Similar coding of freely chosen and externally cued intentions in a fronto-parietal network

Abbreviated title: Similar coding of free and cued intentions

David Wisniewski [1,2], Thomas Goschke [2], John-Dylan Haynes [1,3]

1 Bernstein Center for Computational Neuroscience Berlin, and Berlin Center for Advanced Neuroimaging, and Cluster of Excellence NeuroCure, and Department of Neurology, Charité – Universitätsmedizin Berlin, Germany

2 Department of Psychology and Neuroimaging Center, and Collaborative Research Center 940 "Volition and Cognitive Control", Technische Universität Dresden, Germany

3 Berlin School of Mind and Brain, and Department of Psychology, Humboldt-Universität zu Berlin, Germany

Correspondence should be addressed to David Wisniewski, Charité – Universitätsmedizin Berlin, Bernstein Center for Computational Neuroscience, Haus 6, Philippstrasse 13, 10115 Berlin, Germany, email: david.wisniewski@bccn-berlin.de

Conflict of interest: The authors declare no competing financial interest

ABSTRACT

Intentional action is essential to human behavior, yet its neural basis remains poorly understood. In order to identify neural networks specifically involved in intentional action, freely chosen and externally cued intentions have previously been contrasted. This has led to the identification of a fronto-parietal network, which is involved in freely choosing one's intentions. However, it remains unclear whether this network encodes specific intentions, or whether it merely reflects general preparatory or control processes correlated with intentional action. Here, we used MVPA on fMRI data to identify brain regions encoding non-motor intentions that were either freely chosen or externally cued. We found that a fronto-parietal network, including the lateral prefrontal cortex, premotor, and parietal cortex, contained information about both freely chosen and externally cued intentions. Importantly, MVPA cross-classification indicated that this network represents the content of our intentions similarly, regardless of whether these intentions are freely chosen or externally cued. This finding suggests that the intention network has a general role in processing and representing intentions independent of their origin.

KEYWORDS

INTRODUCTION

Intentional action is an essential part of everyday human behavior (Goschke, 2013; Haggard, 2008). In cognitive neuroscience, two types of intentions are often contrasted: internally vs. externally guided intentions (Beck et al., 2014; Brass et al., 2013; Cunnington et al., 2002; Forstmann et al., 2006; Gilbert et al., 2009; Jahanshahi et al., 1995; Mueller et al., 2007, for a recent meta-analysis see Rae et al., 2014). Internally guided or "free" intentions are generated in the absence of direct external trigger stimuli and result from an internal choice process. Externally guided or "cued" intentions in contrast are generated in direct response to external stimuli. Note that when we speak of free intentions, we do not mean to imply that these intentions are less causally determined, but that they their direct cause is not an external stimulus. In typical experiments, subjects can either freely choose which of two tasks to perform (Soon et al., 2013), or they are externally cued as to which of two tasks to perform (Monsell, 2003; Ruge et al., 2013). By comparing the neural representations of free and cued intentions it is possible to assess whether they are processed similarly in the brain or not (Deiber et al., 1991; Forstmann et al., 2006; Passingham et al., 2010; but see Nachev and Husain, 2010).

Previous research using functional magnetic resonance imaging (fMRI) suggests that different networks might be involved in processing free and cued intentions, with e.g. the dACC (Bengtsson et al., 2008; Forstmann et al., 2006; Mueller et al., 2007) or preSMA (Bengtsson et al., 2008; Rae et al., 2014) being more strongly activated while processing free intentions. However, regions found to be associated with free intentions have also been found to be associated with general preparatory processes (Fedorenko et al., 2013), such as conflict monitoring (Botvinick et al., 2001). Whenever intentions are freely chosen, two comparable alternatives need to be maintained and conflict between them resolved (Brass et al., 2013). This conflict resolution might explain increased activation in the dACC as well, although in contrast to other conflict paradigms (Eriksen, 1995; Stürmer et al., 2002) there is no "correct" or "incorrect" response for free choices. Furthermore, working memory demands might also play a role (Lau et al., 2004a).

3

A solution to ruling out unspecific processes when comparing free and cued intentions, is to use multivoxel pattern analysis (MVPA; Haynes and Rees, 2006; Haynes, 2015; Kriegeskorte et al., 2006). Here, spatial activation patterns which encode the content of specific free (or cued) intentions can be identified.

Recent studies identified a fronto-parietal network which encoded cued intentions, including the anterior medial PFC (Gilbert, 2011), lateral PFC, and parietal cortex (Bode and Haynes, 2009; Momennejad and Haynes, 2013; Wisniewski et al. 2015). A partly overlapping brain network was found to encode freely chosen intentions, including the dACC (Wisniewski et al., 2014), frontopolar cortex, precuneus (Soon et al., 2008), medial and lateral PFC (Haynes et al., 2007). However, it is still an open question whether free and cued intentions are represented similarly in the brain, although the overlapping networks seem to suggest this. Within a single region the same task might nonetheless have a different neural representation depending on whether it is free or cued. Thus, in order to identify regions where neural task representations under free and cued conditions are similar it is essential to directly compare the specific patterns coding individual intentions under both cueing and free conditions. Furthermore, comparing results from previous studies is complicated by the different tasks, stimuli and designs used. One recent study, which compared free and cued intentions within-subjects using the same task (Zhang et al., 2013), demonstrated that a fronto-parietal intention network represents free and cued intentions. Different parts of this network were functionally specialized, with the premotor and parietal cortex representing both free and cued intentions, whereas the lateral PFC only represented free intentions. However, the tasks in this experiment were based on *attending* to different features of one stimulus, making it difficult to exclude attentional confounds in task coding. Moreover, it remains unclear whether results generalize to more abstract intentions, which are not specified in terms of direct stimulus-response-mappings between perceptual features and motor responses, but rather refer to more abstract cognitive tasks such as mental calculation as used here (see Discussion for details).

Here, we report an experiment in which we directly compared the representations of free and cued intentions by applying MVPA to fMRI data in a mental calculation task. This research was motivated by two central questions: First, is the fronto-parietal intention network also involved in representing abstract, non-motor intentions especially when differences in feature-based attention can be ruled out? Second, does this network represent freely chosen and externally cued intentions similarly, suggesting a general role for the fronto-parietal network in intention processing? Or does this network exhibit functional specialization with respect to freely chosen vs. externally cued intentions?

MATERIALS AND METHODS

Participants

35 participants took part in the experiment (24 females). All subjects volunteered to participate and had normal or corrected-to-normal vision. Subjects gave written informed consent and received 25€ for participation. The experiment was approved by the local ethics committee. Subjects were right-handed, and no subject had a history of neurological or psychiatric disorders. Four subjects showed excessive head movements inside the MR scanner (>5mm) and were excluded from further fMRI analyses.

Experimental Paradigm

The experiment was implemented using Matlab Version 8.1.0 (The MathWorks) and the Cogent Toolbox (http://www.vislab.ucl.ac.uk/cogent.php). Trials started with the presentation of a single visual cue centrally on screen (Figure 1). In half of the trials this cue specified one of two possible calculation task to be performed (adding or subtracting). This was the *cued* condition. In the other half of the trials, subjects were given the free choice between addition and subtraction (*free* condition). In half of the *cued* trials, subjects were cued to add, in the other half they were cued to subtract. In *free* trials, subjects received no explicit instruction on which task to perform in each trial.
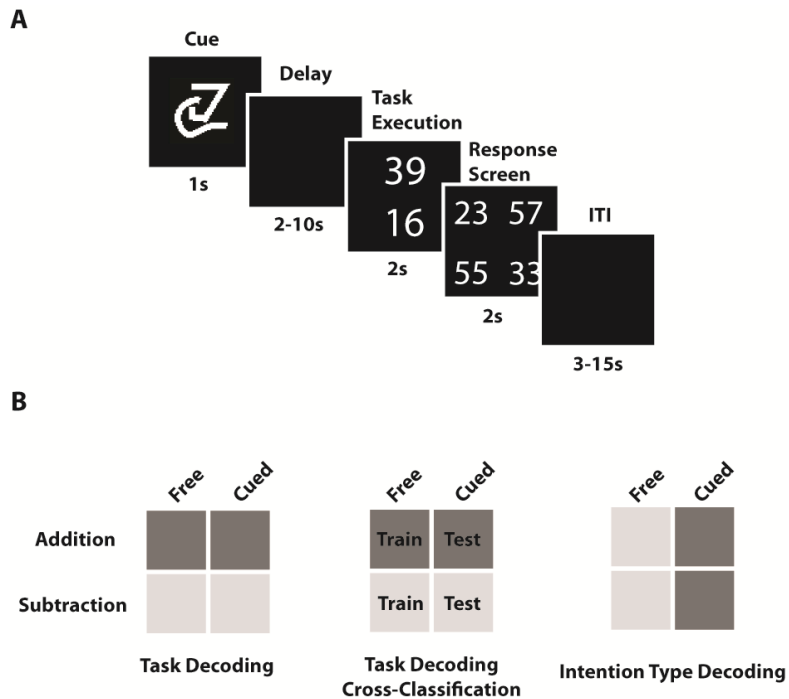
Figure 1: A. Trial structure. At the beginning of each trial, subjects were presented with a cue indicating that they could either choose freely between the two tasks (free trials), or indicating which of the two tasks to perform (cued trials). After a variable delay, subjects were presented with two numbers on the screen and either added or subtracted them, depending on the current trial condition. The response screen was used to indicate the correct response, and in free trials the task performed was determined by subjects' responses. Trials were separated by a variable inter-trial-interval (ITI). **B. Analyses**. The three main analyses are presented. In the task decoding, addition (dark grey) and subtraction (light grey) trials were contrasted. In the task decoding cross-classification, a classifier was trained to distinguish addition and subtraction trials in the free condition only. Classifier performance was then tested on cued trials. Only brain regions in which tasks are represented similarly in free and cued trials will show above chance decoding accuracies. In the intention type decoding, free trials (light grey) and cued trials (dark grey) were contrasted.

The tasks (addition and subtraction) and intention types (free and cued) were orthogonalized in the experiment. The cues used were abstract line drawings that were designed to minimize a priori semantic associations to the subjects before the start of the experiment (see Wisniewski et al. 2014; Reverberi et al. 2012 for a similar approach). Furthermore, two cues each were associated with the free, cued addition, and cued subtraction conditions, respectively. This was done in order to allow dissociating the neural representation of the task choice from the visual identity of the cues (see below for details). Which of the two semantically identical cues was presented was pseudo-

randomized across trials. The cue was presented visually for 1000ms, after which a variable delay followed (between 2000 and 10000ms, mean duration 6000ms, distributed uniformly in steps of 2000ms). This delay allowed to dissociate in time the cue presentation and intention maintenance phase from the task execution, which followed after the delay. It also made the time of the task screen onset unpredictable to the subjects, forcing them to maintain a task representation throughout the whole delay period (Haynes et al., 2007). Two numbers were presented on the task screen, one above and the other below the fixation cross. Depending on the current condition subjects either added or subtracted these two numbers. The numbers were randomly chosen in each trial, from a set of numbers between 11 and 59. Integer multiples of 10 were excluded, as calculations with these numbers would be too easy. The numbers were presented for 2000ms. Then, a response screen was presented, giving subjects four different response options: the correct response for addition, the correct response for subtraction and two wrong responses. Response options were presented on four fixed positions on screen (Figure 1), which were mapped onto four buttons which subjects operated using their left and right index and middle fingers. The response-mapping was pseudo-randomized in each trial, dissociating task execution from motor preparation processes. (This is in contrast with Zhang et al. 2013, where response-mappings were fixed for each subject, making a dissociation of task execution and motor preparation processes more difficult). The response screen was presented for 2000ms, irrespective of the actual reaction time (RT) in each trial. Note, that in free trials the chosen task was inferred from the responses. If for instance subjects chose the correct response for addition, we treated this trial as an addition trial. After the response screen, a variable inter-trial-interval was presented (mean duration 5000ms, between 3000ms and 15000ms). Inter-trial-interval durations were distributed geometrically, to de-correlate individual trials in time.

Overall, each subject performed 240 trials, divided into five runs of 48 trials each. The whole experimental session lasted around 90min on average. Subjects also underwent a one hour training session one to four days before the fMRI session. They learned the meaning of all abstract cues, and

familiarized themselves with the numerical tasks. This was done to minimize training effects during the fMRI session. At the end of the training session, subjects performed three runs of 48 trials each of the task, using the same parameters as would be used later in the fMRI session. In the training session, they were asked to choose addition and subtraction equally often across the whole experiment. Large choice biases severely complicate MVPA analyses, and we therefore screened subjects' training data to rule out subjects that chose one task in more than 60% of the trials. None of our subjects showed a choice bias according to this criterion, and thus all pilot subjects were included in the fMRI scanning.

Image Acquisition

Functional imaging was conducted on a 3-Tesla Siemens Trio (Erlangen, Germany) scanner with a 12-channel head coil. In each run, 417 T2*-weighted echo-planar images (EPI) were acquired (TR = 2000ms, TE = 30ms, flip angle 90°). Each volume consisted of 33 slices, separated by a gap of 0.6mm. Matrix size was 64 x 64, and field of view (FOV) was 192mm, which resulted in a voxel size of 3 x 3 x 3mm. The first three images of each run were discarded.

Data Analysis: Behavior

Behavioral data were analyzed using Matlab (Version 8.1.0). For each subject, task performance was assessed by calculating the mean RT and mean error rate across all runs. For all further analyses, error trials were discarded. We further quantified how often subjects chose addition and subtraction in *free* trials. We expected both values to be close to 50% (Allefeld et al., 2013). We also tested whether RTs and error rates differed between tasks, intention types (free or cued), and visual cues. We did not expect any reliable differences between these conditions.

Data Analysis: fMRI

Functional data analysis was performed using SPM8 (http://www.fil.ion.ucl.ac.uk/spm), unless stated otherwise. We first unwarped, realigned, and slice time corrected all volumes. Data were also screened for possible scanner artifacts, using the Artifact Detection Tool (http://www.nitrc.org/projects/artifact_detect/, ART version 2011-07). In three subjects, some volumes (between one and five) were marked as showing spiking artifacts. Variance that was due to these artifacts was removed in the affected subjects by explicitly modeling the artifact in all first level general linear model (GLM) analyses (as suggested in the ART documentation). Preprocessed data were entered into a GLM (Friston et al., 1994), after which multivoxel pattern analysis (MVPA, Cox and Savoy, 2003; Haynes and Rees, 2006; Haynes, 2015; Haxby, 2012; Kriegeskorte et al., 2006; Tong and Pratte, 2012) was performed. We investigated which brain regions coded for the tasks during the delay period and whether the codes used for freely chosen and externally cued tasks were similar. We also tested which brain regions encoded the intention type (freely chosen, externally cued) and whether the codes used were similar for the two tasks. Full details on the analyses are presented below.

Multivariate decoding I – Tasks: In order to investigate the neural encoding of task-sets, a GLM was estimated for each subject. For each of the 5 runs we estimated regressors for addition trials and subtraction trials, collapsed across freely chosen and cued intentions. All regressors were locked to the onset of the cue presentation, while the duration was set to cover the whole delay period, encompassing the whole maintenance phase during which subjects had to represent their intentions, without yet implementing them. The task execution period was not explicitly modeled. Given that the delay period duration was variable, modeled signals were dissociated from task execution processes. We also added the current reaction time (RT) as a parametric modulation. This way, we used RTs to explicitly model any potential differences in task difficulty between addition and subtraction trials in order to minimize potential task difficulty effects. If not explicitly modelled, task

difficulty might partly explain task encoding effects in MVPA (Todd et al., 2013), although in other cases difficulty did not play a large role in explaining task decoding effects (Wisniewski et al., 2014; Woolgar et al., 2014). We decided to conservatively control for this potential confounding variable. We further added estimated movement parameters as separate regressors to the model. Brain responses were modeled using the SPM canonical haemodynamic response basis function.

In the next step, we performed a decoding analysis on the parameter estimates of the GLM (Figure 1B). A support vector classifier (SVC) was applied (Cox and Savoy, 2003; Haynes and Rees, 2006; Kamitani and Tong, 2005; Mitchell et al., 2004), implemented in *The Decoding Toolbox* (Hebart et al. 2014). Using a fixed regularization parameter (C=1), we performed a searchlight decoding analysis (Haynes et al., 2007; Kriegeskorte et al., 2006) which makes no a priori assumptions about informative brain areas. Around each measured voxel a sphere with a radius of three voxels was defined in the acquired volumes. For both conditions (addition and subtraction), we extracted the parameter estimates for each of the N voxels in a given sphere, resulting in an N-dimensional pattern vector. This was done for each run independently. Pattern vectors from four of the five runs (training dataset) were used to train the SVC to discriminate between the two task patterns. Classifier performance was tested on the remaining independent run (test dataset). This procedure was repeated five times, with each run being the test dataset once (leave-one-run-out-cross-validation). This resulted in a five-fold cross-validation of the classifier performance, which is necessary to control for potential problems of overfitting. The mean prediction accuracy was calculated across all five cross-validation steps and assigned to the central voxel of the sphere. SVC was repeated for each voxel in the acquired brain volume, resulting in a 3-D accuracy map for each subject. These accuracy maps were normalized to a standard brain (Montreal Neurological Institute (MNI) EPI template as implemented in SPM8) and resampled to an isotropic resolution of 3x3x3mm. The images were then smoothed with a Gaussian kernel (FWHM = 6mm) in order to account for potential differences in localization across subjects. The group analysis was performed using a random effects model on the accuracy maps, which were statistically tested using voxelwise t-tests against chance level. Given that

the SVC was performed on two conditions, chance level was 50%. We applied a statistical threshold of p < 0.05 (FWE corrected at the cluster level, initial voxel threshold p < 0.005).

Multivariate decoding II – Cross-classification of tasks: We then performed a second searchlight decoding analysis, testing whether tasks were represented similarly in the free and cued conditions (or "intention types", Figure 1B). In this analysis, we first estimated a GLM with the following regressors: freely chosen addition trials, freely chosen subtraction trials, externally cued addition trials, and externally cued subtraction trials. Again, RTs were added as a parametric modulator to account for potential differences in task difficulty. We also added the estimated movement parameters as separate regressors to the model. Regressors were locked to the cue onset and duration was set to cover the whole delay. We then trained a first SVC to discriminate between freely chosen addition and freely chosen subtraction trials. Classifier performance was then tested on externally cued addition and subtraction trials. A second SVC was trained on externally cued trials and tested on freely chosen trials. This resulted in two accuracy maps per subject. Note that only brain regions that encoded tasks similarly in freely chosen and externally cued trials would show above chance accuracies in this analysis. Furthermore, training on one set of trials and testing on an independent set of trials counteracted potential problems of overfitting. The *cross-classification* approach used in this analysis can be used to identify brain regions in which activation patterns are invariant with respect to the free and cued conditions (see Kaplan et al., 2015 for more details). The group analysis was performed using a voxelwise one-factorial ANOVA (factor: training dataset, i.e. train on freely chosen trials vs. train on externally cued trials). A t-contrast was computed, testing which brain regions showed above chance (50%) accuracies for both training datasets. Please note, that this analysis has less power compared to the task decoding analysis described above. This is due to the fact that only half of the trials were used to estimate the regressors used for training the classifier (i.e. only freely chosen or externally cued trials). We applied a statistical threshold of p < 0.05 (FWE corrected at the cluster level, initial threshold p < 0.005).

Multivariate decoding III – intention types: We also investigated the neural coding of intention types, i.e., whether subjects freely chose a task or whether it was externally cued (Figure 1B). We first estimated a GLM for each subject, using the following regressors in each run: freely chosen trials and externally cued trials. Reaction times were added as a parametric modulator in order to control for potential differences in difficulty between the conditions. Movement parameters were added as separate regressors. Regressors were locked to the cue onset and duration was set to the whole delay period of the trial. A searchlight decoding analysis similar to the task decoding was applied to the parameter estimates, identifying brain regions that code for intention type within each subject. The decoding analysis was performed on two conditions, resulting in a chance level of 50%. Resulting accuracy maps were normalized and smoothed and entered into a random effects group level analysis. A t-test was used to assess the statistical significance of the decoding results, using a statistical threshold of $p < 0.05$ (FWE corrected at the voxel level, minimal cluster size 20 voxels). The threshold was chosen more conservatively, as we expected a strong univariate signal difference between freely chosen and externally cued trials in this analysis (Forstmann et al. 2006; Mueller et al., 2007).

Control analysis - multivariate decoding of visual cues: As stated above, we used two different visual cues in free, cued addition, and cued subtraction trials. This was done in order to dissociate neural responses to tasks and intention types from those to the visual features of the cues. We therefore ran an additional control analysis, directly assessing which brain regions represented the visual features of the cues. We started by estimating a GLM with the following regressors for each subject: cued addition trials with visual cue 1, cued addition trials with visual cue 2. We added RTs as a parametric modulator and entered movement parameters as separate regressors to the model. Regressors were locked to the cue onset and duration was set to the whole delay phase in the trial, in order to make this analysis comparable to the task and intention type decoding analyses. We then

conducted a searchlight decoding analysis, in which a SVC was trained to discriminate between the two visual cues used in cued addition trials. Note that the visual features of the cues were the only difference between the conditions contrasted here. Trials were identical in all other respects. The same analysis was repeated for cued subtraction trials, free addition trials, and free subtraction trials, resulting in four accuracy maps per subject. Accuracy maps were normalized and smoothed and assessed at the group level using a one factorial ANOVA (four analyses). Given that the classification was performed on two conditions, chance level was 50%. Results were thresholded at $p < 0.05$ (FWE corrected at the cluster level, initial threshold $p < 0.005$). The overlap with task and intention type decoding analyses was assessed. This approach was chosen in order to mimic task and intention type analyses as closely as possible. For instance, running a six-class decoding analysis, with each cue being assigned to one class would have changed the chance level to 16%, reducing our ability to directly compare these results with the task and intention type decoding analyses.

RESULTS

Behavioral Results

The mean reaction time (RT) across all trials was 960ms (SE = 28ms). The mean error rate across all trials was 5.0% (SE = 0.5%). This includes cued trials in which a wrong button was pressed, free trials in which a button was pressed that was wrong for both tasks, and trials in which no button was pressed. In free trials, subjects chose addition in 51.2% (SE = 9.2%) of the trials and subtraction in 48.8% (SE = 8.8%) of the trials, the difference not being significant, $t(30) = 1.37$, $p = 0.18$. Furthermore, subjects rarely chose a response option that was wrong for both tasks in free trials (3.4%, SE = 0.4%). This indicates that subjects indeed chose one of the two available tasks and did not merely pick responses randomly. RTs did not differ significantly between the visual cues, $t(30) = 0.24$, $p = 0.81$, and intention types (freely chosen vs. externally cued), $t(30) = -0.74$, $p = 0.46$. RTs differed between tasks $t(30) = -4.5$, $p = 0.001$, indicating that subtraction ($RT_{sub} = 978$ms) was slightly more difficult than addition ($RT_{add} = 942$ms). Note that we corrected for this slight difficulty difference

13

between tasks in all fMRI analyses (see Methods above). Difficulty differences are thus unlikely to explain our neuroimaging results.

Task information

Distributed activation patterns in the left inferior parietal cortex and a large cluster encompassing the left inferior and medial frontal gyrus, and premotor cortex represented whether subjects intended to add or subtract the numbers presented on screen (Figure 2, $p < 0.05$ FWE corrected at the cluster level, initial voxel threshold $p < 0.005$, all results are also reported in Table 1). As a control analysis, we tested whether either addition or subtraction were associated with a univariate signal increase or decrease. We computed univariate contrast maps (addition < subtraction, and subtraction < addition), but found no brain region to show significant univariate signal differences between these two conditions ($p < 0.05$ FWE corrected at the cluster level, initial voxel threshold $p < 0.005$).
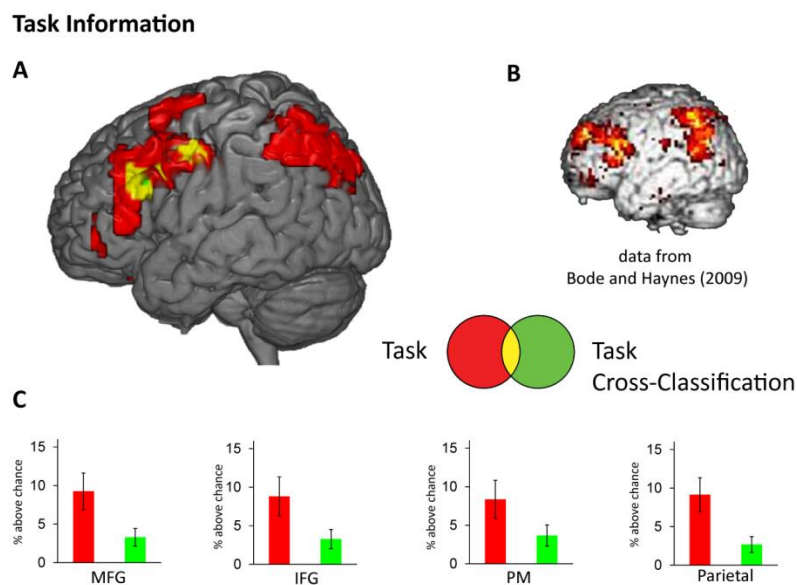
Figure 2. Task information. **A.** Regions in which local spatial activation patterns in the delay phase represented tasks are shown in red. Green regions represented tasks similarly in both freely chosen and externally cued trials, as assessed by cross-classifying freely chosen and externally cued tasks.

Yellow regions show the overlap of both analyses. **B.** Data from Bode & Haynes (2009). In their study, the authors found information about tasks in a highly similar network as in the current study. The image is included with permission of the authors. **C.** The plots depict region of interest analyses results for the parietal cortex and three sub-regions of the frontal cluster identified (inferior frontal gyrus IFG, medial frontal gyrus MFG, premotor cortex PM). The red bars represent decoding accuracies from the task decoding analysis. The green bars represent decoding accuracies from the task decoding using additional cross-classification across intention types. Given that all green bars are above chance, every region identified contains at least some neuronal populations that represent tasks similarly in freely chosen and externally cued trials. Error bars represent SEM.

We then used a cross-classification approach (Kaplan et al., 2015) to test which brain regions represented tasks similarly for both freely chosen and externally cued tasks. Results indicated that only the ventral premotor cortex and inferior frontal gyrus represented tasks similarly for both intention types (Figure 2, $p < 0.05$ FWE corrected at the cluster level, initial voxel threshold $p < 0.005$). At first glance this seems to be consistent with previously reported findings, which identified a highly similar task network and found similar coding only in parts of that network (Zhang et al. 2013). However, we further investigated whether tasks were encoded similarly across intention types using a more sensitive leave-one-subject-out region of interest (ROI) analysis (for more details on this analysis method see Estermann et al., 2010). We repeated the task decoding group level analysis (without cross-classification) excluding a single subject. Voxel coordinates from all significant regions identified in this analysis were extracted. In the remaining subject, we then extracted the mean decoding accuracy from the task decoding analysis with and without cross-classification. This procedure was repeated until every subject was left out. This ensured that the data used to define the ROI remained independent from the data used to statistically assess the accuracy values inside this ROI, which was done using t-tests against chance level across all subjects (Bonferroni corrected for the number of tests performed). As in the whole-brain decoding results, the left lateral prefrontal cortex (including the inferior and middle frontal gyrus and premotor cortex) represented tasks similarly in freely chosen and externally cued trials, $t(30) = 2.86$, $p = 0.003$. Importantly, we found the same patterns of results for the inferior parietal cortex as well, $t(30) = 2.58$, $p = 0.007$ (Figure 2C). The decoding accuracies between the original and cross-classified task decoding analyses differed

significantly in the prefrontal cortex, t(30) = 2.81, p = 0.009, and parietal cortex, t(30) = 3.13, p = 0.004, showing that the accuracies were higher in the original task decoding analysis without cross-classification.

Given that the prefrontal cluster identified in the task decoding analysis was fairly large, we repeated the ROI analysis, now taking into account anatomically defined sub-regions of the prefrontal cortex (as defined in the WFU_Pickatlas software, Maldjian et al., 2003). Specifically, ROIs were masked with anatomical masks encompassing the inferior frontal gyrus, medial frontal gyrus, and premotor cortex, respectively. Then, the leave-one-subject-out ROI analysis was performed as described above. All three prefrontal sub-regions thus tested represented the two tasks similarly (all ps < 0.05, corrected), demonstrating that there are no regional differences within the large prefrontal cluster with respect to the task cross-classification analysis. Furthermore, paired t-tests revealed that decoding accuracies were higher in the original decoding analysis in the left IFG and MFG (both ps < 0.05, corrected), but not in the premotor cortex (p > 0.05, corrected). The more sensitive ROI-based analysis therefore revealed that every brain region that encodes tasks during the delay period does so similarly for freely chosen and externally cued intentions, which is in contrast with some previous findings (Zhang et al., 2013). It should be noted that in all brain regions except the premotor cortex the decoding accuracies of the cross-classified task decoding analysis yielded significantly lower accuracies than the original task decoding analysis (without cross-classification). This difference possibly stems from the fact that applying cross-classification to MVPA reduces the number of data points used to train the classifier, thus reducing the power of the analysis. This in turn might lead to lower overall decoding accuracies. Alternatively, this might indicate that only some parts of the identified brain regions (e.g. only a subset of parietal voxels) encoded tasks similarly in free and cued trials. This might also have led to lower decoding accuracies in the cross-classification analysis. Regardless, these results show that each region identified contains at least some neuronal populations which encode free and cued tasks similarly.

In order to ensure that neuroimaging results were not contaminated by any systematic response strategies in the free choice conditions (Arrington & Logan 2004, Demanet et al., 2013), two further control analyses were performed. First, training data was screened for strong and stable choice biases and task preferences. No subjects were found to exhibit such preferences (see Methods for more details). Second, choices might be affected by the previous trial. It has been shown previously that subjects tend to repeat tasks more often than expected under free choice conditions ("repetition bias", Arrington and Logan, 2004). In some cases, the opposite bias ("switch bias") can be observed (Allefeld et al., 2013). Although such sequential biases do not necessarily indicate non-random choices (Allefeld et al., 2013), they possibly affect multivariate decoding results (Lages et al., 2013, but see Soon et al., 2014). We therefore tested whether any such biases were present in each subject individually. We used mutual information (MI) measures in order to test whether the currently performed freely chosen task was associated with the task performed on the previous trial. MI is not sensitive to the direction of possible biases and can detect both switch and repetition biases. In order to assess the statistical significance of the MI value for each subject, we performed permutation tests. We ran 10.000 random choice simulations, calculating the same MI value for each random simulation and used these simulations to extract a null distribution. The statistical significance of the MI value for each subject was assessed by computing its percentile value in the null distribution. We found that choices in six subjects were in fact influenced by the previous task condition ($p < 0.05$). We then excluded these subjects and repeated the ROI analysis described above. Even after this additional control, we still found every brain region identified in the task decoding analysis (lateral PFC and parietal cortex) to encode tasks similarly during the delay period for freely chosen and externally cued intentions.

We did not find medial prefrontal information in the task decoding analysis as has been reported previously using analyses that did not correct for task difficulty (e.g. Haynes et al. 2007; Momennejad et al. 2013; but see Zhang et al. 2013). We therefore used the medial prefrontal regions found in Haynes et al. (2007) in order to construct ROIs and test whether we find any task information in

these a priori defined ROIs. The authors in this study used similar tasks as we did here, and results should therefore be comparable (but note that Haynes et al., 2007 did not mix free and cued trials, as was the case in this experiment). We extracted the peak voxel coordinates from the anterior and posterior medial prefrontal cluster found in Haynes et al. 2007. We defined a sphere (radius 10mm) around these peak coordinates and used all voxels within the sphere as our ROI. For each subject, we extracted the accuracy values for the task decoding during the delay period in both ROIs. Using a t-test against chance level (Bonferroni corrected for the two models we compared) we assessed the significance of the results. We found no task information during the delay period in either the anterior or posterior medial prefrontal cortex (ps > 0.05).

Intention Type information

Distributed activation patterns in most of the lateral and medial prefrontal cortex, and parietal cortex encoded whether subjects freely chose a task or were externally cued to perform a task during the cue and delay phase (Figure 3A, p < 0.05 FWE corrected at the voxel level, minimal cluster size 20 voxels). The large effect likely stems from a global signal difference between freely chosen and externally cued trials. We therefore computed the univariate contrast of freely chosen minus externally cued trials, and found a widespread signal increase in a fronto-parietal network, including the rostral cingulate zone, lateral prefrontal cortex, and parietal cortex (p < 0.05 FWE corrected at the voxel level, minimal cluster size 20 voxels, Figure 3B). This is largely in line with previous work (Bengtsson et al., 2008; Forstmann et al., 2006; Mueller et al., 2007). The opposite contrast yielded no significant results. Given that multivariate analyses also reflect univariate signal contributions, albeit with a higher sensitivity (Davis et al., 2014), this explains the strong effect in our intention type decoding analysis.
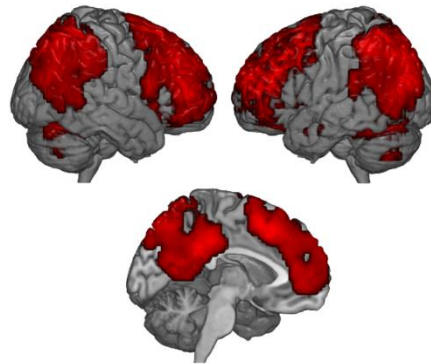
18

Control analysis: Visual cue decoding

In order to assess whether the visual features of the cues used in the experiment influenced task and intention type decoding results, we ran four separate control analyses, directly decoding the visual features of the cues in free addition, free subtraction, cued addition, and cued subtraction trials, respectively. All other variables were kept constant. We found no significant results in any of the four analyses ($p < 0.05$ FWE corrected at the cluster level, initial threshold $p < 0.005$). Furthermore, we also tested whether any of the ROIs defined in the task decoding analyses contained information about visual cues. This approach is more sensitive than whole-brain corrected analyses, yet it yielded no significant results in any ROI for any control analysis (all $ps > 0.05$, corrected). This confirms that our task and intention type decoding results do not merely reflect the visual features of the cues used in the experiment.
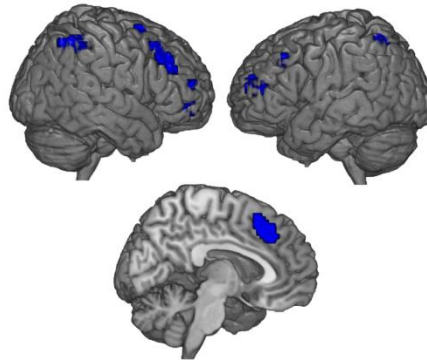
DISCUSSION

Our ability to form, maintain, and implement intentions is critical for goal-directed behavior. Intentions can be broadly classified according to how they are formed (Brass et al., 2013; Passingham et al., 2010): intentions can be determined largely by external cues, or they can be chosen freely. It remains an open question whether these different types of intentions are represented similarly in the brain or not. In this experiment, we applied MVPA to fMRI data in order to identify brain regions that represent freely chosen and externally cued abstract intentions. Similarly to previous findings (Bode et al., 2009; Reverberi et al., 2012; Zhang et al., 2013), we found that free and cued intentions were represented in a left-lateralized fronto-parietal network, including the IFG, MFG, premotor and parietal cortex. Interestingly, the whole network identified in this study represented freely chosen and externally cued intentions using similar spatial activation patterns, suggesting that these regions have a general role in representing the content of intentions under varying conditions.

Figure 3. A. Intention type information. Regions which contained information about whether subjects freely chose a task or were externally cued to perform the task during the cue presentation and delay phase. As can be seen, most parts of the frontal and parietal cortex encoded the intention type. **B. Intention type activation.** Shown in blue are brain regions in which there was a univariate global signal increase when subjects freely chose their task, as compared to being externally cued to perform the task. We found differences in the medial prefontal cortex, dorso-lateral and anterior lateral prefrontal cortex and the parietal cortex, similar to previous work (Forstmann et al., 2006, Passingham et al., 2010). These univariate signal differences partly explain the strong effect in the multivariate intention type decoding analysis.

General preparatory vs. specific intention-related processes

Previous experiments demonstrated that freely choosing intentions increases activity in the dmPFC,

dACC (Bengtsson et al., 2008; Forstmann et al., 2006), and the parietal cortex (Mueller et al., 2007).

This might indicate that these brain regions are involved in encoding and maintaining freely chosen

intentions. But it might equally reflect general preparatory processes merely correlated with freely choosing intentions (Fedorenko et al., 2013), such as increased demands on conflict monitoring (Botvinick et al., 2001, Brass and Haggard, 2008). Using classical univariate analyses, it is difficult to determine whether the identified network is related to specific intention-related processes or whether it merely reflects general preparatory processes. This issue is somewhat ameliorated in MVPA studies, where specific free or cued intentions can be directly contrasted. For instance, brain regions that represent whether subjects choose to add or subtract two numbers differ in the content of the chosen intention, but not in conflict monitoring demands. A growing number of studies which investigated intention representations using MVPA indicate that the anterior medial PFC (Gilbert, 2011), lateral PFC, and parietal cortex (Bode et al., 2009; Momennejad and Haynes, 2012, 2013; Waskom et al., 2014; Wisniewski et al., 2015) represent externally cued intentions. On the other hand, freely chosen intentions have been shown to be represented in the dmPFC, dACC (Wisniewski et al., 2014), frontopolar cortex, precuneus (Soon et al., 2008), medial and lateral PFC (Haynes et al., 2007). Our results emphasize the role of the lateral PFC, premotor and parietal cortex in representing both freely chosen and externally cued intentions, with relatively little involvement of the medial PFC (see below for a possible explanation). These brain regions, which have previously been associated with cognitive control more generally (Duncan, 2001; Miller and Cohen, 2001), appear to be specifically involved in the representation and maintenance of intentions. Please note that this result does not rule out that the same regions have functions beyond intention representation, in fact they are known to be involved in a variety of different contexts (e.g. Fedorenko et al., 2013; Yeo et al., 2015), such as cognitive control (Miller, 2000) or difficulty processing (Crittenden and Duncan, 2014). Our results help isolating their specific role in intention processing, controlling for other processes associated with the same brain regions.

Role of the medial and lateral PFC in intention processing

It has previously been suggested that the intention network is dissociable along the medial-lateral axis, especially in the PFC: The medial PFC is thought to be more strongly involved in processing freely chosen intentions, while the lateral PFC is thought to be more strongly involved processing externally cued intentions (Brass et al., 2013; Brass and Haggard, 2008; Passingham et al., 2010). This hypothesis is consistent with previous findings which showed that freely chosen intentions are encoded in the medial PFC (e.g. Wisniewski et al., 2014), while externally cued intentions are encoded in the lateral PFC (e.g. Bode et al., 2009). However, this picture is complicated by findings that show both medial and lateral PFC involvement in cued (Momennejad and Haynes, 2013) and free intentions (Haynes et al., 2007), or a dissociation between free and cued intentions in the opposite direction (Gilbert et al., 2009). Note that these divergent results might be partly due to the differences in tasks, stimuli and designs used across studies. Here, we found free intentions to activate both medial and lateral prefrontal as well as parietal regions, supporting no strong medial vs. lateral dissociation. The functional organization of the intention network thus remains an open issue. Our current data allow a direct within-subject comparison of freely chosen and externally cued intentions, using the same tasks and stimuli. Our findings emphasize the role of the lateral prefrontal cortex in both free and cued intention representation (see also Zhang et al., 2013). This is in line with previous research demonstrating that the lateral prefrontal cortex is involved in the free selection of task sets (Forstmann et al., 2006; Rowe et al., 2008). It has further been shown that both the dlPFC (Bengtsson et al., 2009), as well as the vlPFC (Bunge et al., 2003; Woolgar et al., 2011) are involved in the maintenance of task sets (for a review see e.g. Sakai, 2008). This important role of the PFC in task processing might be supported by highly flexible neuronal populations (Duncan, 2001), which are able to represent abstract task sets (Wallis et al., 2001). This is a possible neural basis of the similar representation of free and cued intentions in the same brain regions in our experiment (see below for more details). An unexpected result was that our attempts to find information about intentions in the medial PFC, even when using ROI analyses, were unsuccessful. One possible explanation is that

22

previous studies did not control for task difficulty effects on the MVPA results as conservatively as we did in the current study (Todd et al., 2013). In a previous study, over half of all grey matter voxels were correlated with RTs, likely reflecting difficulty effects (Todd et al., 2013). These effects potentially influence task decoding results, and need to be controlled for carefully (but see Woolgar et al., 2014). An alternative explanation might be that previous MVPA studies did not intermix free and cued trials (Bode et al., 2009; Haynes et al., 2007), whereas subjects alternated between performing free and cued trials in the current experiment. In the related task switching literature, it has been demonstrated that intermixing different tasks changes cognitive control processes, as compared to performing two tasks in a block-wise fashion (Monsell, 2003). Possibly, a similar effect is present when intermixing free and cued tasks, as compared to performing blocks of free or cued tasks in isolation. In this case, the cognitive control processes might differ between the current and previous studies (Haynes et al., 2007), and this might lead to different engagement of medial PFC, which is known to be involved in cognitive control (e.g. Ridderinkhof et al., 2004). A third explanation for this result is that the medial PFC is known to be involved in learning and performing new tasks, with its involvement decreasing as the new task is more routinely performed (Chein and Schneider, 2005; Koch et al., 2008). Subjects in our experiment underwent an extensive training session and performed the task routinely in the scanner, as evidenced by the very low error rates. This might have reduced our chances to find intention-related in formation in medial prefrontal brain regions (see Zhang et al., 2013 for a similar explanation).

In some previous studies, the frontopolar cortex (FPC) has been implicated in voluntary selection as well (Orr and Banich, 2014, Soon et al., 2008). In contrast, our results do not provide further evidence for its role in representing either free or cued intentions. In one study (Orr and Banich, 2014), the FPC has been shown to be uniquely activated under free intention conditions, but not under cued intention conditions. This FPC activation might be related to general preparatory processes associated with voluntary selection, and does not necessarily reflect intention representations. In the other study (Soon et al., 2008), the FPC has been found to encode free intentions before the chosen

intention became conscious. In contrast, the current study only investigated representations of intentions that are already conscious. This suggests that the FPC might be specialized to represent pre-conscious intentions, but more research is clearly needed to resolve this issue.

Functional specialization within the fronto-parietal intention network

In a recent study, Zhang et al., (2013) found freely chosen and externally cued intentions to be represented in a fronto-parietal network that is highly similar to our current findings. In their experiment, subjects were instructed to attend to one of three different stimulus features, or were able to freely choose to which feature to attend to. Task performance thus strongly rested on feature-based attention, leaving unclear whether the results reflect intentions or feature-based attention. In our experiment, subjects performed more abstract tasks and conditions did not systematically differ in terms of feature-based attention. Interpretation of these previous results was further complicated by how motor and RT effects on MVPA results were controlled, and by the choice reporting procedure used. In contrast to our experiment, Zhang et al. (2013) did not model RT effects in the first-level GLM analyses, which is thought to be a highly conservative control procedure for difficulty-related effects in MVPA studies (see Todd et al., 2013). Also, stimulus-response mappings were fixed for each participant, while they were pseudo-randomized on a trial-by-trial basis in our study. Both of these controls allow us to more clearly dissociate intention representations from motor-related processes. Furthermore, subjects had to explicitly report each choice in the previous study. This procedure increases the attention to one's own intentions, which is associated with increased activity in the lateral PFC and parietal cortex (Lau et al., 2004b). It might be that this additional attentional component affected intention representations found in Zhang et al. (2013). In contrast, we did not ask our subjects to explicitly report their choices, but rather inferred them from their responses in each trial. Despite these differences in experimental design and analysis methods, the intentions network identified in both studies is remarkably similar. This comparison rather shows that the fronto-parietal intention network has a general function in

processing and representing intentions, which is not specific to either attention-based or abstract, non-motor intentions.

Although the intention network identified in the present study is similar to the one reported by Zhang et al. (2013), we found no evidence for functional specialization within this network with respect to different types of intentions. Whereas Zhang et al. (2013) reported that the premotor and parietal cortex represent both freely chosen and externally cued intentions, the dlPFC was found to be specialized in representing freely chosen intentions only. Based on this finding, the authors argued that some regions within the intention network contain neuronal populations which are able to encode different types of intentions and therefore have a general role in intention processing. Other regions are characterized by more specialized neuronal populations, which represent either freely chosen or externally cued intentions. Our findings support a different interpretation.

In every region of the intention network identified, we found at least some neuronal populations which represent both freely chosen and externally cued intentions using similar spatial activation patterns. In light of our results, the whole intention network identified in this study seems to contain neuronal populations which are able to encode different types of intentions, suggesting that the whole network potentially plays a general role in intention processing. This extends previous findings which showed that similar regions can code different tasks flexibly (Stiers et al., 2010) and under varying conditions (Crittenden and Duncan, 2014, Waskom et al., 2014). Recently, it has been suggested that even the underlying mechanism for the selection of free and cued intentions is similar (Zhang et al., 2012). Both seem to be implemented using a common accumulation-to-threshold mechanism (Brown and Heathcote, 2008). Our findings suggest that a fronto-parietal network holds the intention representations that are the outcome of this selection mechanism. And given that this mechanism is similar for free and cued intentions, one might also expect the resulting intention representations to be similar as well. Relatedly, some recent research demonstrated that acquiring new skills is easier when their representations are similar to already learned skills (Stadtler et al.,

2014). Taken together, these results suggest that representational similarity might be beneficial to implement behavior in a variety of different, changing contexts.

In sum, our results demonstrate the generality of the fronto-parietal network in intention processing. Representations of specific intentions were found to be largely unaffected by the conditions under which they are formed, whether they were freely chosen or externally cued. Future studies will have to show whether this further generalizes to other types of intentions, such as reward-guided (Hampton and O'Doherty, 2007) or difficulty-guided (Wisniewski et al., 2014) intentions.

REFERENCES

Allefeld C, Soon CS , Bogler C, Heinzle J, Haynes JD. 2013. Sequential Dependencies between Trials in Free Choice Tasks. *arXiv:1311.0753 [q-Bio]*, http://arxiv.org/abs/1311.0753.

Arrington CM, Logan, GD. 2004. The cost of a voluntary task switch. *Psychological Science* 15: 610–15.

Beck SM, Ruge H, Walser M, Goschke T. 2014. The functional neuroanatomy of spontaneous retrieval and strategic monitoring of delayed intentions. *Neuropsychologia, 52:* 37-50.

Bengtsson SL, Haynes JD, Sakai K, Buckley MJ, Passingham RE. 2009. The Representation of Abstract Task Rules in the Human Prefrontal Cortex. *Cerebral Cortex*, *19:* 1929–1936.

Bode S, Haynes JD. 2009. Decoding sequential stages of task preparation in the human brain. *NeuroImage*, *45:* 606–613.

Botvinick MM, Braver TS, Barch DM, Carter CS, Cohen JD. 2001. Conflict monitoring and cognitive control. *Psychological Review*, *108:* 624–652.

Brass M, Haggard P. 2008. The What, When, Whether Model of Intentional Action. *The Neuroscientist*, *14:* 319–325.

Brass M, Lynn MT, Demanet J, Rigoni D. 2013. Imaging volition: what the brain can tell us about the will. *Experimental Brain Research*, *229:* 301–312.

Brown SD, Heathcote A. 2008. The Simplest Complete Model of Choice Response Time: Linear Ballistic Accumulation. *Cognitive Psychology,* 57: 153–78.

Bunge SA, Kahn I, Wallis JD, Miller EK, Wagner AD. 2003. Neural Circuits Subserving the Retrieval and Maintenance of Abstract Rules. *Journal of Neurophysiology,* 90: 3419–3428.

Chein JM, Schneider W. 2005. Neuroimaging studies of practice-related change: fMRI and meta-analytic evidence of a domain-general control network for learning. *Cognitive Brain Research*, *25:* 607–623.

Cox DD, Savoy RL. 2003. Functional magnetic resonance imaging (fMRI) "brain reading": detecting and classifying distributed patterns of fMRI activity in human visual cortex. *NeuroImage*, *19:* 261–270.

Crittenden BM, Duncan J. 2014. Task Difficulty Manipulation Reveals Multiple Demand Activity but No Frontal Lobe Hierarchy. *Cerebral Cortex, 24*: 532–40.

Cunnington R, Windischberger C, Deecke L, Moser E. 2002. The Preparation and Execution of Self-Initiated and Externally-Triggered Movement: A Study of Event-Related fMRI. *NeuroImage*, *15:* 373–385.

Davis T, LaRocque KF, Mumford JA, Norman KA, Wagner AD, Poldrack RA. 2014. What do differences between multi-voxel and univariate analysis mean? How subject-, voxel-, and trial-level variance impact fMRI analysis. *NeuroImage*, *97:* 271–283.

Deiber MP, Passingham RE, Colebatch JG, Friston KJ, Nixon PD, Frackowiak RSJ. 1991. Cortical Areas and the Selection of Movement: A Study with Positron Emission Tomography. *Experimental Brain Research, 84*: 393–402.

Demanet J, De Baene W, Arrington CM, Brass M. 2013. Biasing Free Choices: The Role of the Rostral Cingulate Zone in Intentional Control. *NeuroImage,* 72: 207–13.

Duncan J. 2010. The multiple-demand (MD) system of the primate brain: mental programs for intelligent behaviour. *Trends in Cognitive Sciences*, *14*: 172–179.

Duncan J. 2001. An Adaptive Coding Model of Neural Function in Prefrontal Cortex. *Nature Reviews Neuroscience,* 2: 820–829.

Eriksen CW. 1995. The Flankers Task and Response Competition: A Useful Tool for Investigating a Variety of Cognitive Problems. *Visual Cognition ,* 2: 101–18.

Esterman M, Tamber-Rosenau BJ, Chiu YC, Yantis S. Avoiding Non-Independence in fMRI Data Analysis: Leave One Subject out. *NeuroImage,* 50: 572–76.

Fedorenko E, Duncan J, Kanwisher N. 2013. Broad Domain Generality in Focal Regions of Frontal and Parietal Cortex. *Proceedings of the National Academy of Sciences, 110*: 16616–21.

Forstmann BU, Brass M, Koch I, von Cramon DY. 2006. Voluntary selection of task sets revealed by functional magnetic resonance imaging. *Journal of Cognitive Neuroscience*, *18*: 388–398.

Friston KJ, Holmes AP, Worsley KJ, Poline JP, Frith CD, Frackowiak RS. 1994. Statistical parametric maps in functional imaging: a general linear approach. *Human Brain Mapping*, *2:* 189–210.

Gilbert SJ. 2011. Decoding the Content of Delayed Intentions. *Journal of Neuroscience*, *31*: 2888–2894.

Gilbert SJ, Gollwitzer PM, Cohen AL, Oettingen G, Burgess PW. 2009. Separable brain systems supporting cued versus self-initiated realization of delayed intentions. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *35*: 905–915.

Goschke T. 2013. Volition in action: Intentions, control dilemmas and the dynamic regulation of cognitive control. In W Prinz, A Beisert, A Herwig (Eds.), *Action science: Foundations of an emerging discipline* (pp. 409-434).

Haggard P. 2008. Human volition: towards a neuroscience of will. *Nature Reviews Neuroscience, 9:* 934-946.

Hampton AN, O'Doherty JP. 2007. Decoding the Neural Substrates of Reward-Related Decision Making with Functional MRI. *Proceedings of the National Academy of Sciences* 104: 1377–82.

Haxby JV. 2012. Multivariate pattern analysis of fMRI: The early beginnings. *NeuroImage*, *62:* 852–855.

Haynes JD, Rees G. 2006. Decoding mental states from brain activity in humans. *Nature Reviews Neuroscience*, *7*: 523–534.

Haynes JD, Sakai K, Rees G, Gilbert S, Frith C, Passingham RE. 2007. Reading Hidden Intentions in the Human Brain. *Current Biology*, *17*:323–328.

Haynes JD. 2015. A Primer on Pattern-Based Approaches to fMRI: Principles, Pitfalls, and Perspectives. *Neuron* 87: 257–70.

Hebart MN, Görgen K, Haynes JD. 2014. The Decoding Toolbox (TDT): A versatile software package for multivariate analyses of functional imaging data. *Frontiers in Neuroinformatics*, *8*, 88.

Jahanshahi M, Jenkins IH, Brown RG, Marsden CD, Passingham RE, Brooks DJ. 1995. Self-initiated versus externally triggered movements I. An investigation using measurement of regional cerebral blood flow with PET and movement-related potentials in normal and Parkinson's disease subjects. *Brain*, *118:* 913–933.

Kamitani Y, Tong F. 2005. Decoding the visual and subjective contents of the human brain. *Nature Neuroscience*, *8:* 679–685.

Kaplan JT, Man K, Greening SG. 2015. Multivariate Cross-Classification: Applying Machine Learning Techniques to Characterize Abstraction in Neural Representations. *Frontiers in Human Neuroscience* 9, doi:10.3389/fnhum.2015.00151.

Koch K, Schachtzabel C, Wagner G, Reichenbach JR, Sauer H, Schlösser R. 2008. The neural correlates of reward-related trial-and-error learning: An fMRI study with a probabilistic learning task. *Learning & Memory*, *15:* 728–732.

Kriegeskorte N, Goebel R, Bandettini P. 2006. Information-based functional brain mapping. *Proceedings of the National Academy of Sciences of the United States of America*, *103:* 3863–3868.

Lages M, Stephanie CB, Jaworska K. 2013. Flipping a Coin in Your Head without Monitoring Outcomes? Comments on Predicting Free Choices and a Demo Program. *Frontiers in Psychology* 4, doi:10.3389/fpsyg.2013.00925.

Lau H, Rogers R, Ramnani N, Passingham R. 2004a. Willed action and attention to the selection of action. *NeuroImage*, *21:* 1407–1415.

Lau H. 2004b. Attention to Intention. *Science*, *303*: 1208–1210.

Maldjian JA, Laurienti PJ, Burdette JB, Kraft RA. 2003. An Automated Method for Neuroanatomic and Cytoarchitectonic Atlas-based Interrogation of fMRI Data Sets. *NeuroImage*, 19:1233-1239.

Miller EK, Cohen JD. 2001. An integrative theory of prefrontal cortex function. *Annual Review of Neuroscience*, *24:* 167–202.

Miller EK. 2000. The Prefontral Cortex and Cognitive Control. *Nature Reviews Neuroscience,* 1: 59–65.

Mitchell TM, Hutchinson R, Niculescu RS, Pereira F, Wang X, Just M, Newman S. 2004. Learning to decode cognitive states from brain images. *Machine Learning*, *57*: 145–175.

Momennejad I, Haynes JD. 2012. Human anterior prefrontal cortex encodes the "what" and "when" of future intentions. *NeuroImage*, *61:* 139–148.

Momennejad I, Haynes JD. 2013. Encoding of Prospective Tasks in the Human Prefrontal Cortex under Varying Task Loads. *Journal of Neuroscience*, *33:* 17342–17349.

Monsell S. 2013. Task Switching. *Trends in Cognitive Sciences* 7: 134–40.

Mueller VA, Brass M, Waszak F, Prinz W. 2007. The role of the preSMA and the rostral cingulate zone in internally selected actions. *NeuroImage*, *37:* 1354–1361.

Nachev P, Husain M. 2010. Action and the fallacy of the "internal": Comment on Passingham et al. *Trends in Cognitive Sciences*, *14:* 192–193.

Passingham RE, Bengtsson SL, Lau H. 2010. Medial frontal cortex: from self-generated action to reflection on one's own performance. *Trends in Cognitive Sciences*, *14:* 16–21.

Rae C, Hughes LE, Weaver C, Anderson MC, Rowe JB. 2014. Selection and Stopping in Voluntary Action: A Meta-Analysis and Combined fMRI Study. *NeuroImage,* 86: 381–91.

Reverberi C, Görgen, K, Haynes JD. 2012. Distributed Representations of Rule Identity and Rule Order in Human Frontal Cortex and Striatum. *Journal of Neuroscience*, *32:* 17420–17430.

Ridderinkhof KR, Ullsperger M, Crone EA, Nieuwenhuis S. 2004. The Role of the Medial Frontal Cortex in Cognitive Control. *Science*, 306: 443–47.

Rowe JB, Hughes L, Eckstein D, Owen AM. 2008. Rule-Selection and Action-Selection Have a Shared Neuroanatomical Basis in the Human Prefrontal and Parietal Cortex. *Cerebral Cortex,* 18: 2275–2285.

Ruge H, Jamadar S, Zimmermann U, Karayanidis F. 2013. The many faces of preparatory control in task switching: reviewing a decade of fMRI research. *Human Brain Mapping, 34:* 12-35.

Sakai K. 2008. Task Set and Prefrontal Cortex. *Annual Review of Neuroscience,* 31: 219–245.

Soon CS, Brass M, Heinze HJ, Haynes JD. 2008. Unconscious determinants of free decisions in the human brain. *Nature Neuroscience*, *11:* 543–545.

Soon CS, Allefeld C, Bogler C, Heinzle J, Haynes JD. 2014. Predictive Brain Signals Best Predict Upcoming and Not Previous Choices. *Frontiers in Psychology* 5, doi:10.3389/fpsyg.2014.00406.

Soon CS, He AH, Bode S, Haynes JD. 2013. Predicting free choices for abstract intentions. *Proceedings of the National Academy of Sciences*, *110:* 6217–6222.

Sadtler PT, Quick KM, Golub MD, Chase SM, Ryu SI, Tyler-Kabara EC, Yu BM, Batista AP. 2014. Neural Constraints on Learning. *Nature,* 512: 423–26.

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.

Stürmer B, Leuthold L, Soetens E, Schröter H, Sommer W. 2002. Control over Location-Based Response Activation in the Simon Task: Behavioral and Electrophysiological Evidence. *Journal of Experimental Psychology: Human Perception and Performance* 28: 1345–63.

Todd MT, Nystrom LE, Cohen JD. 2013. Confounds in multivariate pattern analysis: Theory and rule representation case study. *NeuroImage*, *77:* 157–165.

Tong F, Pratte MS. 2012. Decoding Patterns of Human Brain Activity. *Annual Review of Psychology*, *63:* 483–509.

Wallis JD, Anderson KC, Miller EK. 2001. Single Neurons in Prefrontal Cortex Encode Abstract Rules. *Nature,* 411: 953–956.

Waskom ML, Kumaran D, Gordon AM, Rissman J, Wagner AD. 2014. Frontoparietal Representations of Task Context Support the Flexible Control of Goal-Directed Cognition. *Journal of Neuroscience,* 34: 10743–55.

Wisniewski D, Reverberi C, Tusche A, Haynes JD. 2014. The Neural Representation of Voluntary Task-Set Selection in Dynamic Environments. *Cerebral Cortex*. doi:10.1093/cercor/bhu155

Wisniewski D, Reverberi C, Momennejad I, Kahnt T, Haynes JD. 2015. The Role of the Parietal Cortex in the Representation of Task–Reward Associations. *The Journal of Neuroscience* 35: 12355–65.

Woolgar A, Golland P, Bode S. 2014. Coping with confounds in multivoxel pattern analysis: What should we do about reaction time differences? A comment on Todd, Nystrom & Cohen 2013. *NeuroImage 98: 506-512*

Woolgar A, Thompson R, Bor D, Duncan J. 2011. Multi-Voxel Coding of Stimuli, Rules, and Responses in Human Frontoparietal Cortex. *NeuroImage,* 56: 744–752.

Yeo BTT, Krienen FM, Eickhoff SB, Yaakub SN, Fox PT, Buckner RL, Asplund CL, Chee MWL. 2015. Functional Specialization and Flexibility in Human Association Cortex. *Cerebral Cortex,* 25: 3654–3672.

Zhang J, Hughes LE, Rowe JB. 2012. Selection and Inhibition Mechanisms for Human Voluntary Action Decisions. *NeuroImage,* 63: 392–402.

Zhang J, Kriegeskorte N, Carlin JD, Rowe JB. 2013. Choosing the Rules: Distinct and Overlapping Frontoparietal Representations of Task Rules for Perceptual Decisions. *Journal of Neuroscience*, *33*: 11852–11862.

Tables

Table 1. Brain regions identified in task and intention type analyses

Brain Region	Side	Cluster Size	Peak MNI-coordinates		
			x	y	z
Task Decoding (no cross-classification)					
PFC	L	1685	-33	-58	43
parietal cortex	L	1406	-54	-4	37
Task Decoding (cross-classification across intention types)					
PFC	L	403	-57	26	25
Intention Type Decoding					
PFC	B	12515	-39	47	22
parietal cortex	B	11754	-36	-49	46
cerebellum	R	463	39	-61	-41
Intention Type Activation (Univariate Contrast: Free - Cued)					
inferior parietal cortex	R	185	39	-58	40
posterior MFG	R	168	45	32	31
dorso-medial PFC / preSMA	B	146	0	26	43
anterior MFG	L	74	-39	53	13
inferior parietal cortex	L	49	-42	-55	55
posterior MFG	L	44	-42	29	34
orbital MFG	R	25	39	50	-8
anterior MFG	R	24	36	50	10

Task decoding results are shown for a statistical threshold of $p < 0.05$ (FWE corrected at the cluster level, initial threshold $p < 0.005$). Intention type results are shown for a statistical threshold of $p < 0.05$ (FWE corrected at the voxel level, minimal cluster size 20 voxels). Abbreviations: L left, R right, B bilateral, MFG middle frontal gyrus, PFC prefrontal cortex, SMA supplementary motor area