define subject-specific ROIs. To test for multivariate results in subject-specific ROIs defined using data combined from two different localizers,

18

we repeated the decoding analysis using contrast data comprised from two runs, one of each localizer. We used the first run of each localizer and created three combination contrasts from pairs of localizer tasks: spatial WM + verbal WM, spatial WM + Stroop, and verbal WM + Stroop. We also used all 6 localizer runs to create a spatial WM + verbal WM + Stroop contrast. The combination contrasts were used to define subject-specific ROIs for the decoding analysis in a similar way to the individual localizers' data.

One advantage of using group-constrained subject-specific ROIs for multivariate analysis in particular is that it allows to keep the number of voxels in each ROI fixed and controlled. This may be important as it has been previously demonstrated in the visual system that the size of ROI may affect decoding accuracy levels (Eger et al., 2008; Said et al., 2010; Walther et al., 2009). To test for potential effect of ROI size on the decoding results, the MVPA was repeated using a range of ROI sizes (50, 100, 150, 200, 250, 300, 350 and 400 voxels).

To ensure that our decoding results did not depend on the choice of classifier, we repeated the MVPA using a representational similarity analysis (RSA) approach, with linear discriminant contrast (LDC) as a measure of dissimilarity between rule patterns (Carlin & Kriegeskorte, 2017; Nili et al., 2014). Cross-validated Mahalanobis distances were calculated for all 15 pairwise rule combinations and averaged to get within- and between-category rule pairs, for each ROI and participant, using all the voxels in the ROI and subject-specific ROIs defined using the different localizer tasks. For each pair of conditions, we used one run as the training set and another run as the testing set. This was done for all pairwise combinations of the 4 runs and LDC values were then averaged across them. Larger LDC values indicate more distinct patterns of the tested conditions, while the LDC value itself is non-indicative for level of discrimination. The choice of using LDC rather than LD-t (associated t-value) meant that we could meaningfully look at differences between distances, and particularly the distinction between within- and between-category rule discriminations in the criterion task. We therefore used the difference of between- and within-category rule pairs compared to 0 as indication for representation of rule information.

2.5.8 Statistical testing and code

19

To compare performance of the localizers to the group levels as well as to each other, we first used a repeated measures ANOVA as a statistical model with factors as appropriate for each question and as detailed in the Results. The main factor of interest in all these ANOVAs is the localizer factor (with levels for the three localizers and the group template when required), and we report all the results related to this factor and its interactions, but not other interactions that are not of interest for our questions. To directly test for our research questions which may not have been captured by the ANOVA model, we also used separate t-tests for each localizer compared to the group template when appropriate, as well as for all possible pairs of localizers, both corrected for multiple comparisons.

We used an alpha level of .05 for all statistical tests. Bonferroni correction for multiple comparisons was used when required, and the corrected p-values and uncorrected t-values are reported. To quantify the evidence for difference in pattern discriminability when using the group template and subject-specific ROIs defined by functional localizers, we conducted a complementary Bayes factor analysis (Rouder et al. 2009). We used JZS Bayes factor for one-sample *t*-test and square-root(2)/2 as the Cauchy scale parameter, therefore using medium scaling. The Bayes factor is used to quantify the odds of the alternative hypothesis being more likely than the null hypothesis, thus enables the interpretation of null results. A Bayes factor greater than 3 is considered as some evidence in favor of the alternative hypothesis. All analyses were conducted using custom-made MATLAB (The Mathworks, Inc) scripts, unless otherwise stated.

2.6. Data and code availability statement

Anonymized data and code will be available in a public repository before publication. Data and code sharing are in accordance with institutional procedures and ethics approval.

3. Results

3.1. Behavioral results

The mean accuracies and reaction times (RT) for the Easy and Hard conditions of the spatial WM, verbal WM, and Stroop localizer tasks are listed in Table 1. As expected, there was a significant increase in RT during the Hard compared to the Easy condition for all localizers (two-tailed paired t-test: spatial WM: $t_{17} = 10.03$, $p < 0.001$; verbal WM: $t_{17} = 25.01$, $p < 0.001$; Stroop: $t_{17} = 9.89$, $p < 0.001$), as well as a significant decrease in accuracy (spatial WM: $t_{17} = 8.65$, $p < 0.001$; verbal WM: $t_{17} = 6.95$, $p < 0.001$; Stroop: $t_{17} = 7.47$, $p < 0.001$). These results confirmed that the task manipulation of Easy and Hard conditions worked as intended.

Accuracy levels for the rule-based criterion task were high (mean ± SD: 95.3% ± 2.3), indicating that the participants were able to learn the different rules and apply them correctly. The mean accuracies (mean ± SD): 93.5±3.4, 95.7±3.2, 94.1±4.2, 95.4±3.3, 95.4±4.6, 97.7±3.0, and the mean RTs (mean ± SD):1.33±0.4 s, 1.46±0.5 s, 1.55±0.5 s, 1.55±0.6 s, 1.27±0.4 s, 1.30±0.3 s, for age, gender, building type, viewpoint, first letter, and last letter rules, respectively. Since the focus of this study was to compare imaging results across different methods rather than linking them to behavior or making inference about the underlying cognitive construct, we did not analyze the accuracy levels and the RTs for the criterion task any further.

Table 1. RTs and accuracies in the localizer tasks. Values are means ± standard errors.

	Spatial WM		Verbal WM		Stroop	
	Easy	Hard	Easy	Hard	Easy	Hard
RT (sec)	1.15±0.05	1.63±0.06	1.02±0.04	1.91±0.04	0.78±0.02	0.90±0.02
Accuracy (% correct)	94.2±1.3	71.4±2.5	96.8±0.9	79.2±2.6	94.4±0.9	68.5±4.1

*3.2 Whole-brain and ROI univariate analysis of the localizer tasks*

To test for the recruitment of the MD network, a whole-brain random effects analysis was conducted for the Hard versus Easy contrasts of each localizer task (**Figure 2A**). The whole-brain patterns of activity clearly showed recruitment of the MD network by all localizers. Areas of increased activity included the anterior-posterior axis along the middle frontal gyrus, anterior insula, and the area anterior to the FEF on the lateral surface; preSMA on the medial surface; and IPS on both the lateral and medial surfaces. An additional visual component was observed, as expected from the nature of the tasks. The pattern of activity for the Stroop localizer was sparser, in particular on the right hemisphere.

An ROI analysis further confirmed the recruitment of the MD network with increased task difficulty (**Figure 2B**). All localizers showed a significant increase in activation for the Hard compared to the Easy condition in each of the ROIs with Bonferroni correction (3) for multiple comparisons (one-sample two-tailed t-test against 0: $t_{17} \geq 2.57$, $p \leq 0.02$ for all). A three-way repeated measures ANOVA with factors task (spatial WM, verbal WM, Stroop), ROI (7) and hemisphere (left, right) revealed a main effect of task ($F_{2, 34} = 3.96$, $p = 0.028$). There was a significant main effect of ROIs ($F_{6, 102} = 7.6$, $p < 0.001$), indicating some differences in activity levels between the ROIs. Although the recruitment of the two hemispheres was broadly similar, there was also an effect of hemisphere ($F_{1, 17} = 15.54$, $p = 0.001$), with larger activity on the left than the right hemisphere. There were significant interactions between tasks and ROIs, tasks and hemispheres, ROIs and hemispheres and a three-way interaction ($F > 4.1$ $p < 0.009$). Post-hoc pairwise comparisons with Bonferroni correction (3 comparisons) across ROIs demonstrated a marginal difference in activity between the verbal WM and Stroop task ($t_{17} = 2.64$, $p = 0.051$), but the overall activity did not differ between the spatial WM and verbal WM tasks or the spatial WM and Stroop tasks ($t_{17} \leq 1.25$, $p \geq 0.3$ for both). Overall, all the tasks recruited the MD network, with some differences in the activation patterns across ROIs for the different localizers. Importantly, the univariate results confirmed a significant increase in activity in the MD network with increased task difficulty for all three localizers, as expected and designed.
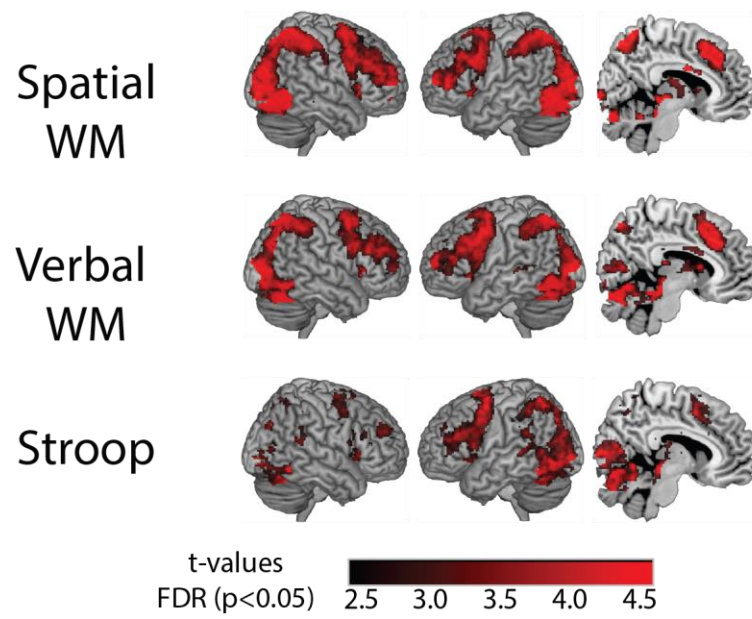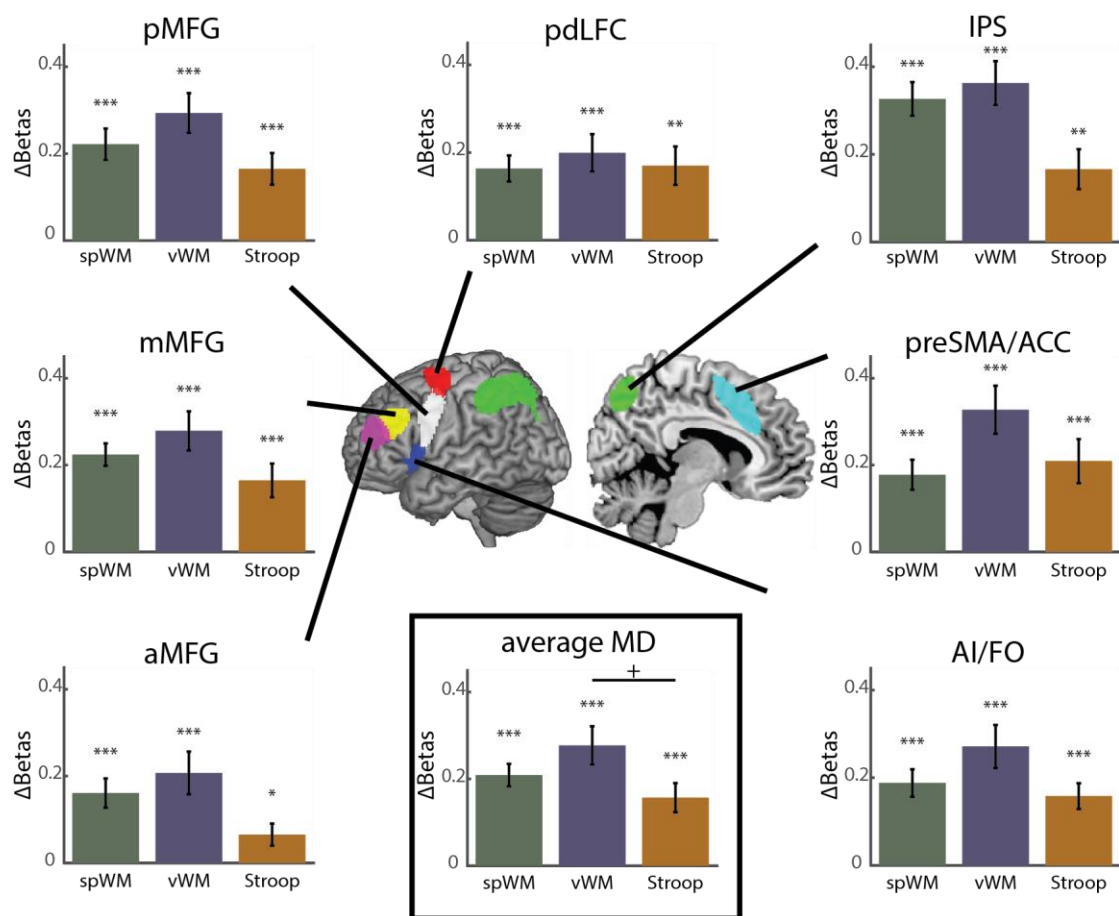
Figure 2. Increased activity across the MD network with increased difficulty level for all three localizers. A: Whole-brain t-maps of the contrast between Hard and Easy conditions for the localizer tasks. WM = working memory. t-maps are FDR corrected, p<0.05. **B:** Univariate results for the contrast between the Hard and Easy conditions for the three localizer tasks, per MD ROI and averaged across ROIs. Significance levels above zero in each ROI and for each localizer (with Bonferroni correction for 3 comparisons) are shown to demonstrate recruitment of the MD network. For the average across MD ROIs, pairwise statistical testing between localizers were also done using a paired two-tailed t-test with Bonferroni correction (3 comparisons). Error bars indicate SEM. The MD template that was used for ROI analysis is shown in the middle for reference (adapted from Fedorenko et al. 2013). spWM = spatial working memory, vWM = verbal working memory. pdLFC = posterior/dorsal lateral prefrontal cortex, IPS = intraparietal sulcus, preSMA = pre-supplementary motor area, AI = anterior insula, aMFG = anterior middle frontal gyrus, mMFG = middle frontal gyrus, pMFG = posterior middle frontal gyrus. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$, + $p < 0.06$.

3.3 Comparisons of activity patterns between localizers

We further quantified and compared the variability of activity patterns across subjects and localizers. To test for variability across subjects, two methods were used: A whole-brain overlay of subjects' activation per voxel and correlations of activity across voxels between runs of the same subject versus different subjects. Variability across localizers was tested by correlating the activity across voxels between runs of the same localizer versus different localizers.

3.3.1 Variability of activity patterns across subjects

While the group averages of the Hard versus Easy contrasts of the three localizers closely resembled the MD template, there was substantial variability between activation patterns of individual participants. To visualize this spread of activations across individuals, we computed a whole-brain overlay map for each localizer (**Figure 3A**). Each voxel shows the number of participants with significant activations after applying FDR ($p < 0.05$) correction across voxels and subjects. The overlay maps show peaks of activation for many subjects in areas similar to those observed in the group

24

averages. However, it is also clear from these maps that there is substantial variability across subjects with activations that extend to large parts of the frontal and parietal cortices, as well as the visual cortex as expected from the visual nature of the tasks.

Correlations of activity of the different localizers for each subject and between subjects further demonstrated that activity patterns are subject-specific. For each localizer, we computed the reliability of activation as the correlation between runs of the same localizer and same subject using all voxels within each ROI of the MD template. These reliabilities were compared to correlation between two runs from different subjects and same localizer. The correlations were averaged across ROIs to get the correlations of activity across the MD network (**Figure 3B**). The reliability of activation patterns for each subject was high for all localizers (Mean ± SD: spatial WM: 0.72 ± 0.13; verbal WM: 0.75 ± 0.14; Stroop: 0.53 ± 0.34), and their similarity was substantially lower for different subjects (Mean ± SD: spatial WM: 0.19 ± 0.05; verbal WM: 0.15 ± 0.04; Stroop: 0.08 ± 0.06). A two-way repeated measures ANOVA with task (3: spatial WM, verbal WM and Stroop), and correlation type (2: within- or between-subjects) as factors showed that the within-subject reliabilities were significantly larger than the between-subject correlations ($F_{1, 17} = 322.4$, $p < 0.001$), providing another support for the variability of activity patterns across subjects. There was also a significant main effect of task ($F_{2, 34} = 8.5$, $p < 0.001$), with no interaction between the factors ($F_{2, 34} = 2.3$, $p = 0.11$). Post-hoc pairwise comparisons with Bonferroni correction (3 comparisons) showed that the Stroop task had lower correlations than the spatial WM ($t_{17} = 2.8$, $p = 0.037$) and verbal WM ($t_{17} = 3.34$, $p = 0.012$) tasks.

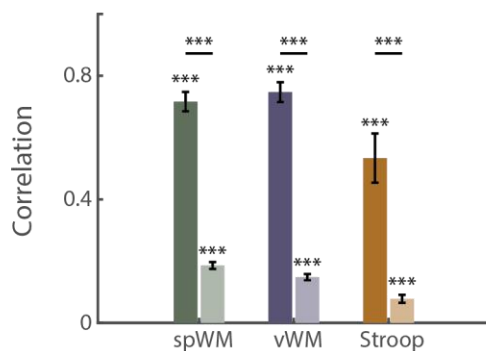*3.3.2 Variability of activity patterns across localizers*

To assess the similarity of activity patterns between localizers, we computed the within-subject correlation of activity pattern across voxels between the first runs of pairs of localizers (**Figure 3C**). As expected, there was a substantial positive correlation between the localizers (Mean ± SD: spatial WM with verbal WM: 0.67 ± 0.36; spatial WM with Stroop: 0.27 ± 0.43; verbal WM with Stroop: 0.46 ± 0.33), with correlations between the working memory tasks and between verbal WM and Stroop being well above 0 (two-tailed t-test against zero corrected for 3 comparisons: $t > 6$, $p < 0.001$), and a marginally significant correlation between the spatial WM and Stroop correlation following Bonferroni

25

correction ($t_{17} = 2.65$, $p = 0.051$). These correlations were significantly smaller than the within-subject within-localizer reliabilities (two-tailed t-test of the average reliabilities across localizers vs. the average correlations of pairs of localizers: $t_{17} = 9.37$, $p < 0.001$). These reduced correlations demonstrated that although all localizers recruited the MD network, there was some spatial variability in their activation patterns.
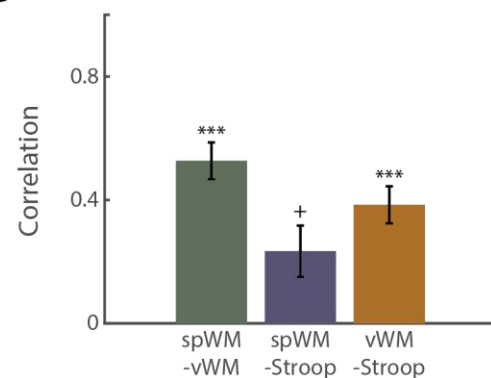
A
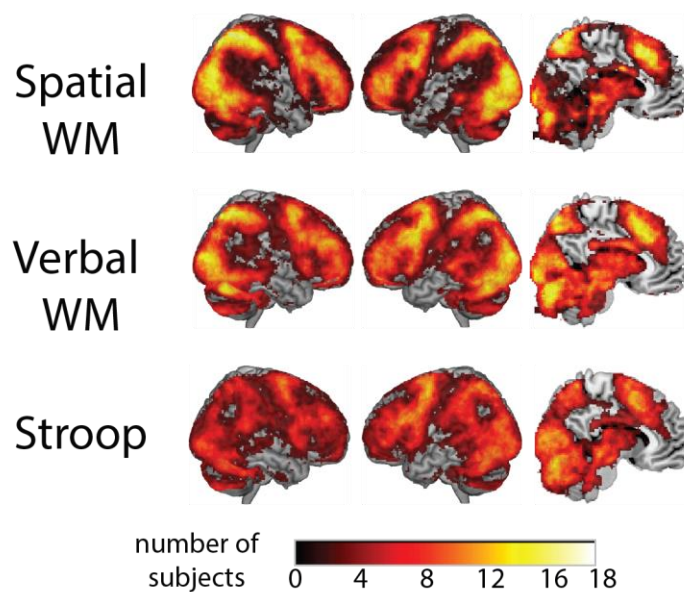
B

C

Figure 3. Activity pattern are variable across subjects and localizers. A: Activation patterns of single subjects only partially overlap, demonstrating variability across participants. The color of each voxel shows the number of subjects that had significant activation in that voxel in the Hard versus Easy contrast, with the color bar indicating number of subjects up to 18 (sample size), thresholded at 1 subject. **B.** Correlation

26

of activity within runs of the same subject (reliabilities, darker bars) and runs of different subjects (lighter bars) for each localizer, averaged across MD ROIs. **C.** Within-subject correlation of activity patterns between pairs of localizers, averaged across MD ROIs. Pearson correlations are presented, while Fisher transformed correlations were used for statistical inference. Asterisks above bars show significance levels (two-tailed t-test against zero, corrected for 3 comparisons). Asterisks above horizontal lines between bars show significance levels of differences (paired two-tailed t-test against zero, corrected for 3 comparisons). Error bars indicate SEM. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

3.4 Subject-specific ROIs and univariate activity in the rule-based criterion task

The main aim of our study was to test for the effect of subject-specific ROIs on univariate and particularly multivariate activity measures in the MD network. We used the rule-based criterion task to extract both univariate and multivariate measures and tested whether using subject-specific ROIs using the independent localizers' data affects the ability to identify activity as expected in the MD network, and whether such changes depend on the choice of localizer.

For the univariate activity, we computed two task switch contrasts in the criterion task: the more demanding between-category Switch versus Stay trials, and the less demanding within-category Switch versus Stay trials. The analysis was done using subject-specific ROIs based on the localizers' data as well as using all voxels within each ROI and averaged across all MD ROIs (**Figure 4**). A three-way repeated measures ANOVA with task (spatial WM, verbal WM, Stroop, all-voxels), contrast type (within-category Switch versus Stay, between-category Switch versus Stay) and ROI (7) was set to test for the effect of using subject-specific ROIs and localizer choice on activity levels. There was a main effect of task ($F_{3, 51} = 14.21$, $p < 0.001$) and post-hoc tests with Bonferroni correction for 6 comparisons showed that activity when using the group template was lower than when using subject-specific ROIs, for all localizers ($t_{17} > 3.5$, $p < 0.02$). The activation using subject-specific ROIs based on the spatial WM was lower than that of verbal WM ($t_{17} = 3.40$, $p = 0.02$) and other comparisons between localizers were not significant ($t_{17} > 1.4$, $p > 0.9$). As expected for activity in the MD network, there was a main effect of contrast type ($F_{1, 17} = 12.12$, $p = 0.003$), with activity for the more demanding between-category Switch being larger than the within-category Switch. There was also an interaction between task and

27

contrast type ($F_{3,\,51}$ = 7.92, $p < 0.001$), but post-hoc tests with Bonferroni correction for 4 comparisons showed that activity for the between-category Switch was larger than for the within-category Switch for all localizers and when the group template was used ($t_{17} > 2.9$, $p < 0.036$). Overall, these results demonstrate that subject-specific ROIs leads to an increase in the observed univariate results, with similar effects for the different localizers.
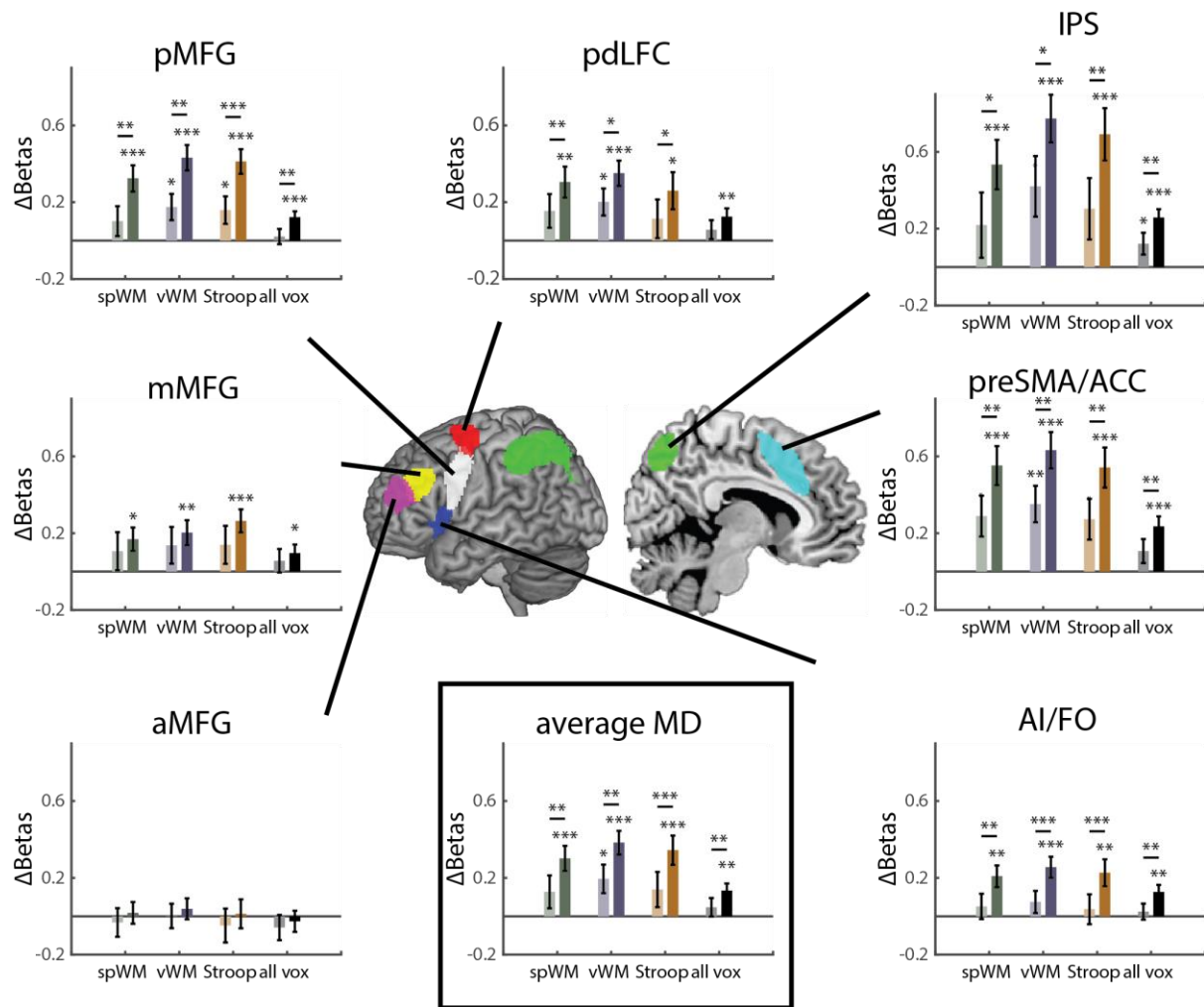
Figure 4. Univariate activity in the criterion task increases when using subject-specific ROIs compared to the group template. Within- (lighter bars) and between- (darker bars) category versus Stay trials contrast values using subject-specific ROI defined by the different localizer tasks and using all voxels (group template, no localizer), per ROI and averaged across MD ROIs. Δbeta values were larger when using subject-specific ROIs compared to when using the group template, and similar for all three localizers. Between-category Switch activity was larger than the within-category Switch for all

28

localizers and when using all voxels, as expected for the MD network. spWM = spatial working memory, vWM = verbal working memory. pdLFC = posterior/dorsal lateral prefrontal cortex, IPS = intraparietal sulcus, preSMA = pre-supplementary motor area, AI = anterior insula, aMFG = anterior middle frontal gyrus, mMFG = middle frontal gyrus, pMFG = posterior middle frontal gyrus. Significant activation levels are shown above bars (two-tailed t-test against 0). For visualization, asterisks above horizontal lines between bars show significance levels of differences (paired two-tailed t-test against zero). Error bars indicate SEM. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

3.5 Subject-specific ROIs and rule decoding in the rule-based criterion task

We next tested for the effect of subject-specific ROIs on decoding levels in the rule-based criterion task across the MD network. We computed the decoding accuracy of all pairwise discriminations between rules when using each localizers' data to select the 200 most responsive voxels within each ROI for each subject and when using all the voxels in each ROI. Overall decoding accuracies above chance (50%) across all MD ROIs and rule-pairs were 2.85 ± 14.53, 2.71 ± 14.75, 3.17 ± 15.04, 5.08 ± 15.34 for subject-specific ROIs defined using the spatial WM, verbal WM and Stroop localizers, and when using the group template, respectively (mean ± SD). These decoding levels were similar to previous studies that used a similar experimental paradigm (Crittenden et al., 2015; Smith et al., 2018). We first tested whether overall decoding accuracy across all pairs of rules differed between the localizers and compared to when all voxels within each ROI were used. A two-way repeated measures ANOVA with task (spatial WM, verbal WM, Stroop, all-voxels) and ROI (7) showed a main effect of task ($F_{3, 51} = 2.9$, $p = 0.042$). However, none of the post-hoc tests to compare pairs of tasks survived correction (Bonferroni correction for 6 comparisons, $t_{17} < 2.7$, $p > 0.09$). There was a main effect of ROI ($F_{6, 102} = 2.5$, $p = 0.025$) but no interaction with task ($F_{18, 306} = 0.8$, $p = 0.7$). Overall, decoding across all pairs of conditions was similar for all localizers and when the group template was used.

Our choice of criterion task has enabled us to not only test for the effect of localizer type on the overall discrimination between rules, which might be too coarse to depict, but also for the relative decoding levels of the two types of discriminations, thus potentially picking up more subtle effects. The

29

criterion task included discriminations between rules applied on the same category (within-category discriminations, e.g., between the gender and age rules, both applied on the faces category) and discriminations between rules applied on different categories (between-category discriminations, e.g., between the gender rule applied on the faces category and the viewpoint rule applied on the building category). To get this more fine-grained picture of the effect of subject-specific ROIs using localizer data on rule decoding, as well as when the groups template is used, we split the decoding accuracy to within- and between-category rule discriminations (**Figure 5**). A three-way repeated measures ANOVA with task (spatial WM, verbal WM, Stroop, all-voxels), distinction type (within-category, between-category) and ROI (7) as within-subject factors was set to test for differences between localizers related to decoding of the two distinction types. There was no main effect of task ($F_{3, 51} = 1.0$, $p = 0.4$), and no interaction between distinction type and task ($F_{3, 51} = 0.9$, $p = 0.4$). To test for differences between each of the localizers and the group template as well as pairs of localizers based on our pre-defined questions, additional t-tests between all pairs of tasks were conducted, but no significant differences were found ($t_{17} < 1.8$, $p > 0.5$, corrected for 6 comparisons). There was no main effect of distinction type ($F_{1, 17} = 3.2$, $p = 0.089$), though a numerical trend was consistent with the previously reported results (Crittenden et al., 2016). There was no main effect of ROI or interaction of ROI with distinction type or task ($F < 1.8$, $p = 0.09$). Taken together, this indicates that decoding results in the criterion task were similar for all localizers as well as when the group template was used, similar to the results obtained across all pairs of conditions.

Based on these ANOVA results and to further establish that using localizer data for subject-specific ROIs did not change decoding levels, we conducted a Bayes factor analysis (Rouder et al. 2009), separately for each functional localizer compared to decoding with all voxels. First, the difference in classification accuracy between the between- and within-category distinctions averaged across ROIs for each functional localizer was compared to the classification accuracies with all voxels using a paired two-tailed t-test. The t-value was then entered into a one-sample Bayes factor analysis with a Cauchy scale parameter of 0.7. The Bayes factors for the spatial WM, verbal WM and Stroop localizer when compared to the decoding levels with all voxels were 0.40, 0.56 and 0.79, respectively.

These results demonstrate little evidence for difference in decoding levels when subject-specific ROIs and the group template are used.

Individual-subject localization of the MD network for MVPA can be done not only using one localizer task, but also using two runs of two different localizer, thus benefitting from the activation patterns evoked by both localizer to more robustly identify MD-like voxels. To test for the effect of such an approach on decoding results, we repeated the analysis using subject-specific ROIs based on localizer data combined from two different localizers (spatial WM + verbal WM, spatial WM + Stroop, verbal WM + Stroop). A three-way repeated measures ANOVA with task (4: three combination contrasts and all-voxels), distinction type (within/between category) and ROI (7) as within-subject factors did not show a main effect of task ($F_{3, 51} = 2.5$, $p = 0.07$), similarly to the results when individual localizers were used to define subject-specific ROI. There was no effect of distinction type ($F_{1, 17} = 3.3$, $p = 0.08$), despite a numerical trend, and there was no interaction of task and distinction type ($F_{3, 51} = 1.4$, $p = 0.25$). Similarly, subject-specific ROI were also defined using all localizers with one contrast using all 6 localizer runs (spatial WM + verbal WM + Stroop). A three-way repeated measures ANOVA with task (2: one combination contrast and all-voxels), distinction type (within/between category) and ROI (7) as within-subject factors showed no main effect of task or distinction type or an interaction of the two ($F < 3.2$, $p > 0.05$). Overall, these results indicate that using combinations of localizer tasks to define subject-specific ROIs yielded decoding results similar to the ones obtained when using the group template, at least with the range of localizers that we used here.
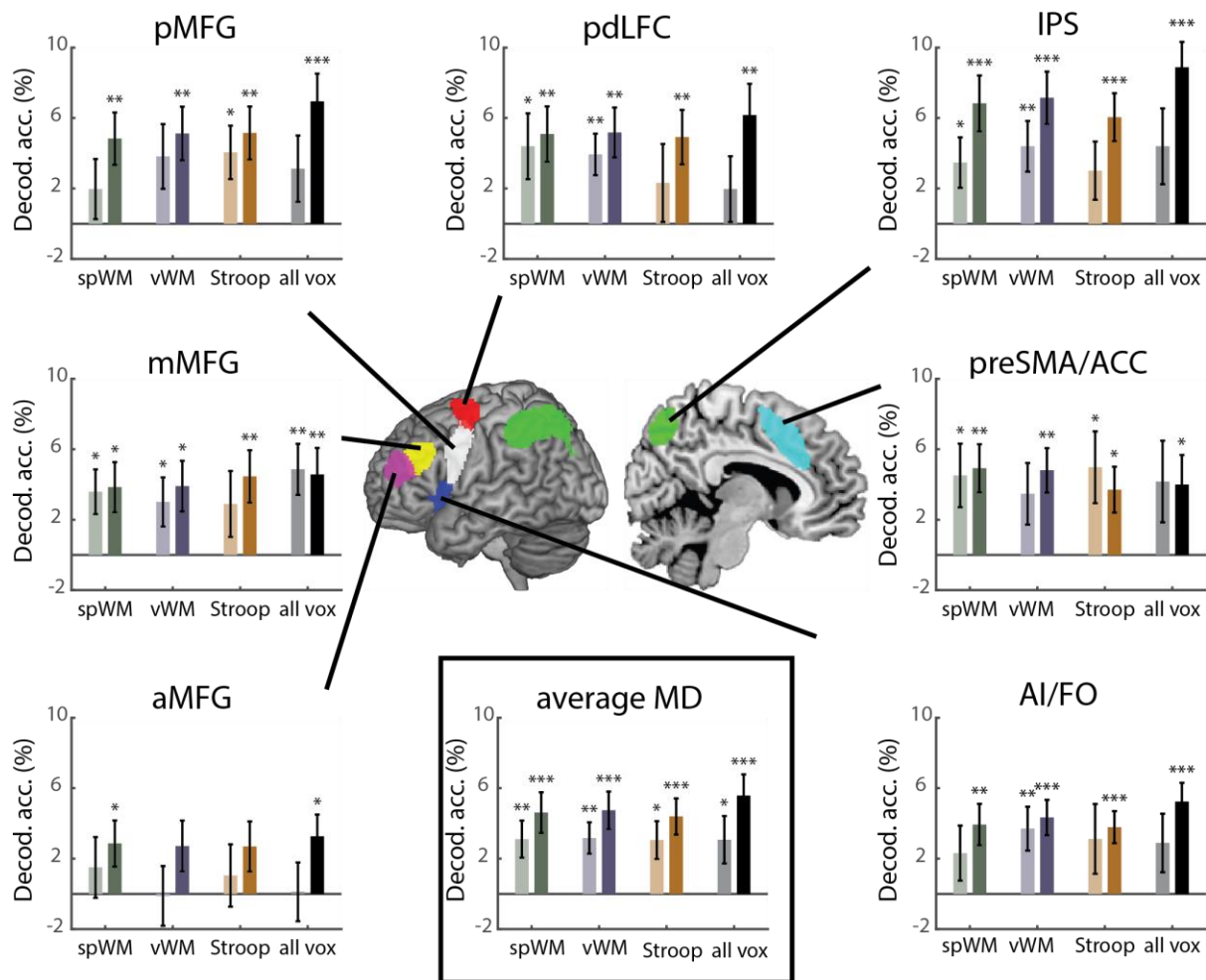
Figure 5. Decoding accuracy for within- and between-category rule pairs. Within- (lighter bars) and between- (darker bars) category rule decoding accuracy values above chance (50%) for subject-specific ROIs defined using the different localizer tasks and using all voxels (group template), per ROI and averaged across MD ROIs. Decoding accuracies were similar for all three localizers and when using all voxels within each template ROI, with similar differences between within- and between-category rule decoding (see Text (*3.5*) for statistical details). spWM = spatial working memory, vWM = verbal working memory. pdLFC = posterior/dorsal lateral prefrontal cortex, IPS = intraparietal sulcus, preSMA = pre-supplementary motor area, AI = anterior insula, aMFG = anterior middle frontal gyrus, mMFG = middle frontal gyrus, pMFG = posterior middle frontal gyrus. To demonstrate rule decoding, significant decoding accuracy above chance (50%) is shown (one-tailed t-test against 0). Error bars indicate SEM. * p < 0.05, ** p < 0.01, *** p < 0.001.

32

To ensure that our results are robust and not limited to the choice of classifier, we repeated the analysis using a representational similarity analysis (RSA) approach (N. Kriegeskorte, Mur, & Bandettini, 2008; Nili et al., 2014) and linear discriminant contrasts (LDC). An LDC value between pair of conditions indicates the level of their discriminability, with larger values meaning better discrimination. The difference between the average LDC of all the between-category rule pairs and all the within-category rule pairs (ΔLDC) was calculated per participant and per ROI, for both subject-specific ROIs and group template, with ΔLDC larger than 0 as an indication for rule information. For all localizers as well as when all voxels within the group template were used, the ΔLDC was greater than zero (two tailed: $t_{17} > 3.3$, $p < 0.016$, with Bonferroni correction for multiple (4) comparisons), indicating that the distributed patterns of activity conveyed rule information. This was comparable to the trend of distinction type effect seen in the SVM analysis. Importantly, a two-way repeated measures ANOVA with ΔLDC as dependent variable and task (spatial WM, verbal WM, Stroop, all-voxels) and ROI (7) as within-subject factors showed no main effect of task ($F_{3, 51} = 0.3$, $p = 0.7$) and no interaction ($F_{18, 306} = 1.1$, $p = 0.35$), further supported by individual t-tests of each localizer compared to the group template as well as all possible pairs of localizers ($t_{17} < 0.8$, $p > 0.8$, Bonferroni corrected for 6 comparisons). These results indicate that discriminability was similar when using the three localizers to define ROIs in individual subjects and when using all voxels within the group template, in line with the SVM decoding results.

3.6 Effect of ROI size on decoding results

In order to examine whether decoding results depend on the number of selected voxels in subject-specific ROIs, as has been observed in the visual system, we performed MVPA for the decoding accuracies across all MD ROIs using different ROI sizes. ROI sizes ranged from 50 to 400, in steps of 50. A three-way repeated measures ANOVA was done with factors: task (spatial WM, verbal WM, Stroop), ROI size (8) and distinction type (2: within- and between-category). There was a main effect of ROI size ($F_{7, 119} = 11.6$, $p < 0.001$), no main effect of task ($F_{2, 34} = 0.1$, $p = 0.9$) and no interaction between ROI size and task ($F_{14, 238} = 0.7$, $p = 0.7$), with the latter indicating that the difference between

33

ROI sizes was the same for the different localizers. We then pooled the data across the three localizers and rule distinctions in order to visualize the main effect of ROI size (**Figure 6**). Classification accuracies tended to be lower for the smaller ROI sizes, and particularly for 50 and 100 voxels, but overall decoding levels were stable with similar decoding accuracies for ROI size of 150 voxels and above. For all ROI sizes, classification accuracy was above chance ($t > 3.4$, $p < 0.003$).
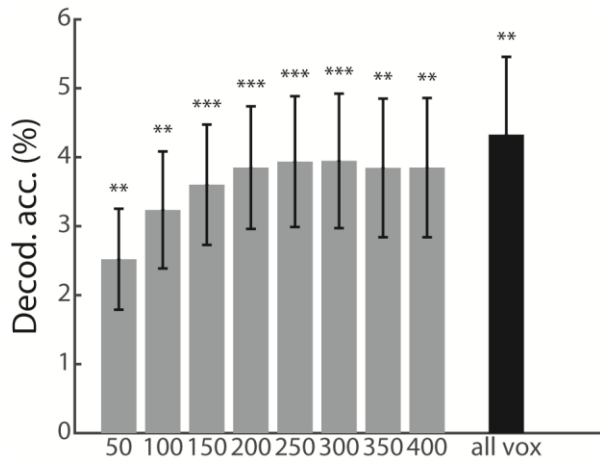
Figure 6. Decoding accuracy for rule distinction for different ROI sizes. Decoding accuracy above chance (50%) is presented for average of all the localizers, MD ROIs and distinction types. The decoding accuracy level using all the voxels in the MD ROIs (group template) is shown for reference at the rightmost bar. Error bars indicate SEM. ** = p < 0.01, *** = p < 0.001.

4. Discussion

In this study we tested for the effect of using individually-defined ROIs within the cognitive control frontoparietal MD network compared to a group template on both univariate and multivariate results in a rule-based criterion task. We systematically tested three localizer tasks (spatial WM, verbal WM, and Stroop) and used a group-constrained subject-specific approach to define ROIs at the single subject level using a conjunction of the independent localizer data and a group mask (E. Fedorenko et al., 2010). The primary benefit of the proposed approach is the use of both a group template that ensures spatial consistency across participants, as well as individual-subject activation patterns within the group template that provide more focused targeting of MD voxels. We showed a clear benefit for using the individual ROIs compared to when using the group template for univariate contrasts of activity. For multivariate discriminability measures, however, the results were similar for the individually-defined ROIs and for the group template, with no clear benefit, or cost, for the subject-specific ROIs approach. Despite differences between the localizers in their spatial activation patterns of the MD network at the individual subject level, we observed similar performance for both univariate and multivariate measures for all three localizers, demonstrating that the choice of localizer did not make a difference to the obtained results. Overall, our results demonstrate that using individually-defined ROIs is a useful way to maintain, or even increase in the case of univariate activity, the sensitivity of broad group templates with an added specificity of activations at the individual subject level.

Our univariate results show that the observed activity is larger when ROIs are defined at the individual level compared to using a group template. These findings replicate and generalize previous findings in other systems such as language and vision, as well as in simulated data (E. Fedorenko et al., 2010; Nieto-Castañón & Fedorenko, 2012; Saxe et al., 2006). When using the group template, many voxels outside the subject-specific functional signature of the studied region at the individual level are included in the analysis. These voxels average-out the overall activity level, reducing its sensitivity to detect actual changes (Nieto-Castañón & Fedorenko, 2012). Previous studies showed that boundaries between parcellated networks varied across individuals (Glasser et al., 2016; Schaefer et al., 2018; Yeo et al., 2011), further supporting the idea that overall average of activity within a group template may not capture changes that can be observed when ROIs are localized in individual subjects.

35

Our main focus in this study was the multivariate results in the MD network and whether individually-localized ROIs will result in a benefit, as seen in pattern discriminability measures, similar to the one observed in the univariate results. Most studies that used MVPA for fMRI data across the frontoparietal network used a group template as ROIs, implemented in a variety of ways. These include, among others, using all voxels within the functionally-defined group-average MD ROIs (Muhle-Karbe, Duncan, De Baene, Mitchell, & Brass, 2017; Woolgar, Hampshire, Thompson, & Duncan, 2011), defining areas of interest based on univariate or multivariate effects of part of the data of the main task and testing on another (Ester, Sprague, & Serences, 2015; Etzel et al., 2016), centering spheres on peak activation loci (M D. Fox, Snyder, Barch, Gusnard, & Raichle, 2005), resting-state networks (Cole et al., 2013), searchlight algorithm (Cole, Ito, & Braver, 2016), and anatomical landmarks in conjunction with group-level univariate contrast (Curtis, Cole, Rao, & D'Esposito, 2005). Because boundaries between specialized areas and networks vary between individuals (Glasser et al., 2016; Schaefer et al., 2018; Yeo et al., 2011), and specifically for the MD network (Evelina Fedorenko et al., 2012), the individual ROIs approach has the potential to improve our ability to reliably discriminate patterns of activity within this network. Our results, however, showed that this is not the case. In contrast to the univariate results, pattern discriminability was largely similar when using the individual ROIs and the group template. Our criterion task was chosen to allow for more subtle distinctions between between-category and within-category rule pairs that have been previously observed across the MD network (Crittenden et al. 2016). In the task used by Crittenden et al. (2016), the stimuli were presented at the same time as the colored frames, therefore the between-category and within-category rule pairs were confounded with the category of the stimuli. To avoid this confound, a later study used a variation of the task presenting only colored frames first, and the stimuli later without the cue (Smith et al., 2018). Using this cue and stimulus separation resulted in no difference between the between- and within-category rule pairs, despite strong decoding overall across the MD network. We investigated whether the subject-specific ROIs might show more subtle distinctions between rule pairs. However, the differences between the between-category and within-category rule pairs were similar when using the individual ROIs and the group template, with no benefit for the subject-specific ROIs. Importantly, using the individual ROIs for MVPA did not lead to reduced discriminability. The preserved accuracy

36

levels demonstrated that the increased localization at the individual subject level did not come at a cost of reduced decodability and maintained the sensitivity to detect task-related neural representations. We note that these results do not depend on the choice of classifier. To ensure the robustness of our results, we used linear discriminant contrast (LDC) in addition to SVM to measure discriminability and differences between localizers and observed similar results. Therefore, the similar multivariate performance when using individual ROIs and the group template was independent of the choice of the MVPA method.

One possible explanation for the similarity of multivariate results obtained for the individual ROIs and the group template is that voxelwise distributed pattern across the entire MD template captured the information related to rule decoding well, in the criterion task that we used, with no need for further refinement of the voxels selected for this decoding. Previous studies indeed showed that multivoxel discrimination may also be driven by voxels outside the focused regions of increased univariate activity (Haxby et al., 2001; N. Kriegeskorte et al., 2006). Another related point is the relatively high decoding accuracies that we observed when using the group template. Decoding levels were at a level similar to what has been shown to be the base rate for the frontoparietal network (Bhandari et al., 2018), therefore potentially limiting our ability to identify increases in decoding levels. Notably, an important limitation of our data, is the use of only one criterion task due to limited time in the scanner. It is possible that multivariate benefits or costs of the individual ROIs will be observed for other tasks, and future studies will be required to generalize our results in that respect. We note, that since our analysis focused on the fixed-duration cue phase of the trials, our results are well controlled and not driven by behavioral responses such as reaction times. More generally, the underlying factors that contribute to pattern discriminability in the frontoparietal network are not yet well understood (Bhandari et al., 2018), with some previous data showing clear limitations of fMRI decodability compared to what is observed in single-unit data in other brain systems (Dubois, de Berker, & Tsao, 2015). Our data provide another tier of evidence to better understand the relationship between the spatial organization and activity of the MD network at the micro and macro levels and pattern decodability using fMRI.

In this study, we systematically tested three localizers: a spatial WM, verbal WM and a Stroop task. We used tasks that capture a core cognitive aspect associated with the MD network and have the potential to be used as general functional localizers. As expected, and in line with previous results (E. Fedorenko et al., 2013), all three localizers showed increased activity in the MD network, thus confirming their suitability to serve as localizer tasks. Activation patterns of the localizers in individual subjects were highly reliable, as reflected in high correlations across voxels between the two runs of each localizer. In contrast, these correlations were substantially reduced across subjects, as is also shown in the whole-brain overlay maps, demonstrating the need for a subject-specific ROI definition approach. In line with previous data (E. Fedorenko et al., 2013), there were substantial correlations between activation patterns of the different localizers (computed within each subject), providing further support these correlations were lower than those between runs of the same localizer. The Stroop task evoked weaker, and less MD-focused pattern of activity, and had lower reliability between runs. These differences compared to the other two localizers were small and in some cases only marginally significant, but may imply that the Stroop task captured the MD network less well. The spatial differences in recruitment between the localizers may reflect differential functional preferences for cognitive demands and constructs across MD regions (Assem, Glasser, Essen, & Duncan, 2019; Crittenden et al., 2016; Dosenbach et al., 2007; Nomura et al., 2010; Shashidhara et al., 2019). Specifically, the two working memory tasks might reflect a more similar cognitive construct compared to the Stroop task that involves conflict monitoring and inhibition. Another possible explanation for the difference could be related to the difficulty manipulation in the tasks. While increase in difficulty level in the working memory tasks was simply controlled by increasing the number of highlighted cells in the grid or numbers, this manipulation in the Stroop task was operationally less well defined.

An important aspect of the individual ROI approach that we used is the ability to control for the ROI size. It has been previously shown that increased ROI size leads to increased classification levels in the visual system, highlighting the need to control for ROI size when comparing results across ROIs (Eger et al., 2008; Erez & Yovel, 2014; Said et al., 2010; Walther et al., 2009). It was not clear whether this is the case for MD regions, which are different from visual regions in multiple respects, and whether more generally the choice of this parameter affects decoding levels. In our data (Figure 6),

we observed slightly lower levels of classification for the smaller ROI sizes (50 and 100 voxels), but these stabilized for ROI sizes of 150 voxels or more, in line with previous reports (Erez & Duncan, 2015; Shashidhara & Erez, 2019). This does not necessarily mean that such high dimensionality is required to reach maximal decodability without over-fitting, an issue that other studies have looked into more formally (Ahlheim & Love, 2018). Importantly, controlling for ROI size may be essential when comparing MD regions to each other as well as to other brain systems, such as visual areas, further emphasizing the importance of using a method that enables such control.

Different variations of the individual ROI approach that we used here can be designed for future studies, offering a balance between the need of a consistent definition of regions across participants, and perhaps studies, and the localization at the individual participant level. Such variations can be designed depending on the research question, and can be used with both univariate and multivariate analyses while avoiding double-dipping (Nikolaus Kriegeskorte, Kyle Simmons, F Bellgowan, & Baker, 2009). For example, data from two or more localizer tasks can be combined, as we demonstrated here. Combining data across localizers could lead to capturing core parts of the MD system, thus reflecting the multiple-demand nature of the selected voxels (Assem et al., 2019; John Duncan, 2010, 2013; E. Fedorenko et al., 2013). An even more cognitively diverse variation can be a localizer that consists of multiple tasks within the same run with a similar manipulation of difficulty level. On the other end of this scale, a localizer task can be designed to target a specific cognitive aspect of interest, and constraining activation patterns by a group template will ensure that the areas of interest are within the boundaries of the MD network.

In summary, we used three independent localizer tasks to define subject-specific ROIs and test the effect of using the individual ROIs compared to a group template on univariate and multivariate effects in a rule-based criterion task. The univariate results in the criterion task greatly benefitted from using individual ROIs compared to using the group template. In contrast, multivoxel task-related representations did not vary with localizers were similar for the subject-specific ROIs and the group template, and for all localizers, with no benefit of increased pattern discriminability, as well as no cost of reduced discriminability. The group-constrained individually-defined ROIs offer a refined and targeted localization of each participant's MD regions based on the individual's unique functional

pattern of activity, while ensuring that similar brain regions are studied in all participants. Pushing forward towards standardization in the field, our study provides important empirical evidence for researchers using both univariate and multivariate analysis of fMRI data to study the functional organization of the MD network.

Acknowledgements

The authors declare no competing financial interests.

5. References

Ahlheim, C., & Love, B. C. (2018). Estimating the functional dimensionality of neural representations. *NeuroImage*, *179*, 51–62. https://doi.org/10.1016/J.NEUROIMAGE.2018.06.015

Assem, M., Glasser, M. F., Essen, D. C. Van, & Duncan, J. (2019). A Domain-general Cognitive Core defined in Multimodally Parcellated Human Cortex. *BioRxiv*, 517599. Retrieved from https://www.biorxiv.org/content/early/2019/01/10/517599

Berman, M. G., Park, J., Gonzalez, R., Polk, T. A., Gehrke, A., Knaffla, S., & Jonides, J. (2010). Evaluating functional localizers: The case of the FFA. *NeuroImage*, *50*(1), 56–71. https://doi.org/10.1016/j.neuroimage.2009.12.024

Bhandari, A., Gagne, C., & Badre, D. (2018). Just above Chance: Is It Harder to Decode Information from Prefrontal Cortex Hemodynamic Activity Patterns? *Journal of Cognitive Neuroscience*, *30*(10), 1473–1498. https://doi.org/10.1162/jocn_a_01291

Blank, I., & Fedorenko, E. (2017). Domain-general brain regions do not track linguistic input as closely as language-selective regions. *The Journal of Neuroscience*, *37*(41), 9999–10011. https://doi.org/10.1523/JNEUROSCI.3642-16.2017

Blank, I., Kanwisher, N., & Fedorenko, E. (2014). A functional dissociation between language and multiple-demand systems revealed in patterns of BOLD signal fluctuations. *Journal of Neurophysiology*, *112*(5), 1105–1118. https://doi.org/10.1152/jn.00884.2013

Brainard, D. H. (1997). The Psychophysics Toolbox. *Spatial Vision*, *10*(4), 433–436. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/9176952

Brett, M., Anton, J.-L., Valabregue, R., & Poline, J.-B. (2002). Region of interest analysis using an SPM toolbox. *NeuroImage*, *16*. Retrieved from http://matthew.dynevor.org/research/abstracts/marsbar/marsbar_abstract.pdf

Brett, M., Johnsrude, I. S., & Owen, A. M. (2002). The problem of functional localization in the human brain. *Nature Reviews Neuroscience*, *3*(3), 243–249. https://doi.org/10.1038/nrn756

Carlin, J. D., & Kriegeskorte, N. (2017). Adjudicating between face-coding models with individual-face fMRI responses. *PLoS Computational Biology*, *13*(7), e1005604.

https://doi.org/10.1371/journal.pcbi.1005604

Cole, M. W., Ito, T., & Braver, T. S. (2016). The Behavioral Relevance of Task Information in Human Prefrontal Cortex. *Cerebral Cortex*, *26*(6), 2497–2505. https://doi.org/10.1093/cercor/bhv072

Cole, M. W., Reynolds, J. R., Power, J. D., Repovs, G., Anticevic, A., & Braver, T. S. (2013). Multi-task connectivity reveals flexible hubs for adaptive task control. *Nature Neuroscience*, *16*(9), 1348–1355. https://doi.org/10.1038/nn.3470

Cole, M. W., & Schneider, W. (2007). The cognitive control network: Integrated cortical regions with dissociable functions. *NeuroImage*, *37*(1), 343–360. https://doi.org/10.1016/j.neuroimage.2007.03.071

Crittenden, B. M., Mitchell, D. J., & Duncan, J. (2015). Recruitment of the default mode network during a demanding act of executive control. *ELife*, *4*, e06481. https://doi.org/10.7554/eLife.06481

Crittenden, B. M., Mitchell, D. J., & Duncan, J. (2016). Task Encoding across the Multiple Demand Cortex Is Consistent with a Frontoparietal and Cingulo-Opercular Dual Networks Distinction. *Journal of Neuroscience*, *36*(23), 6147–6155. https://doi.org/10.1523/JNEUROSCI.4590-15.2016

Curtis, C. E., Cole, M. W., Rao, V. Y., & D'Esposito, M. (2005). Canceling Planned Action: An fMRI Study of Countermanding Saccades. *Cerebral Cortex*, *15*(9), 1281–1289. https://doi.org/10.1093/cercor/bhi011

Cusack, R., Vicente-Grabovetsky, A., Mitchell, D. J., Wild, C. J., Auer, T., Linke, A. C., & Peelle, J. E. (2014). Automatic analysis (aa): efficient neuroimaging workflows and parallel processing using Matlab and XML. *Frontiers in Neuroinformatics*, *8*, 90. https://doi.org/10.3389/fninf.2014.00090

Desimone, R., & Duncan, J. (1995). Neural Mechanisms of Selective Visual Attention. *Annual Review of Neuroscience*, *18*(1), 193–222. https://doi.org/10.1146/annurev.ne.18.030195.001205

Dosenbach, N. U. F., Fair, D. A., Miezin, F. M., Cohen, A. L., Wenger, K. K., Dosenbach, R. A. T., … Petersen, S. E. (2007). Distinct brain networks for adaptive and stable task control in humans. *Proc. NatL Acad. Sci. USA*, *104*(26), 11073–11078. https://doi.org/10.1073/PNAS.0704320104

Dove, A., Pollmann, S., Schubert, T., Wiggins, C. J., & Yves von Cramon, D. (2000). Prefrontal cortex

activation in task switching: an event-related fMRI study. *Cognitive Brain Research*, *9*(1), 103–109. https://doi.org/10.1016/S0926-6410(99)00029-4

Dubois, J., de Berker, A. O., & Tsao, D. Y. (2015). Single-unit recordings in the macaque face patch system reveal limitations of fMRI MVPA. *Journal of Neuroscience*, *35*(6), 2791–2802. https://doi.org/10.1523/JNEUROSCI.4037-14.2015

Duncan, J. (2006). EPS mid-career award 2004: Brain mechanisms of attention. *Quarterly Journal of Experimental Psychology*, *59*(1), 2–27. https://doi.org/10.1080/17470210500260674

Duncan, J. (2010). The multiple-demand (MD) system of the primate brain: mental programs for intelligent behaviour. *Trends in Cognitive Sciences*, *14*(4), 172–179. https://doi.org/10.1016/j.tics.2010.01.004

Duncan, J. (2013). The Structure of Cognition: Attentional Episodes in Mind and Brain. *Neuron*, *80*(1), 35–50. https://doi.org/10.1016/j.neuron.2013.09.015

Duncan, J., & Owen, A. M. (2000). Common regions of the human frontal lobe recruited by diverse cognitive demands. *Trends in Neurosciences*, *23*(10), 475–483. https://doi.org/10.1016/S0166-2236(00)01633-7

Eger, E., Ashburner, J., Haynes, J., Dolan, R. J., & Rees, G. (2008). Functional Magnetic Resonance Imaging Activity Patterns in Human Lateral Occipital Complex Carry Information about Object Exemplars within Category. *Journal of Cognitive Neuroscience*, *20*(2), 356–370.

Erez, Y., & Duncan, J. (2015). Discrimination of Visual Categories Based on Behavioral Relevance in Widespread Regions of Frontoparietal Cortex. *Journal of Neuroscience*, *35*(36), 12383–12393. https://doi.org/10.1523/JNEUROSCI.1134-15.2015

Erez, Y., & Yovel, G. (2014). Clutter Modulates the Representation of Target Objects in the Human Occipitotemporal Cortex. *Journal of Cognitive Neuroscience*, *26*(3), 490–500. https://doi.org/10.1162/jocn_a_00505

Ester, E. F., Sprague, T. C., & Serences, J. T. (2015). Parietal and Frontal Cortex Encode Stimulus-Specific Mnemonic Representations during Visual Working Memory. *Neuron*, *87*(4), 893–905. https://doi.org/10.1016/J.NEURON.2015.07.013

Etzel, J. A., Cole, M. W., Zacks, J. M., Kay, K. N., & Braver, T. S. (2016). Reward Motivation Enhances

Task Coding in Frontoparietal Cortex. *Cerebral Cortex*, *26*(4), 1647–1659. https://doi.org/10.1093/cercor/bhu327

Fedorenko, E., Duncan, J., & Kanwisher, N. (2013). Broad domain generality in focal regions of frontal and parietal cortex. *Proc. NatL Acad. Sci. USA*, *110*(41), 16616–16621. https://doi.org/10.1073/pnas.1315235110

Fedorenko, E., Hsieh, P.-J., Nieto-Castañón, A., Whitfield-Gabrieli, S., & Kanwisher, N. (2010). New method for fMRI investigations of language: defining ROIs functionally in individual subjects. *Journal of Neurophysiology*, *104*(2), 1177–1194. https://doi.org/10.1152/jn.00032.2010

Fedorenko, E., Nieto-Castañon, A., & Kanwisher, N. (2012). Lexical and syntactic representations in the brain: An fMRI investigation with multi-voxel pattern analyses. *Neuropsychologia*, *50*(4), 499–513. https://doi.org/10.1016/j.neuropsychologia.2011.09.014

Feinberg, D. A., Moeller, S., Smith, S. M., Auerbach, E., Ramanna, S., Gunther, M., … Yacoub, E. (2010). Multiplexed echo planar imaging for sub-second whole brain FMRI and fast diffusion imaging. *PloS One*, *5*(12), e15710. https://doi.org/10.1371/journal.pone.0015710

Fox, M. D., Snyder, A. Z., Barch, D. M., Gusnard, D. A., & Raichle, M. E. (2005). Transient BOLD responses at block transitions. *NeuroImage*, *28*(4), 956–966. https://doi.org/10.1016/J.NEUROIMAGE.2005.06.025

Fox, M. D., Snyder, A. Z., Vincent, J. L., Corbetta, M., Van Essen, D. C., & Raichle, M. E. (2005). The human brain is intrinsically organized into dynamic, anticorrelated functional networks. *Proc. NatL Acad. Sci. USA*, *5*, 9673–9678.

Garrison, K. A., Rogalsky, C., Sheng, T., Liu, B., Damasio, H., Winstein, C. J., & Aziz-Zadeh, L. S. (2015). Functional MRI preprocessing in lesioned brains: Manual versus automated region of interest analysis. *Frontiers in Neurology*, *6*, 196. https://doi.org/10.3389/fneur.2015.00196

Glasser, M. F., Coalson, T. S., Robinson, E. C., Hacker, C. D., Harwell, J., Yacoub, E., … Van Essen, D. C. (2016). A multi-modal parcellation of human cerebral cortex. *Nature*, *536*, 171–178. https://doi.org/10.1038/nature18933

Hampshire, A., Highfield, R. R., Parkin, B. L., & Owen, A. M. (2012). Fractionating Human Intelligence. *Neuron*, *76*(6), 1225–1237. https://doi.org/10.1016/j.neuron.2012.06.022

Haxby, J. V, Gobbini, M. I., Furey, M. L., Ishai, A., Schouten, J. L., & Pietrini, P. (2001). Distributed and Overlapping Representations of Faces and Objects in Ventral Temporal Cortex Downloaded from. *Science (New York, N.Y.)*, *293*(5539), 2425–2430. Retrieved from www.sciencemag.orghttp://science.sciencemag.org/

Haynes, J.-D., & Rees, G. (2006). Decoding mental states from brain activity in humans. *Nature Reviews Neuroscience*, *7*(7), 523–534. https://doi.org/10.1038/nrn1931

Hebart, M. N., Gorgen, 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

Julian, J. B., Fedorenko, E., Webster, J., & Kanwisher, N. (2012). An algorithmic method for functionally defining regions of interest in the ventral visual pathway. *NeuroImage*, *60*, 2357–2364. https://doi.org/10.1016/j.neuroimage.2012.02.055

Kanwisher, N., McDermott, J., & Chun, M. M. (1997). The fusiform face area: a module in human extrastriate cortex specialized for face perception. *The Journal of Neuroscience*, *17*(11), 4302–4311. https://doi.org/10.1098/Rstb.2006.1934

Kriegeskorte, N., Goebel, R., & Bandettini, P. (2006). Information-based functional brain mapping. *Proc. NatL Acad. Sci. USA*, *103*(10), 3863–3868. https://doi.org/10.1073/pnas.0600244103

Kriegeskorte, N., Kyle Simmons, W., F 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

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

Krishnan, S., Slavin, M. J., Tran, T.-T. T., Doraiswamy, P. M., & Petrella, J. R. (2006). Accuracy of spatial normalization of the hippocampus: Implications for fMRI research in memory disorders. *NeuroImage*, *31*, 560–571. https://doi.org/10.1016/j.neuroimage.2005.12.061

Lafer-Sousa, R., Conway, B. R., & Kanwisher, N. (2016). Color-Biased Regions of the Ventral Visual Pathway Lie between Face-and Place-Selective Regions in Humans, as in Macaques. *Journal of*

Neuroscience, *36*(5), 1682–1697. https://doi.org/10.1523/JNEUROSCI.3164-15.2016

Malach, R., Reppas, J. B., Benson, R. R., Kwong, K. K., Jiang, H., Kennedy, W. A., … Tootell, R. B. (1995). Object-related activity revealed by functional magnetic resonance imaging in human occipital cortex. *Proc. NatL Acad. Sci. USA*, *92*(18), 8135–8139. https://doi.org/10.1073/PNAS.92.18.8135

Mineroff, Z., Blank, I., Mahowald, K., & Fedorenko, E. (2018). A robust dissociation among the language, multiple demand, and default mode networks: Evidence from inter-region correlations in effect size. *Neuropsychologia*, *119*, 501–511. https://doi.org/10.1016/j.neuropsychologia.2018.09.011

Mitchell, D. J., & Cusack, R. (2018). Visual short-term memory through the lifespan: Preserved benefits of context and metacognition. *Psychology and Aging*, *33*(5), 841–854. https://doi.org/10.1037/pag0000265

Muhle-Karbe, P. S., Duncan, J., De Baene, W., Mitchell, D. J., & Brass, M. (2017). Neural Coding for Instruction-Based Task Sets in Human Frontoparietal and Visual Cortex. *Cerebral Cortex*, *27*(3), 1891–1905. https://doi.org/10.1093/cercor/bhw032

Nee, D. E., & Brown, J. W. (2012). Rostral-caudal gradients of abstraction revealed by multi-variate pattern analysis of working memory. *NeuroImage*, *63*(3), 1285–1294. https://doi.org/10.1016/j.neuroimage.2012.08.034

Nelissen, N., Stokes, M. G., Nobre, A. C., & Rushworth, M. F. S. (2013). Frontal and parietal cortical interactions with distributed visual representations during selective attention and action selection. *Journal of Neuroscience*, *33*(42), 16443–16458. https://doi.org/10.1523/JNEUROSCI.2625-13.2013

Nieto-Castañón, A., & Fedorenko, E. (2012). Subject-specific functional localizers increase sensitivity and functional resolution of multi-subject analyses. *NeuroImage*, *63*(3), 1646–1669. https://doi.org/10.1016/j.neuroimage.2012.06.065

Nili, H., Wingfield, C., Walther, A., Su, L., Marslen-Wilson, W., & Kriegeskorte, N. (2014). A Toolbox for Representational Similarity Analysis. *PLoS Computational Biology*, *10*(4), e1003553. https://doi.org/10.1371/journal.pcbi.1003553

Nomura, E. M., Gratton, C., Visser, R. M., Kayser, A., Perez, F., & D'Esposito, M. (2010). Double dissociation of two cognitive control networks in patients with focal brain lesions. *Proc. NatL Acad. Sci. USA*, *107*(26), 12017–12022. https://doi.org/10.1073/pnas.1002431107

Paunov, A. M., Blank, I. A., & Fedorenko, E. (2019). Functionally distinct language and Theory of Mind networks are synchronized at rest and during language comprehension. *J Neurophysiol*, *121*, 1244–1265. https://doi.org/10.1152/jn.00619.2018

Reddy, L., & Kanwisher, N. (2007). Category Selectivity in the Ventral Visual Pathway Confers Robustness to Clutter and Diverted Attention. *Current Biology*, *17*(23), 2067–2072. https://doi.org/10.1016/j.cub.2007.10.043

Rouder, J. N., Speckman, P. L., Sun, D., Morey, R. D., & Iverson, G. (2009). Bayesian t tests for accepting and rejecting the null hypothesis. *Psychonomic Bulletin & Review*, *16*(2), 225–237. https://doi.org/10.3758/PBR.16.2.225

Said, C. P., Moore, C. D., Engell, A. D., Todorov, A., & Haxby, J. V. (2010). Distributed representations of dynamic facial expressions in the superior temporal sulcus. *Journal of Vision*, *10*(5), 1–12. https://doi.org/10.1167/10.5.11

Saxe, R., Brett, M., & Kanwisher, N. (2006). Divide and conquer: A defense of functional localizers. *NeuroImage*, *30*(4), 1088–1096. https://doi.org/10.1016/J.NEUROIMAGE.2005.12.062

Schaefer, A., Kong, R., Gordon, E. M., Laumann, T. O., Zuo, X.-N., Holmes, A. J., … Yeo, T. B. T. (2018). Local-Global Parcellation of the Human Cerebral Cortex from Intrinsic Functional Connectivity MRI. *Cerebral Cortex*, *28*(9), 3095–3114. https://doi.org/10.1093/cercor/bhx179

Seeley, W. W., Menon, V., Schatzberg, A. F., Keller, J., Glover, G. H., Kenna, H., … Greicius, M. D. (2007). Dissociable Intrinsic Connectivity Networks for Salience Processing and Executive Control. *Journal of Neuroscience*, *27*(9), 2349–2356. https://doi.org/10.1523/JNEUROSCI.5587-06.2007

Shashidhara, S., & Erez, Y. (2019). Reward motivation modulates representation of behaviorally-relevant information across the frontoparietal cortex. *BioRxiv*, 609537. https://doi.org/10.1101/609537

Shashidhara, S., Mitchell, D. J., Erez, Y., & Duncan, J. (2019). Progressive Recruitment of the

Frontoparietal Multiple-demand System with Increased Task Complexity, Time Pressure, and Reward. *Journal of Cognitive Neuroscience*, *31*(11), 1617–1630. https://doi.org/10.1162/jocn_a_01440

Smith, V., Mitchell, D. J., & Duncan, J. (2018). Role of the Default Mode Network in Cognitive Transitions. *Cerebral Cortex*, *28*(10), 3685–3696. https://doi.org/10.1093/cercor/bhy167

Stiers, P., Mennes, M., & Sunaert, S. (2010). Distributed task coding throughout the multiple demand network of the human frontal-insular cortex. *NeuroImage*, *52*, 252–262. https://doi.org/10.1016/j.neuroimage.2010.03.078

Tibon, R., Greve, A., & Henson, R. N. (2018). The missing link? Testing a schema account of unitization. *Memory & Cognition*, *46*(7), 1023–1040. https://doi.org/10.3758/s13421-018-0819-3

van der Kouwe, A. J. W., Benner, T., Salat, D. H., & Fischl, B. (2008). Brain morphometry with multiecho MPRAGE. *NeuroImage*, *40*(2), 559–569. https://doi.org/10.1016/j.neuroimage.2007.12.025

Walther, D. B., Caddigan, E., Fei-Fei, L., & Beck, D. M. (2009). Natural Scene Categories Revealed in Distributed Patterns of Activity in the Human Brain. *Journal of Neuroscience*, *29*(34), 10573–10581. https://doi.org/10.1523/JNEUROSCI.0559-09.2009

Weiner, K. S., Barnett, M. A., Witthoft, N., Golarai, G., Stigliani, A., Kay, K. N., … Grill-Spector, K. (2018). Defining the most probable location of the parahippocampal place area using cortex-based alignment and cross-validation. *NeuroImage*, *170*, 373–384. https://doi.org/10.1016/j.neuroimage.2017.04.040

Wisniewski, D., Goschke, T., & Haynes, J.-D. (2016). Similar coding of freely chosen and externally cued intentions in a fronto-parietal network. *NeuroImage*, *134*, 450–458. https://doi.org/10.1016/J.NEUROIMAGE.2016.04.044

Woolgar, A., Hampshire, A., Thompson, R., & Duncan, J. (2011). Adaptive coding of task-relevant information in human frontoparietal cortex. *Journal of Neuroscience*, *31*(41), 14592–14599. https://doi.org/10.1523/JNEUROSCI.2616-11.2011

Woolgar, A., Thompson, R., Bor, D., & Duncan, J. (2011). Multi-voxel coding of stimuli, rules, and responses in human frontoparietal cortex. *NeuroImage*, *56*(2), 744–752.

https://doi.org/10.1016/J.NEUROIMAGE.2010.04.035

Yeo, T. B. T., Krienen, F. M., Sepulcre, J., Sabuncu, M. R., Lashkari, D., Hollinshead, M., … Buckner, R. L. (2011). The organization of the human cerebral cortex estimated by intrinsic functional connectivity. *Journal of Neurophysiology*, *106*(3), 1125–1165. https://doi.org/10.1152/jn.00338.2011