



## Multimodal MRI cerebral correlates of verbal fluency switching and its impairment in women with depression

L. Domain<sup>a</sup>, M. Guillery<sup>a</sup>, N. Linz<sup>b</sup>, A. König<sup>c,d</sup>, J.M. Batail<sup>a</sup>, R. David<sup>e</sup>, I. Corouge<sup>f</sup>, E. Bannier<sup>f</sup>, J.C. Ferré<sup>f</sup>, T. Dondaine<sup>g</sup>, D. Drapier<sup>a</sup>, G.H. Robert<sup>a,f,\*</sup>

<sup>a</sup> *Universitary Department of Psychiatry, Centre Hospitalier Guillaume Régnier, Rennes, France*

<sup>b</sup> *ki:elements, Saarbrücken, Germany*

<sup>c</sup> *Stars Team, Institut National de Recherche en Informatique et en Automatique (INRIA), Sophia Antipolis, France*

<sup>d</sup> *CoBTeK (Cognition-Behaviour-Technology) Lab, FRIS-University Côte d'Azur, Nice, France*

<sup>e</sup> *Old-age Psychiatry DEPARTMENT, Geriatriy Division, University of Nice, France*

<sup>f</sup> *U1228 Empenn, UMR 6074, IRISA, University of Rennes 1, France*

<sup>g</sup> *Univ. Lille, Inserm, CHU Lille, LilNCog, Lille Neuroscience & Cognition, F-59000 Lille, France*

### ARTICLE INFO

#### Keywords:

Major depressive disorder  
Automatic speech analysis  
Switching  
Verbal fluency  
Cortical Thickness  
Fractional Anisotropy

### ABSTRACT

**Background:** The search of biomarkers in the field of depression requires easy implementable tests that are biologically rooted. Qualitative analysis of verbal fluency tests (VFT) are good candidates, but its cerebral correlates are unknown.

**Methods:** We collected qualitative semantic and phonemic VFT scores along with grey and white matter anatomical MRI of depressed (n = 26) and healthy controls (HC, n = 25) women. Qualitative VFT variables are the “clustering score” (i.e. the ability to produce words within subcategories) and the “switching score” (i.e. the ability to switch between clusters). The clustering and switching scores were automatically calculated using a data-driven approach. Brain measures were cortical thickness (CT) and fractional anisotropy (FA). We tested for associations between CT, FA and qualitative VFT variables within each group.

**Results:** Patients had reduced switching VFT scores compared to HC. Thicker cortex was associated with better switching score in semantic VFT bilaterally in the frontal (superior, rostral middle and inferior gyri), parietal (inferior parietal lobule including the supramarginal gyri), temporal (transverse and fusiform gyri) and occipital (lingual gyri) lobes in the depressed group. Positive association between FA and the switching score in semantic VFT was retrieved in depressed patients within the corpus callosum, right inferior fronto-occipital fasciculus, right superior longitudinal fasciculus extending to the anterior thalamic radiation (all p < 0.05, corrected).

**Conclusion:** Together, these results suggest that automatic qualitative VFT scores are associated with brain anatomy and reinforce its potential use as a surrogate for depression cerebral bases.

### 1. Introduction

Major depressive disorder (MDD) affects 350 million people (women predominantly) worldwide and is disabling due to chronicity (James et al., 2018). In practice, it remains challenging for clinicians to predict the course of MDD given its clinical heterogeneity and the lack of biomarkers. The development of magnetic resonance neuroimaging (MRI)

technologies has enabled significant advances towards the quest for such biomarkers. Specifically, anatomical MRI (e.g. structural, diffusion-weighted MRI), by allowing visualization and quantification of brain alterations in MDD have played a critical role in deciphering the pathogenesis of the disease. From structural MRI, cortical thickness (i.e. the distance between the white matter-gray matter surface) can be estimated. Cortical thickness (CT) permits a higher registration accuracy

**Abbreviations:** AES, Apathy Evaluation Scale; CT, Cortical Thickness; CAT, Computational Anatomy Toolbox; DSM-IV-TR, Diagnostic and Statistical Manual of Mental Disorders; DSST, digit-symbol substitution test; FA, Fractional Anisotropy; FWE, Familywise error rate; FDR, false discovery rate; GLM, general linear model; HC, healthy controls; JHU, John Hopkins University; MADRS, Montgomery Åsberg Depression Rating Scale; MDD, Major depressive disorder; MINI, Mini International Neuropsychiatric Interview; MRI, Magnetic Resonance neuroimaging; ROI, region of interest; SPM, Statistical Parametric Mapping; TMT-A & B, Trail Making Test; TBSS, Tract-Based Spatial Statistics; TIV, total intracranial volume; VFT, verbal fluency tasks.

\* Corresponding author at: University Department of Psychiatry, Centre Hospitalier Guillaume Régnier, 108 Boulevard du Général Leclerc, 35700 Rennes, France.

<https://doi.org/10.1016/j.nicl.2021.102910>

Received 30 July 2021; Received in revised form 3 December 2021; Accepted 4 December 2021

Available online 6 December 2021

2213-1582/© 2021 Published by Elsevier Inc. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

than volume-based registration (Desai et al., 2005), reduces the risk of possible false positives (Greve and Fischl, 2018) and is considered as a sensitive (Lerch and Evans, 2005) and specific (Lerch et al., 2008) measure of brain structure. Moreover, other standard approaches such as surface and volume estimation should be normalized to the total intra cranial volume, leaving room to possible variance in the estimated effect sizes (Schwarz et al., 2016). Recent results have reported widespread cortical thinning in depressed patients in comparison to healthy controls (HC) in the orbitofrontal cortex, insula, anterior and posterior cingulate, and temporal lobes (Schmaal et al., 2017), providing candidates MRI-derived biomarkers (Jiang et al., 2020). Diffusion weighted imaging measures water molecule diffusion to estimate fractional anisotropy (FA) along the white matter bundles. Large sample analyses refer to small effect sizes but widely distributed reduced FA in MDD compared to HC within the superior and inferior longitudinal fasciculus (S/ILF, the corpus callosum (CC), the uncinate, the inferior fronto-occipital fasciculus (IFOF) (van Velzen et al., 2020). These findings have led to the suggestion that structural and diffusion-weighted MRI could be a promising tool for identifying imaging biomarkers of MDD.

However, these methods have contraindications, are costly, and are currently not widely available. Therefore, there is a need for simple tools, doable in routine but which have clear associations with biological, a condition required to be considered as biomarkers (Ruggeri et al., 2014). Cognitive functions are good candidates since: i) executive dysfunction is frequent in patients with MDD (Snyder, 2013; Fossati et al., 2003), ii) it can be performed in clinical settings. Recent development automates their assessments (<https://ki-elements.de/en/start/>, <https://www.cambridgecognition.com/>) which make it easy to quantify and standardize individual performances against population norms. Verbal fluency tasks (VFT) are among the quickest and easiest to acquire and measures both executive and language abilities (Snyder, 2013) participants are required to produce as many words belonging to a semantic or phonemic category as possible within a given time (Troyer et al., 1997). However, in a recent meta-analysis, VFT failed to predict the course of MDD (Pimontel et al., 2016). This might be related to the scoring method, as most studies use the quantitative score (i.e. simply summing correct words) which does not shine any light on the cognitive function. In contrast, the qualitative analysis of the performance provides a more fine-grained measure of cognitive function (Troyer et al., 1997). It consists in measuring the size of word subcategories or cluster (clustering score) and counting the number of transitions (switching score) between clusters (Troyer et al., 1997). Those “clustering” and “switching” scores reflects the integrity of the storage in lexico-semantic memory and executive function respectively (Troyer et al., 1997). Separating between clustering and switching should allow clinicians to distinguish the two cognitive processes occurring during VF: accessing semantic memory to retrieve words and executive search functions active in navigating the larger semantic stock. Typically, the clusters are determined using predefined subcategories, following the approach from Troyer et al. (Troyer et al., 1997). This leaves room for subjective interpretation, and variance in cluster sizes and switch counts. Statistical model-based approaches which allow for automatic cluster identification reduce this variance and showed promising results of identifying early stages of major cognitive disorders (Linz et al., 2017; Linz et al., 2017; Tröger et al., 2017; König et al., 2015). This offers new perspectives on using qualitative VFT performance (i.e. clustering and switching scores) as reliable cognitive marker of MDD. However, this is currently unknown if qualitative VFT performance are associated with cerebral features, especially in MDD (Ruggeri et al., 2014).

Consistently with executive deficits in MDD with preserved language and memory functions, our hypothesis are to find decreased switching performances with similar performances of the clustering (Fossati et al., 2003). Among the MDD participants, we expect switching performances to be associated with thinner cortices among cortical regions involved in executive functions such as the frontal gyri (Yuan and Raz, 2014) and semantic VF (such as the temporal cortices) (Vonk et al., 2019).

Similarly, we expected decreased FA among white matter tracts both involved in cognitive switching, semantic classification and verbal fluency, such as the superior longitudinal fasciculus (SLF) and the inferior fronto-occipital fasciculus (IFOF) (Hoagey et al., 2021; Garcin et al., 2018; Li et al., 2017).

To fill in this gap, the primary objective of this study was to test for association between automatic-computerized measures of clustering and switching in semantic and phonemic VFT, and measures of CT and FA in sample of patients with MDD and in HC. The secondary objective was to compare the groups in term of switching and clustering scores during both VFT, and in term of MRI-derived metrics (CT and FA).

## 2. Material and methods

### 2.1. Participants

We included 51 women (26 patients) between July 2012 and October 2014. All patients were outpatient from an expert center in mood disorders of Rennes University Hospital (France). All participants were native French speakers. All patients were suffering from either chronic or recurrent depression. Patients met Diagnostic and Statistical Manual of Mental Disorders (DSM-IV-TR) criteria for MDD and had a minimum Montgomery Åsberg Depression Rating Scale (MADRS) score of 21. Clinical characteristics of the sample is provided by the mean of the 4 factor structure of the MADRS (Quilty et al., 2013). They had no other psychiatric disorder, as measured with the Mini International Neuropsychiatric Interview (MINI) and were all native French speakers. All assessment (including MRI) were performed within the week after inclusion. Healthy volunteers did not have current psychiatric disorder (according to the MINI). Since VFT are amenable to effects of age (Tombaugh, 1999) and education (Tombaugh, 1999), the HC were matched for age and education level with the MDD group. All participants were right-handed (according to the Edinburgh Handedness Inventory) and did not exhibit any severe cognitive impairment (defined as a score < 130 on the Mattis Dementia Rating Scale) or suicide risk (assessed by the Clinical Global Impression of Severity of Suicidality Scale). Exclusion criteria were potential contraindications to MRI (pacemakers, metal implants, pregnancy, and lactation). The study was approved by the relevant institutional review board (CPP of Nancy number 2759, ID-RCB number 2019-A00111-56). All participants provided their written informed consent.

### 2.2. Procedure

After inclusion, each participant underwent a structural MRI scan, and psychiatric and neuropsychological assessments within the same week.

#### 2.2.1. Neuropsychological assessment

All participants underwent the same neuropsychological battery, in the same order. They were asked to generate as many words as possible within the space of 120 s. For semantic VFT, they had to produce words according to a category criterion (“animal”). For phonemic VFT, they had to produce words beginning with the letter “P”. The words were split into 30-seconds time frames. Additionally, processing speed was assessed with the digit-symbol substitution test (DSST), cognitive switching with the TMT, cognitive inhibition with the Stroop Test, abstract reasoning and set-shifting with the modified Wisconsin Card Sorting Test (WCST) and working memory with the direct/indirect spans.

#### 2.2.2. Fully automatic analysis of the verbal fluency tests

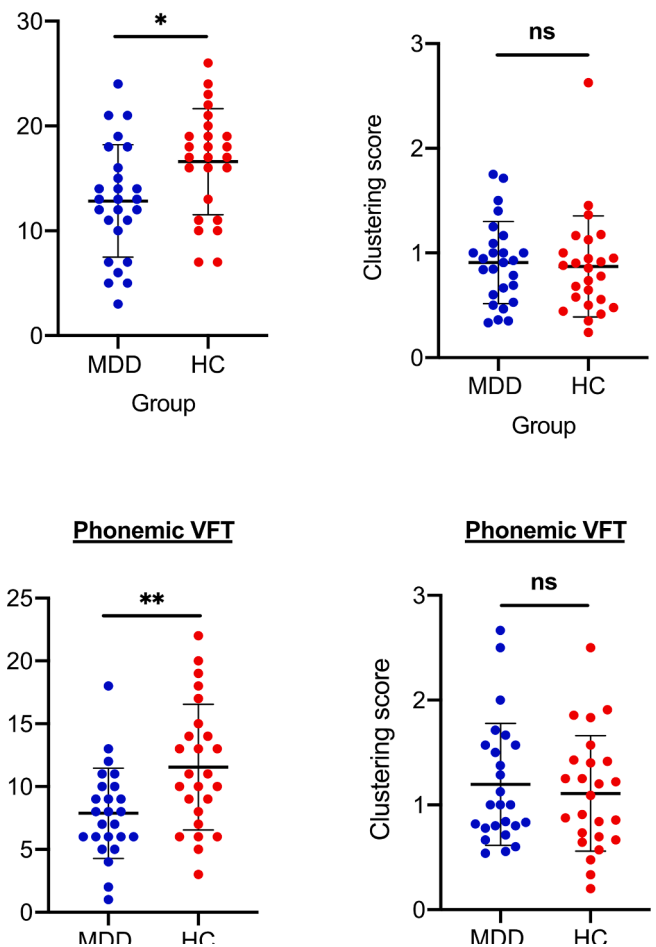
To compute the clusters, we used the statistical method described by Linz et al. (Linz et al., 2017). This automatic method of analysis requires words (textual information) to be converted into a numerical form (vector). This conversion method is known as word vectorization. Each

**Table 1**

Clinical and Demographic Variables. Values given are means (Standard Deviation). Except for the variable “treatment” where values are proportion (%). MADRS = Montgomery Asberg Depression Scale, MADRS-S: sadness factor; MADRS-NV: neurovegetative factor; MADRS-D: detachment factor; MADRS-NT: negative thinking factor; AES: apathy evaluation scale; STAI-B: State Trait Anxiety Inventory – trait; MDRS: Mattis Dementia Rating Scale; W: unpaired Wilcoxon’s statistic.

	MDD group (n = 26)	HC group (n = 25)	Statistic
Age (years)	46.69 (11)	49.5 (9.5)	W = 281, p = 0.41
Education (years)	12.9 (2.3)	13.7 (2.4)	W = 259, p = 0.21
Number of depressive episodes	4.7 (3.5)		
Age at MDD onset (years)	31.9 (13.5)		
Disease duration (years)	14.5 (6.3)		
Treatment			
Antidepressants			
Benzodiazepines	24/26		
Antipsychotics	(92.3%)		
Electroconvulsive Therapy	18/26		
Transcranial Magnetic Stimulation	(69.2%)		
Stimulation	3/26		
Transcranial Direct Current Stimulation	(11.6%)		
Stimulation	2/26		
Stimulation	(7.7%)		
Stimulation	2/26		
Stimulation	(7.7%)		
Stimulation	1/26		
Stimulation	(3.9%)		
Stimulation	1/26		
Stimulation	(3.9%)		
MADRS score	29.92 (6.8)	1.7 (1.9)	W = 598, p < 0.001
• MADRS-S	6.9 (1.9)	0.5 (0.6)	
• MADRS-NV	8.5 (2.8)	0.4 (0.5)	
• MADRS-D	10.1 (2.6)	0.4 (0.6)	
• MADRS-NT	4.3 (1.9)	0.3 (0.7)	
AES score	45.6 (8.1)	25.8 (5.5)	W = 607, p < 0.001
STAI-B score	63.9 (12.7)	34.2 (7.5)	W = 575.5, p < 0.001
MDRS total score	136.2 (9.3)	142 (1.2)	W = 199.5, p = 0.03
Total Intracranial Volume (cm <sup>3</sup> )	1356.7 (98.8)	1362.4 (96.4)	W = 325, p = 1

vector considers the semantic or phonemic similarity of the word (captured by looking at the word’s context in large corpus of data) relatively to the other words produced. Finally, using a similarity cut-off (mean similarity of any two random words), the words were grouped into clusters. Semantic distance is calculated as the semantic distance between all possible word pairs for each subject during the VFT, estimated based on large word corpus. Here, semantic distance was determined using the fastText word embedding pipeline (Linz et al., 2017); pretrained on the Common Crawl® and the Wikipedia® corpora (Grave and Bojanowski, 2018). It acts like a map of the semantic search performed by each subject. We then applied Troyer’s rules for scoring the cluster size and switch numbers (Troyer et al., 1997). The mean cluster size is the average number of words contained in a semantic cluster (a sequential group of words that have a semantic relation). The size of any cluster is the number of words in it minus one (as single word clusters are considered length zero). The cluster size reflects the integrity of the lexico-semantic memory, while the switching score is related to the ability to access and navigate among the semantic stock. Typically, if executive functions are impaired, patients are more likely to latch onto a semantic cue they have discovered and produce words that are closely related (i.e. fewer switches). If semantic memory is impaired, patients will move on more quickly from a given cue and therefore explore more different semantic categories (i.e. smaller clusters). Both are triggered by the time constraint and the goal of the subject to produce as many words as possible. Currently, there are no existing norms for the



**Fig. 1.** Scatter dot plots of the switching and clustering scores stratified by groups and phonemic and semantic VFT.

qualitative measures of VF in French.

### 2.2.3. MRI acquisition

All MRI studies were performed on a 3 T Magnetom Verio Syngo MR B17 scanner (Siemens Healthcare, Erlangen, Germany). An anatomical scan was performed using the following 3D T1-weighted, sagittal, magnetization prepared rapid gradient echo (MPRAGE) sequence: 176 sagittal slabs, voxel size  $1 \times 1 \times 1 \text{ mm}^3$ , repetition time 1900 ms, echo time 2.26 ms, inversion time 900 ms, field of view  $256 \times 256 \text{ mm}^2$ , number of excitations 1, GRAPPA acceleration factor 2. Diffusion weighted images were acquired with an EPI sequence (30 directions, b-value  $1000 \text{ s/mm}^2$ ) with repetition time 11000 ms, echo time 99 ms, field of view  $256 \times 256 \text{ mm}^2$ , 60 slices and voxel size:  $2 \times 2 \times 2 \text{ mm}^3$ .

### 2.2.4. Cortical thickness

Data were preprocessed and analyzed using Statistical Parametric Mapping (SPM) Version 12 (Ashburner and Friston, 2000) (Wellcome Department of Cognitive Neurology, London, UK, <http://www.fil.ion.ucl.ac.uk/spm/>) with MATLAB 2018b (The Mathworks Inc., Sherborn, MA, United States). First, T1-weighted images of each participant were manually reoriented to the anterior commissure. The reoriented structural images were then segmented using the Segment option of the CAT (Computational Anatomy Toolbox) Version 12.6 (Gaser and Dahnke, 2016). The resulting segmented images were then used to estimate the cortical thickness and central surface simultaneously, using the projection based thickness. This approach does not require a reconstruction of the outer boundary or white matter surface deformations for thickness estimation within sulci where partial volume effect and blurred images

**Table 2**

VFT scores in the MDD and HC groups. Mean (SD), W = unpaired Wilcoxon's statistics.

	MDD group	HC group	Comparison
<b>Semantic VFT: animals</b>			
total words	<b>25 (9.01)</b>	<b>30.7 (5.4)</b>	<b>W = 168.5, p = 0.003</b>
0–30 sec	10.3 (3.5)	13.4 (2.4)	W = 137, p = 0.0004
30–60 sec	5.8 (3.1)	7.9 (2.9)	W = 203, p = 0.021
60–90 sec	5.0 (3.1)	5.4 (2.3)	W = 301.5, p = 0.66
90–120 sec	3.9 (2.6)	4.0 (2.6)	W = 363.5, p = 0.47
word frequency	<b>3.7 (0.25)</b>	<b>3.6 (0.14)</b>	<b>W = 369, p = 0.42</b>
0–30 sec	3.88 (0.37)	3.8 (0.21)	W = 343, p = 0.74
30–60 sec	3.57 (0.33)	3.56 (0.26)	W = 319.5, p = 0.92
60–90 sec	3.44 (0.8)	3.37 (0.38)	W = 403, p = 0.15
90–120 sec	3.12 (1)	3.17 (1)	W = 297.5, p = 0.61
semantic distance	<b>0.41 (0.05)</b>	<b>0.4 (0.03)</b>	<b>W = 387, p = 0.25</b>
0–30 sec	0.48 (0.06)	0.45 (0.03)	W = 402, p = 0.15
30–60 sec	0.57 (0.2)	0.47 (0.07)	W = 459, p = 0.012
60–90 sec	0.57 (0.2)	0.54 (0.13)	W = 369, p = 0.41
90–120 sec	0.55 (0.21)	0.52 (0.21)	W = 363.5, p = 0.47
number of switches	<b>12.85 (5.4)</b>	<b>16.6 (5.1)</b>	<b>W = 200, p = 0.019</b>
mean cluster size	<b>0.91 (0.39)</b>	<b>0.87 (0.48)</b>	<b>W = 387.5, p = 0.25</b>
<b>Phonemic VFT : letter P</b>			
total words <sup>a</sup>	<b>18.5 (6)</b>	<b>24.16 (6.32)</b>	<b>W = 145.5, p = 0.001</b>
0–30 sec	6.9 (2.7)	9.1 (2.9)	W = 169.5, p = 0.005
30–60 sec	4.2 (1.6)	6.0 (2.0)	W = 159, p = 0.003
30–60 sec	3.8 (2.3)	4.8 (2.3)	W = 234.5, p = 0.13
60–90 sec	3.5 (2)	4.4 (1.6)	W = 229, p = 0.1
word frequency <sup>a</sup>	<b>3.88 (0.27)</b>	<b>3.86 (0.31)</b>	<b>W = 310, p = 0.96</b>
0–30 sec	3.77 (0.85)	3.87 (0.42)	W = 319, p = 0.9
30–60 sec	3.85 (0.94)	3.89 (0.52)	W = 336, p = 0.66
60–90 sec	3.69 (1.03)	3.92 (0.65)	W = 274.5, p = 0.47
90–120 sec	3.51 (1.15)	3.66 (0.72)	W = 314.5, p = 0.97
semantic distance <sup>a</sup>	<b>0.26 (0.03)</b>	<b>0.24 (0.03)</b>	<b>W = 435, p = 0.017</b>
0–30 sec	0.34 (0.11)	0.32 (0.06)	W = 415, p = 0.047
30–60 sec	0.44 (0.17)	0.36 (0.09)	W = 444, p = 0.01
60–90 sec	0.51 (0.24)	0.42 (0.16)	W = 411, p = 0.06
90–120 sec	0.48 (0.25)	0.42 (0.14)	W = 376, p = 0.22
number of switches <sup>a</sup>	<b>7.9 (3.6)</b>	<b>11.56 (5)</b>	<b>W = 174, p = 0.007</b>
mean cluster size <sup>a</sup>	<b>1.2 (0.58)</b>	<b>1.1 (0.55)</b>	<b>W = 330.5, p = 0.73</b>

<sup>a</sup>One missing values per group phonemic VFT (MDD: N = 25, HC: N = 24).

can over and underestimate CT (Dahnke et al., 2013) We adopted a region of interest approach using the Desikan atlas (Desikan et al., 2006) as a common space onto the surface-based CT and central surface maps (originally in the native space) were registered using both a topological defect correction (Yotter et al., 2011) and reparameterization of the surface mesh (Yotter et al., 2011). The resulting CT data were smoothed with a Gaussian kernel of 15 mm full width at half maximum to increase statistical power, given the small sample size (Pardoe et al., 2013; Liem et al., 2015).

### 2.2.5. Fractional anisotropy

All DWI volumes were visually inspected to discard gross abnormalities such as dropouts and interleave artifacts and none were discarded. After brain extraction, we used ExploreDTI (<http://www.exploreDTI.com/>), running on MATLAB 2018b, to preprocess the data. After having created a matrix which incorporates the b-values and the b-vectors, ExploreDTI applies motion and eddy current-induced geometrical distortion corrections by applying an appropriate B matrix rotation (Leemans and Jones, 2009). Finally, the tensor is estimated on a voxelwise basis using a linear least square approach embedded in the informed Robust Extraction of kurtosis INDices with linear estimation (REKINDLE) algorithm (Tax et al., 2015). Informed REKINDLE is an improved function which additionally removes physiological noise in low redundancy datasets (30 directions). Fractional anisotropy (FA) maps are then calculated accordingly (Basser and Pierpaoli, 1996).

The FA maps were processed using tract-based spatial statistics (Smith et al., 2007) part of FSL (Jenkinson et al., 2012). TBSS performs voxel-based statistics projected onto a “skeleton” of the major tracts to avoid partial volume effects and does not imply any a-priori smoothing

step. Each of the 51 (MDD and HC) FA map was registered to each other and the resulting deformations fields were compared to identify the target FA map. The target FA map was non-linearly registered to the MNI152 space at  $1 \times 1 \times 1 \text{ mm}^3$  resolution. Finally, every FA map was non-linearly wrapped by combining the transformation from the subject's FA map to the representative FA map and the transformations from the latest to the MNI152 space. The mean FA image across all participants was computed, “skeletonized” and thresholded at 0.2. TBSS then projected the maximum FA value found along the normal of the skeleton onto it.

### 2.3. Statistical analyses

#### 2.3.1. Clinical and verbal fluency data analyses

We used Wilcoxon two-sample tests for between-group analyses. As qualitative VFT measures are sparsely documented, we used non-parametric Spearman correlations between qualitative VF measures and clinical characteristics, as long with other executive functions in both groups. We quantified the contribution of MDD to number switches during the semantic VFT, as post-hoc analysis, while accounting to other executive functions (abstract reasoning/set shifting (MCST-time), working memory (backward span), cognitive inhibition (Stroop interference score) and cognitive switching (TMT B-A)), using a Poisson regression.

#### 2.3.2. Cortical thickness analyses

We tested for positive and negative associations between CT (within the 68 ROIs (Desikan et al., 2006) and clustering/switching scores after adjusting for age within each group (MDD and HC) using linear regressions with CT as the dependent variable. Between-group comparisons were performed, while accounting for age. Multiple testing correction was controlled using the False Discovery Rate (FDR) (Benjamini and Hochberg, 1995).

#### 2.3.3. Fractional anisotropy analyses

The skelotonised FA data were analyzed using the FSL “Threshold-Free Cluster Enhancement” option in “randomise” with 5000 permutations (Nichols and Holmes, 2002; SMITH and Nichols, 2009) of the maximum statistic of the cluster as a non-parametric approach of multiple testing correction which reduce false positive for cluster inference (Eklund et al., 2016). Tracts were then identified using the Johns-Hopkins University (JHU) white-matter tractography atlas (Hua et al., 2008). In each group (MDD and HC), we tested for positive and negative correlations between voxel-wised FA values and the switching and clustering scores, accounting for age within MDD and HC groups separately. We also performed between-groups comparisons also accounting for age.

#### 2.3.4. Exploratory group by brain interactions

Our primary goal was to identify associations within the MDD group only to derive VF behavioral markers of cerebral physiopathology. However, to determine that our findings are specific to MDD condition, we also conducted exploratory group by brain interactions on semantic switching scores. We performed 25 models testing for group by CT interactions and 7 models testing for group by FA interactions, all accounting for age. False Discovery rates multiple correction were applied.

For all statistical analyses, multiple testing corrected  $p < 0.05$  type 1 error rate was applied.

## 3. Results

### 3.1. Study population

Participants' clinical and demographic characteristics are displayed in Table 1. Of relevance, patients had moderate to severe apathy with high levels of anxiety, without any general cognitive deficits. Groups



**Table 3**

Correlation analysis between VFT scores and other clinical parameters in each group. 3A: Phonemic VFT, 3B: Semantic VFT. Spearman's rho are displayed. Values in bold are significant \*  $p < 0.05$ ; \*\*  $p < 0.005$  †  $p < 0.05$  after FDR multiple testing correction (76 tests for each group and for both phonemic and semantic VFT). MADRS = Montgomery Asberg Depression Scale, MADRS-S: sadness factor; MADRS-NV: neurovegetative factor; MADRS-D: detachment factor; MADRS-NT: negative thinking factor; AES: apathy evaluation scale; STAI-B: State Trait Anxiety Inventory – trait; MDRS Mattis Dementia Rating Scale, TMT: Trail Making Test, DSST: Digit Symbol Substitution Test, MCST: Modified wisconsin Card Sorting Test.

3A Phonemic VF						
	HC group			MDD group		
	Word count	Cluster size	Number of switches	Word count	Cluster size	Number of switches
Age (years)	-0.02	0.1	0.44	0.17	0.25	-0.07
Education (years)	0.34	-0.16	0.34	0.24	-0.32	0.39
MADRS	-0.26	0.11	-0.24	-0.24	0.07	-0.12
• MADRS-S	-0.24	-0.20	-0.04	-0.1	-0.08	0.09
• MADRS-NV	-0.25	-0.07	-0.14	-0.22	0.27	-0.21
• MADRS-D	-0.16	0.3	-0.26	-0.12	0.07	-0.15
• MADRS-NT	0.19	0.07	0.08	0.07	-0.32	0.17
AES	-0.02	0.13	-0.06	<b>-0.49*</b>	-0.02	-0.22
STAI-B	-0.12	-0.11	-0.06	0.06	0.02	0.07
MDRS	-0.14	0.03	-0.05	<b>0.44*</b>	-0.02	0.31
Cluster size	<b>-0.44*</b>	-	<b>-0.85**</b>	-0.21	-	<b>-0.77** †</b>
Number of switches	<b>0.77** †</b>	<b>-0.85** †</b>	-	<b>0.68** †</b>	<b>0.77** †</b>	-
Mean Frequency	<b>-0.74** †</b>	0.05	-0.13	-0.1	<b>0.53**</b>	<b>-0.47*</b>
Mean Semantic distance	-0.08	0.32	<b>-0.63** †</b>	<b>-0.58**</b>	0.27	<b>-0.53*</b>
Forward Span	<b>0.42*</b>	-0.11	0.29	<b>0.49*</b>	0.19	0.2
Backward Span	-0.09	-0.3	0.18	0.34	0.02	0.35
TMT A	-0.15	0.08	-0.14	<b>-0.63**†</b>	0.16	<b>-0.49*</b>
TMT B	-0.38	0.35	<b>-0.42*</b>	<b>-0.51**</b>	0.14	<b>-0.43*</b>
TMT B-A	-0.16	0.39*	-0.27	-0.2	-0.12	-0.12
Stroop color/word	-0.11	0.02	0	<b>0.6**</b>	-0.09	<b>0.49*</b>
Stroop Interference score	0.08	-0.01	-0.22	0.35	0.18	0.18
DSST	0.17	-0.39	0.34	0.47*	-0.14	<b>0.46*</b>
MCST-Time	-0.28	0.33	-0.34	-0.05	0.2	-0.2
MCST-Categories	0.08	-0.19	0.21	0.06	-0.13	0.07
MCST-Errors	0	0.26	-0.11	0.07	0.16	-0.2
MCST-Perseverations	0	-0.22	0.18	-0.13	0.28	-0.15

3B – Semantic VFT						
	HC group			MDD group		
	Word count	Cluster size	Number of switches	Word count	Cluster size	Number of switches
Age (years)	-0.27	-0.07	0.04	-0.007	-0.05	-0.12
Education (years)	0.07	-0.008	0.06	0.14	-0.03	0.14
MADRS	0.29	0.09	0.12	-0.27	0.3	-0.31
MADRS-S	0.03	0.13	-0.06	-0.14	0.15	-0.12
MADRS-NV	<b>0.42*</b>	0.18	0.18	-0.2	0.36	-0.36
MADRS-D	0.3	-0.05	0.13	<b>-0.46*</b>	0.07	-0.34
MADRS-NT	-0.04	-0.33	-0.08	-0.05	0.13	-0.01
AES	-0.06	0.28	-0.13	<b>-0.48*</b>	0.37	<b>-0.60**</b>
STAI-B	0.29	0.003	0.14	0.05	0.08	0.04
MDRS	0.13	-0.15	0.16	<b>0.54**</b>	-0.13	<b>0.54*</b>
Cluster size	-0.25	-	<b>-0.78** †</b>	0.05	-	-0.34
Number of switches	<b>0.7** †</b>	<b>-0.78** †</b>	-	<b>0.86** †</b>	-0.34	-
Mean Frequency	<b>-0.48*</b>	0.39	<b>-0.48*</b>	<b>-0.81** †</b>	-0.004	<b>-0.72** †</b>
Mean Semantic Distance	-0.38	-0.002	-0.2	<b>-0.79** †</b>	0.07	<b>-0.73** †</b>
Forward Span	-0.06	-0.01	0.11	<b>0.43*</b>	0	<b>0.4*</b>
Backward Span	-0.23	0.03	-0.14	<b>0.45*</b>	-0.02	0.36
TMT A	-0.2	0.3	-0.18	<b>-0.41*</b>	0.13	<b>-0.42*</b>
TMT B	-0.02	0.08	-0.08	<b>-0.58**</b>	0.011	<b>-0.47*</b>
TMT B-A	0.02	-0.05	0.04	-0.35	-0.12	-0.2
Stroop color/word	0.26	-0.31	0.25	<b>0.5**</b>	-0.1	<b>0.47*</b>
Stroop Interference score	0.3	-0.16	0.19	0.33	-0.23	<b>0.41*</b>
DSST	-0.03	-0.36	0.13	<b>0.47*</b>	-0.19	<b>0.52**</b>
MCST-Time	0.12	0.05	0.12	-0.22	-0.1	-0.25
MCST-Categories	-0.35	-0.24	-0.44	0.07	0.16	0.07
MCST-Errors	0.24	0.26	0.04	-0.19	-0.24	-0.17
MCST-Perseverations	-0.18	0.35	-0.33	-0.25	-0.1	-0.26

were comparable in terms of age, education and total intra-cranial volumes. [Supplementary Table 1](#) displays the associations between the MDD clinical characteristics and the performances on the executive functions and with the general cognitive functions. [Supplementary Table 2](#) displays the descriptive statistics of the cognitive performances for each group and the between group comparison. Nineteen percent of the MDD sample had working memory (backward span) considered as

outside the norms, 11.5%, 27%, 15.4% had abnormal cognitive switching (TMT B-A), semantic and phonemic VF, respectively.

### 3.2. Between group comparison

#### 3.2.1. Verbal fluency tests

The switching score was significantly reduced in the MDD group,

**Table 4**  
Poisson regression to disentangle the contribution of MDD, abstract reasoning/set shifting (MCST-time), working memory (backward span), cognitive inhibition (Stroop interference score) and cognitive switching (TMT B-A) on number of switches during semantic VFT. <sup>a</sup>HC group considered as reference.

Variables	Beta	95 IC	Exp(Beta)	p-value
MDD Group <sup>a</sup>	-0.2	-0.37 ; -0.03	0.82	<b>0.02</b>
Backward span	-0.01	-0.81 ; 0.07	1	0.8
Stroop interference	0.15	0.004 ; 0.02	1.015	<b>0.008</b>
TM B-A	0	-0.002 ; 0.003	1	0.8
MCST - Time	0	-0.002 ; 0.001	1	0.6

compared with HC in both VFT (Fig. 1). Quantitative (stratified by 30 s time-window) and qualitative indices for both groups and task are displayed in Table 2. Group differences on both VFT are essentially supported by number of switches without any cluster sizes difference. Raw quantitative number of words are smaller in the MDD group, essentially during the first minute in both VFT.

### 3.2.2. Cortical thickness

The comparison between the two groups on whole-brain CT did not reveal any significant differences (FDR corrected).

### 3.2.3. Fractional anisotropy

After FDR correction, we observed significantly lower FA for MDD patients in comparison with HC in widespread bilateral white matter

tracts corresponding to a cluster of 55 533 voxels including: the forceps minor and major, bilateral inferior fronto-occipital fasciculus (IFOF), bilateral uncinate fasciculi (UF), inferior longitudinal fasciculi and superior longitudinal fasciculi (SLF) (Supplementary Fig. 1).

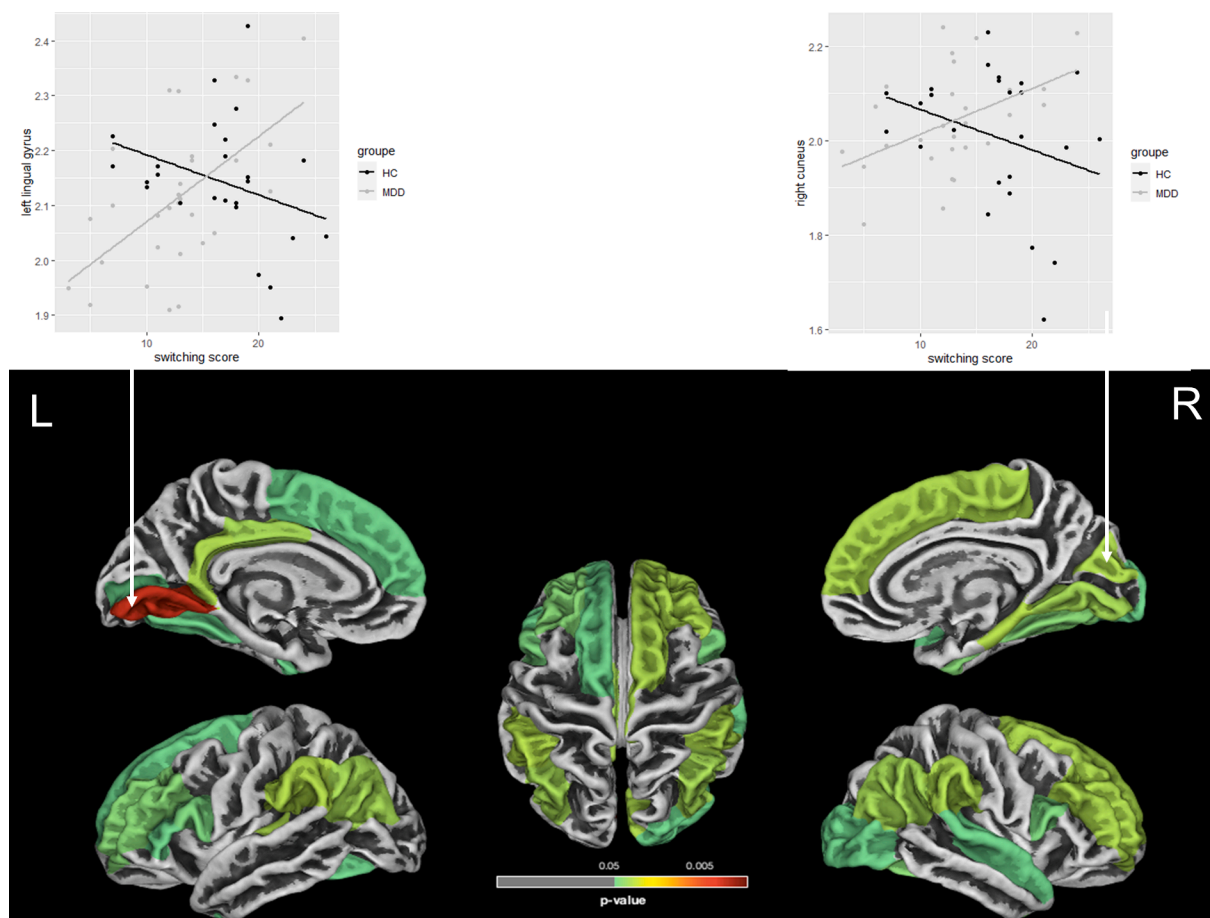
## 4. Clustering and switching scores correlates within MDD and HC groups

### 4.1. Executive functions

Tables 3A and 3B display within-groups clinical and executive correlates of VFT measures for phonemic and semantic VFT, respectively. Both semantic and phonemic number of switches are associated with processing speed (DSST and TMT A and B) and cognitive inhibition. Only number of switches during semantic VFT is associated with AES among MDD patients. Table 4 displays the standardized  $\beta$ , exponentiated  $\beta$  and 95IC of the Poisson regression for each variable. Compared to HC, being part of the MDD group significantly have 18% the number of switches during the semantic VFT. Each increase in Stroop interference score would increase by 1.3% the number of switches during the semantic VFT.

### 4.2. Cortical thickness

In the MDD group, we found significant positive correlations between the switching score in the semantic VFT and CT in 25 regions including the bilateral frontal (inferior, and superior, rostral middle



**Fig. 2.** Regions showing significant positive correlations between switching score during the semantic VFT and CT within the MDD group (age considered as covariate). FDR corrected group by CT interaction plots are also displayed for the left lingual gyrus and the right cuneus (left and right panel, respectively). Color bar depicts the p-value. L: left, R: right. Regions displayed are bilateral lingual gyrus, bilateral superior frontal gyrus, middle frontal gyrus, inferior parietal lobule, transverse and superior temporal and left posterior cingulate.

**Table 5**

Regions showing significant negative associations between the switching score during the semantic VFT and cortical thickness ( $p < 0.05$ , False Discovery Rate (FDR) corrected for multiple comparisons.

Left hemisphere			
P-value (FDR corrected)	T-value	Ze-value	Region
0.003	4.79948	3.95433	Lingual gyrus (BA 18/19)
0.03	3.40297	3.03058	Supramarginal gyrus (BA 40)
0.03	2.89176	2.64258	Isthmus of cingulate gyrus (BA 26/29/30)
0.03	2.88486	2.63715	Transverse Temporal gyrus (BA 41/42)
0.032	2.79163	2.56333	Posterior Cingulate (BA 24)
0.032	2.72878	2.51304	Inferior parietal lobule (BA 40/39)
0.039	2.56065	2.37646	Rostral Middle frontal gyrus (BA 10)
0.043	2.46341	2.29611	Pars Opercularis (BA 44)
0.047	2.32776	2.18239	Superior Frontal gyrus (BA 6)
0.047	2.32177	2.17732	Fusiform gyrus (BA 37)
0.048	2.29225	2.1523	Pars Triangularis (BA 45)
0.048	2.26827	2.13191	Pericalcarine sulcus (BA 17)
Right hemisphere			
P-value (FDR corrected)	T-value	Ze-value	Region
0.029	3.49092	3.09454	Parahippocampal gyrus (BA 28/34/35/36)
0.03	3.27641	2.9371	Supramarginal gyrus (BA 40)
0.03	3.15651	2.84699	Paracentral lobule
0.03	3.01009	2.73486	Superior Frontal gyrus (BA 6)
0.03	2.91241	2.65879	Cuneus (BA 17)
0.03	2.90005	2.6491	Rostral Middle frontal gyrus (BA 10)
0.032	2.78501	2.55805	Lingual gyrus (BA 18/19)
0.032	2.71204	2.49957	Inferior parietal lobule (BA 40/39)
0.039	2.56991	2.38406	Fusiform gyrus (BA 37)
0.043	2.4661	2.29834	Pars Opercularis (BA 44)
0.047	2.37309	2.2206	Lateral Occipital (BA 19)
0.047	2.34753	2.19907	Transverse Temporal (BA 41)
0.047	2.30821	2.16583	Superior Temporal (BA 22)

gyri), parietal (inferior parietal lobule including the supramarginal gyri), temporal (transverse and fusiform gyri) and occipital (pericalcarine and lingual gyri) areas. There were no significant associations between CT and switching score during the semantic VFT within the HC group (all  $p > 0.05$  uncorrected). Only the right cuneus ( $t = 3$ ,  $\beta = 32$ ,  $p = 0.047$ ) and the left lingual gyrus ( $t = 3.5$ ,  $\beta = 36$  Ashburner et al., 2003,  $p = 0.02$ ) showed significant group by CT interaction after FDR correction (Fig. 2 and Table 5). There was no association between the clustering score in the semantic VFT and CT neither in the MDD nor the HC groups. There was no association between any of the phonemic VFT qualitative measures and CT neither in the MDD nor the HC groups.

#### 4.3. Fractional anisotropy

In the MDD group, we found positive correlation between FA values and the switching score in semantic VFT in the MDD group within 7 clusters of 5706 (right superior longitudinal fasciculus extending to the right arcuate fasciculus to the body of the CC (bilateral) and to the right inferior fronto-occipital fasciculus, 134 (the right anterior corpus callosum (CC) extending to the right anterior corona radiata), 74, 33, 11, 7 and 3 voxels including: the right anterior corpus callosum (CC) extending to the right anterior corona radiata, right IFOF, right SLF extending to the anterior thalamic radiation (ATR) (FDR corrected) (Fig. 3 and Table 6). Only voxels in the right superior longitudinal fasciculus, right arcuate, right anterior part of the CC and anterior corona radiata showed significant interaction group by FA interaction, after FDR correction of multiple testing (Table 6). After FDR correction, we did not find any significant correlation between mean FA and the clustering score in the semantic VFT neither in the MDD nor the HC

groups.

There was no association between either score in the phonemic VFT and mean FA neither in the MDD nor the HC groups.

## 5. Discussion

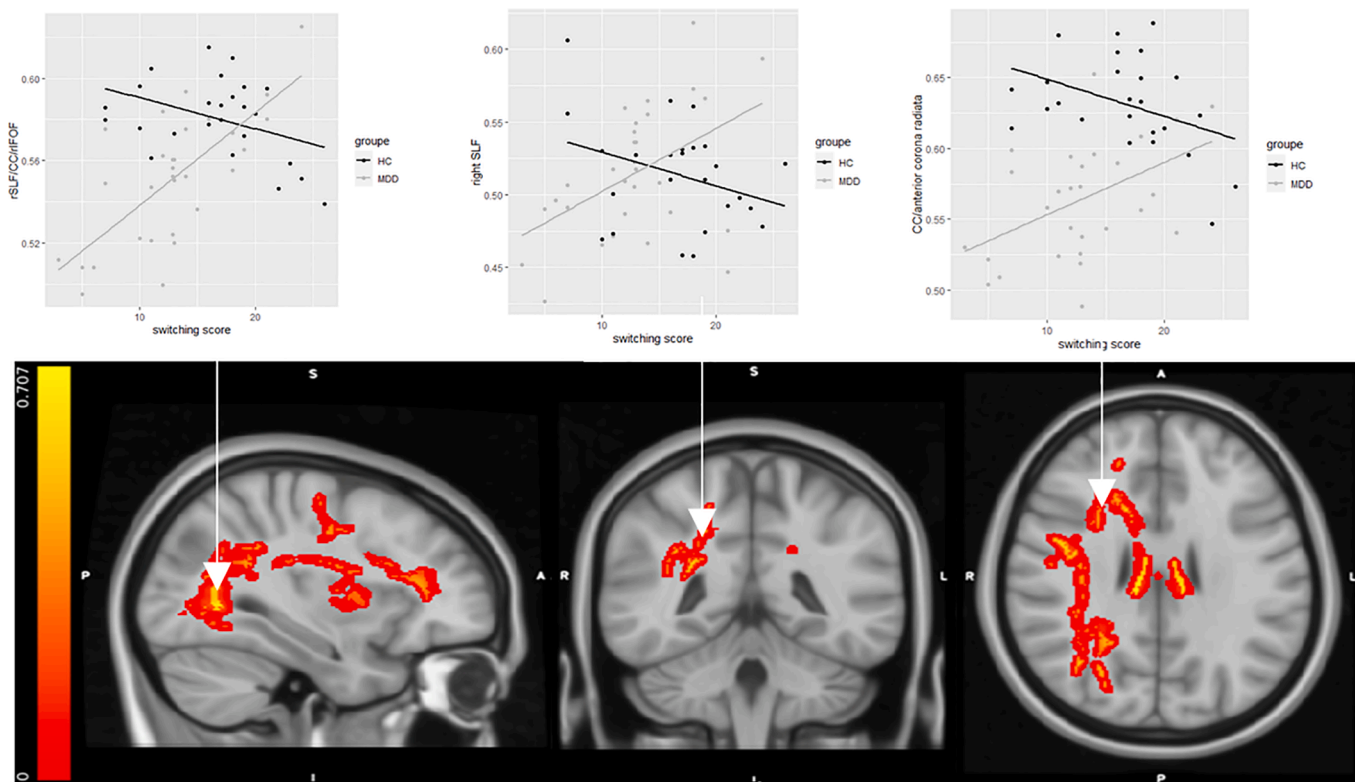
Here, we found that the semantic switching score correlated with CT in various regions distributed bilaterally in the frontal (superior, rostral middle and inferior gyri), parietal (inferior parietal lobule including the supramarginal gyri), temporal (transverse and fusiform gyri) and occipital (lingual gyri) lobes. After FDR correction, only the right cuneus and the left lingual gyrus showed similar pattern of group by cortical thickness interaction. This switching score also correlated with FA in the CC, right SLF extending to the ATR, and IFOF. Only the right superior longitudinal fasciculus, right arcuate, right anterior part of the CC showed group by FA interaction. Our results revealed that MDD subjects had decreased switching scores in both VFT compared to HC.

Our results are consistent with our a-priori hypothesis of thinner cortices among prefrontal regions, supported by recent results which investigated CT basis of semantic fluency in a large sample ( $n = 505$ ) of healthy older participants and found positive correlations with quantitative semantic fluency within the left superior, rostral middle and inferior frontal gyri (Vonk et al., 2019). The left superior, rostral middle and inferior frontal gyri, left supramarginal gyri, left fusiform, left transverse temporal gyri, and left lingual gyrus were also involved, partially overlapping with our results. An increased activation within the middle and inferior frontal gyri, and bilateral parietal cortex (superior and inferior parietal lobule) was also found during both self-reported switching in comparison to constrained switching VFT among healthy individuals (Hirshorn and Thompson-Schill, 2006). Also, cognitive switching with reduced working memory component recruits both the middle frontal gyri along with bilateral inferior parietal lobule and superior/transverse temporal gyri, in line with the current results (Smith et al., 2004).

Switching consists in shifting between clusters when one cluster is exhausted (Trojer et al., 1997). It requires: (i) to identify all the category-relevant items within the semantic memory, and (ii) to retrieve and select the words accordingly to the recommendation (i.e. “animal” or “letter P”) which involves attention and working memory (Ralph et al., 2017). There is substantial evidence that those cognitive processes are supported by different and partially overlapping neural networks (Ralph et al., 2017).

(i) The organization of the semantic memory has been described by the prominent hub-and-spoke theory (Patterson et al., 2007): the knowledge of a semantic concept (e.g. cat) is stored both in a “semantic hub”, and in distributed cortical regions related to its sensorimotor attributes (e.g. shape, name, colors, motion, sound, function) called “the spokes”. The “hub” is assumed to integrate inputs from activated sensorimotor feature sets, categorize, for example, both cat and fish as “animals”. This hub is suggested to be localized in the bilateral anterior temporal lobes (ATL), including the fusiform gyrus (Pobric et al., 2010). The semantic spokes are suggested to be supported by primary and associative auditory (bilateral transverse temporal gyri), motor (the supplementary motor area/paracentral lobule) and visual areas (lingual gyri, middle and lateral occipital gyri and cuneus) cortices. It has been described that the categories of concepts can be dependent on a particular modality. For example, whereas “tools” can be individuated by their associated actions (e.g. hammer/hitting), “animals” are rather individuated by virtue of their constituent sensory features (e.g. shape, colors). Therefore, naming “animals” might recruit greater responses in visual areas compared to naming “tools” (Chouinard and Goodale, 2010).

(ii) Semantic VFT implies a controlled retrieval and selection of the semantic information, also known as “semantic control”, which has been suggested to rely on inferior frontal gyri (Ralph et al., 2017). More precisely, it seems that the most posterior part of the inferior frontal gyri



**Fig. 3.** Fasciculi showing positive correlation between semantic switching score and FA in MDD group (5000 permutations), along with the significant (FDR corrected) group by FA interaction plots (accounting for age) for the cluster including the right SLF, the right IFOF and the body of the CC (5706 voxels, left panel); the right SLF (134 voxels, middle panel) and the right anterior part of the CC extending to the anterior corona radiata (74 voxels, right panel). Clusters are overlaid on a canonical template provided by FSLeves (<https://git.fmrib.ox.ac.uk/fsl/fsleves/fsleves/>) (1 mm resolution).

(i.e. the pars opercularis and pars triangularis) is more critically involved in the selection process (Badre et al., 2005). While inferior frontal gyri seems to be critical, semantic control also relies on posterior middle temporal gyri, the inferior parietal lobule and the intraparietal sulcus within a “semantic control network” (Ralph et al., 2017).

Secondly, there is growing evidence that white matter tracts are involved in quantitative VFT processing. For example, poor quantitative VFT scores have been associated with greater left SLF damage in a study which used voxel-based lesion symptom mapping in patients with penetrating traumatic brain injury (Cristofori et al., 2015). Both mean FA value and lesion of the left SLF, the left IFOF, and left ATR were correlated with the score in semantic VFT in stroke patients (Li et al., 2017). Finally, associations between FA along the IFOF and semantic VFT scores were observed in patients with left diffuse low-grade glioma (Almairac et al., 2015). Although these previous findings align with our current results, they do not address their specific cognitive function in the same way as qualitative VFT scores. Controlled retrieval, selection in accordance with semantic criteria processes have not been associated with white matter MRI-derived measures. Nevertheless, the SLF, the IFOF and the ATR have been associated with other cognitive processes required during semantic switching. For example, the right SLF and IFOF have been related with semantic categorization performance (i.e. the mental operation by which the brain classifies objects and events) in a voxel-based morphometry in healthy subjects (Garcin et al., 2018). The FA of SLF and IFOF have also been found to be positively correlated with processing speed in healthy older adult (Kerchner et al., 2012). A positive association was also found between verbal working memory performance assessed by the letter-number span task and FA in bilateral SLF young healthy subjects (Peters et al., 2012).

Inspection of the interaction plots (Figs. 2 & 3) suggests that for all regions (for either the CT or the FA analyses) there might be a negative correlation between the semantic VF switching score and the MRI-

derived brain features within the HC group. This is quite unexpected since the literature suggests that better cognitive functions are associated with thicker cortices and higher FA. We therefore explored if these associations were statistically significant and none of them were (all  $p > 0.05$ , without any correction for multiple testing). Adding to the fact that (i) none of the whole brain analyses in the HC group revealed significant associations and (ii) that the regions where the interaction terms were statistically significant are very similar to the ones which showed the greater effect size within the MDD group (left lingual gyrus and right SLF, CC and arcuate), it suggests that the interaction terms are related to the significant associations within the MDD group and not to unexpected associations within the HC group. This reinforces and increases our confidence in the specificity of our results within the MDD group.

In the third place, we investigated both CT and FA correlates of switching and clustering scores in phonemic VFT. We did not find any significant association between those scores and the CT, neither the FA. We postulate here that this could be due to a weaker association between the CT and the phonemic performances than with the semantic ones, as previously found in old-age population (Vonk et al., 2019). In addition, the implication of the pars triangularis in both semantic and phonemic performances has been described (Costafreda et al., 2006; Wagner et al., 2014). Focusing on this cortical area, we ran additional post-hoc analysis to explore its association with the semantic and phonemic quantitative scores in the HC (Supplementary Table 3). We retrieved that the CT of the left pars triangularis is associated with semantic and phonemic fluency quantitative scores (effect size:  $-0.39$  and  $-0.06$  respectively), suggesting that we might suffer from low power to reveal significant associations between the brain features and phonemic qualitative measures.

Additionally, to explore the specificity of the association between the qualitative VFT score and the measures of the FA and the CT, we ran additional post-hoc analysis in each group. We tested for associations



**Table 6**

Significant clusters showing associations between the switching score during the semantic VFT and FA within the MDD group (TFCE, 5000 permutations): cluster size, corresponding tracts. Group by FA interactions statistics (t, estimate and 95 confidence interval and false discovery rate p-values) are displayed along with the corresponding exploratory analyses within the healthy control group. CC: Corpus callosum, IFOF: inferior fronto-occipital fasciculus, MLF: Middle Longitudinal Fasciculus, SLF: Superior Longitudinal Fasciculus; 95CI: 95% confidence interval; FDR: False Discovery Rate.

Cluster size (voxels)	Tracts	Group by FA Interaction on semantic switching score (t, $\beta$ [95CI])	FDR corrected p-values	Healthy control group correlation with semantic switching score (FDR corrected p-value)
5706	Right SLF Right MLF Bilateral body CC Right IFOF Right optic radiation	T = 4.18, $\beta$ = 210 [109–311]	0.0008190	T = -2.4, $\beta$ = -134 [-252–-16], p = 0.17
134	Right anterior body CC Right corona radiata	T = 3.4, $\beta$ = 123.4 [51–195]	0.0043050	T = -2.1, $\beta$ = -63 [-125–-0.5], p = 0.17
74	Right SLF	T = 3, $\beta$ = 99 [32–165]	0.0095900	T = -1.6, $\beta$ = -47 [-106–12], p = 0.2
33	Right cortico-spinal tract	T = 0.9, $\beta$ = 31 [-41–104]	0.4515000	T = 0.7, $\beta$ = 24 [-45–92], p = 0.5
11	Right anterior corona radiata	T = 2.6, $\beta$ = 80 [18–142]	0.0216125	T = -1.6, $\beta$ = -37 [-88–14], p = 0.2
7	Right superior thalamic radiation	T = 2.3, $\beta$ = 39 [5–73]	0.0333200	T = -0.5, $\beta$ = -7 [-35–20], p = 0.6
3	Body CC	T = 0.2, $\beta$ = 5 [-56–67]	0.8621000	T = 1.7, $\beta$ = 43 [-9–94], p = 0.2

between the standard VFT score (i.e. the quantitative score) and the CT, and the FA. In both groups, the quantitative score was not significantly associated with the CT or the FA ( $p < 0.05$  corrected). Moreover, we also tested whether the TMT B-A score, which reflect cognitive switching, is associated with CT and FA among the MDD and we could not find any significant results.

These post-hoc exploratory analyses suggest that i) the associations of MRI-derived brain features measures are greater for the automated qualitative indices of semantic VFT than raw quantitative word count. ii) It also suggests that semantic VF switching is more specific to robust anatomic MRI measures than other measures of cognitive switching, in MDD.

As expected, we found that MDD participants performed fewer switches than HC with similar clustering sizes, consistent with executive dysfunctions in MDD without impaired lexico-semantic stocks (Fossati et al., 2003). In this sample, we found moderate to strong correlations between number of switching, processing speed (DSST), attention (TMT-A) and cognitive inhibition and apathy (dor semantic VFT) suggesting shared variance, reinforcing our view that switching during VFT relies on executive functions. Future studies with larger samples might apply data reduction methods to capture the shared variance of cognitive, motivational and possibly other features set and test for cerebral biomarkers. However, data reduction approaches render the interpretation of latent variables difficult to interpret. Our purpose here was rather to test whether automated qualitative indices of VFT could be possible biomarkers for MDD, because easily doable, interpretable and, if replicated, linkable to grey and white matter brain features. As post-hoc

analyses, we also found that while accounting for abstract reasoning/set shifting, working memory, cognitive inhibition and switching, MDD decreases estimation of number of switches during semantic VFT by 18%, suggesting that MDD specifically affects VFT switching, above and beyond other executive functions (Tariot, 1986).

We found reduced FA in CC, bilateral SLF, IFOF, uncinate and the ATR, which concurs with recent results in large sample of MDD, where greater effect sizes were observed among recurrent depression, as in the current sample (van Velzen et al., 2020). In contrast, there were no significant differences in CT between the MDD and the HC groups. This might be due to low power given the small sample size and small effect sizes recently shown when comparing CT between MDD and HC (Schmaal et al., 2017).

## 6. Limitations

The present study had several potential limitations. We only included women in our protocol. This choice was guided by the suggested existence of sex differences in the subtypes and course of MDD (Kuehner, 2017). Sex-related CT (Sowell et al., 2007) and FA (van Hemmen et al., 2016) differences have also been demonstrated, as well as an interaction between sex and performance level in VFT (Scheuringer et al., 2017). Nevertheless, unique gender as well as small sample sizes might reduce the generalizability of the findings. We did not use a standardized task to measure language ability. However, it is unlikely that our findings are related to broad language ability for several reasons. All participants were French native speakers, middle aged and free of cognitive disorders. Moreover, both groups were matched on education level (a strong predictor of language ability). Also, we used the “animal” category of the semantic task, which reduce the impact of education on the performance (Ardila et al. 2016). Finally, the cluster score (which reflects the storage of the lexico-semantic memory) was similar between the groups and not associated with any cerebral measures. Additionally, we did not control for type of treatment in our MDD group. Although previous studies did not find any influence of antidepressants on executive function (Killian et al., 1984), there is some evidence that drugs with an anticholinergic effect may have a cognitive impact (Orzechowska et al., 2015). Antidepressant may also have influenced the CT measurements (Schmaal et al., 2017) and the white matter microstructure as well (Chouiter et al., 2016). We used a SBM technique whereas it has been suggested that left basal ganglia sustain initiation abilities in VFT tasks (van Velzen et al., 2020). While tensor-derived measures of white matter integrity might lack specificity, it is considered as a reasonable measure of axonal density when tracts presumably contain single fiber population, such as long antero-posterior tracts (SLF and the IFOF) (De Santis et al., 2014).

## 7. Conclusion

We elucidated specific cortical and FA signature of switching performance in semantic VFT in a sample of depressed women. Those results provide additional insight about the cognitive impairment which underly VFT deficit in MDD. Moreover, automatic qualitative VFT scores are associated with brain anatomy and reinforces its potential use as a surrogate for depression cerebral bases. Given the brevity and cost profile of the semantic VFT, its use in day-to-day clinical decision making could be recommended.

## Funding

This work was supported by the Fondation de l’Avenir (ET1-628).

## CRedit authorship contribution statement

**L. Domain:** Data curation, Formal analysis, Writing – original draft. **M. Guillery:** Methodology, Validation, Writing – review & editing. N.

**Linz:** Conceptualization, Data curation, Formal analysis, Methodology, Writing – review & editing. **A. König:** Data curation, Formal analysis, Methodology, Writing – review & editing. **J.M. Batail:** Data curation, Writing – review & editing. **R. David:** Methodology, Writing – review & editing. **I. Corouge:** Data curation, Writing – review & editing. **E. Bannier:** Data curation, Writing – review & editing. **J.C. Ferré:** Data curation, Writing – review & editing. **T. Dondaine:** Data curation, Writing – review & editing. **D. Drapier:** Conceptualization, Project administration, Supervision, Funding acquisition, Writing – review & editing. **G.H. Robert:** Conceptualization, Funding acquisition, Formal analysis, Project administration, Writing – review & editing.

## Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

## Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.nicl.2021.102910>.

## References

- James, S.L., Abate, D., Abate, K.H., Abay, S.M., Abbafati, C., Abbasi, N., et al., 2018. Global, regional, and national incidence, prevalence, and years lived with disability for 354 diseases and injuries for 195 countries and territories, 1990–2017: a systematic analysis for the Global Burden of Disease Study 2017. *Lancet* 392 (10159), 1789–1858.
- Desai, R., Liebenthal, E., Possing, E.T., Waldron, E., Binder, J.R. 2005. Volumetric vs. surface-based alignment for localization of auditory cortex activation. *Neuroimage* 26(4), 1019–1029.
- Greve, D.N., Fischl, B., 2018. False positive rates in surface-based anatomical analysis. *Neuroimage* 1 (171), 6–14.
- Lerch, J.P., Evans, A.C., 2005. Cortical thickness analysis examined through power analysis and a population simulation. *Neuroimage* 24 (1), 163–173.
- Lerch, J.P., Carroll, J.B., Dorr, A., Spring, S., Evans, A.C., Hayden, M.R., Sled, J.G., Henkelman, R.M., 2008. Cortical thickness measured from MRI in the YAC128 mouse model of Huntington's disease. *NeuroImage* 41 (2), 243–251.
- Schwarz, C.G., Gunter, J.L., Wiste, H.J., Przybelski, S.A., Weigand, S.D., Ward, C.P., Senjem, M.L., Vemuri, P., Murray, M.E., Dickson, D.W., Parisi, J.E., Kantarci, K., Weiner, M.W., Petersen, R.C., Jack, C.R., 2016. A large-scale comparison of cortical thickness and volume methods for measuring Alzheimer's disease severity. *Neuroimage Clin.* 11, 802–812.
- Schmaal, L., Hibar, D.P., Sämann, P.G., Hall, G.B., Baune, B.T., Jahanshad, N., Cheung, J. W., van Erp, T.G.M., Bos, D., Ikram, M.A., Vernooij, M.W., Niessen, W.J., Tiemeier, H., Hofman, A., Wittfeld, K., Grabe, H.J., Janowitz, D., Bülow, R., Selonke, M., Völzke, H., Grotegerd, D., Dannlowski, U., Arolt, V., Opel, N., Heindel, W., Kugel, H., Hoehn, D., Czisch, M., Couvy-Duchesne, B., Rentería, M.E., Strike, L.T., Wright, M.J., Mills, N.T., de Zubicaray, G.L., McMahon, K.L., Medland, S. E., Martin, N.G., Gillespie, N.A., Goya-Maldonado, R., Gruber, O., Krämer, B., Hatton, S.N., Lagopoulos, J., Hickie, I.B., Frodl, T., Carballo, A., Frey, E.M., van Velzen, L.S., Penninx, B.W.J.H., van Tol, M.-J., van der Wee, N.J., Davey, C.G., Harrison, B.J., Mwangi, B., Cao, B., Soares, J.C., Veer, I.M., Walter, H., Schoepf, D., Zurovski, B., Konrad, C., Schramm, E., Normann, C., Schnell, K., Sacchet, M.D., Gotlib, I.H., MacQueen, G.M., Godlewski, B.R., Nickson, T., McIntosh, A.M., Papey, M., Whalley, H.C., Hall, J., Sussmann, J.E., Li, M., Walter, M., Aftanas, L., Brack, I., Bokhan, N.A., Thompson, P.M., Veltman, D.J., 2017. Cortical abnormalities in adults and adolescents with major depression based on brain scans from 20 cohorts worldwide in the ENIGMA Major Depressive Disorder Working Group. *Mol. Psychiatry* 22 (6), 900–909.
- Jiang, W., Andreassen, O.A., Agartz, I., Lagerberg, T.V., Westlye, L.T., Calhoun, V.D., Turner, J.A., 2020. Distinct structural brain circuits indicate mood and apathy profiles in bipolar disorder. *NeuroImage: Clin.* 26, 101989. <https://doi.org/10.1016/j.nicl.2019.101989>.
- van Velzen, L.S., Kelly, S., Isaev, D., Aleman, A., Aftanas, L.I., Bauer, J., et al., 2020. White matter disturbances in major depressive disorder: a coordinated analysis across 20 international cohorts in the ENIGMA MDD working group. *Mol. Psychiatry*. 25 (7), 1511–1525.
- Ruggeri, B., Sarkans, U., Schumann, G., Persico, A.M., 2014. Biomarkers in autism spectrum disorder: the old and the new. *Psychopharmacology*. 231 (6), 1201–1216.
- Snyder, H.R., 2013. Major depressive disorder is associated with broad impairments on neuropsychological measures of executive function: a meta-analysis and review. *Psychol. Bull.* 139 (1), 81–132.
- Fossati, P., Bastard Guillaume, L.e., Ergis, A.-M., Allilaire, J.-F., 2003. Qualitative analysis of verbal fluency in depression. *Psychiatry Res.* 117 (1), 17–24.
- Troyer, A.K., Moscovitch, M., Winocur, G., 1997. Clustering and switching as two components of verbal fluency: evidence from younger and older healthy adults. *Neuropsychology* 11 (1), 138–146.
- Pimontel, M.A., Rindskopf, D., Rutherford, B.R., Brown, P.J., Roose, S.P., Sneed, J.R., 2016. A meta-analysis of executive dysfunction and antidepressant treatment response in late-life depression. *Am. J. Geriatric Psychiatry*. 24 (1), 31–41.
- Nicklas Linz, Johannes Tröger, Jan Alexandersson, Alexandra König. Using Neural Word Embeddings in the Analysis of the Clinical Semantic Verbal Fluency Task. IWCS 2017 - 12th International Conference on Computational Semantics, Sep 2017, Montpellier, France. Using Neural Word Embeddings in the Analysis of the Clinical Semantic Verbal Fluency Task. In: IWCS 2017 - 12th International Conference on Computational Semantics, Sep 2017, Montpellier, France. 2017. p. pp.1-7. hal-01672593.
- Linz, N., Tröger, J., Alexandersson, J., Wolters, M., König, A., Robert, P., 2017. Predicting Dementia Screening and Staging Scores from Semantic Verbal Fluency Performance. In: In: 2017 IEEE International Conference on Data Mining Workshops (ICDMW), pp. 719–728.
- Tröger J, Linz N, Alexandersson J, König A, Robert P. Automated speech-based screening for alzheimer's disease in a care service scenario. In: Proceedings of the 11th EAI International Conference on Pervasive Computing Technologies for Healthcare [Internet]. New York, NY, USA: Association for Computing Machinery; 2017 [cited 2021 Jul 25]. p. 292–7. (PervasiveHealth '17). Available from: <https://doi.org/10.1145/3154862.3154915>.
- König, A., Satt, A., Sorin, A., Hoory, R., Toledo-Ronen, O., Derreumaux, A., Manera, V., Verhey, F., Aalten, P., Robert, P.H., David, R., 2015. Automatic speech analysis for the assessment of patients with pre-dementia and Alzheimer's disease. *Alzheimers Dement (Amst)*. 1 (1), 112–124.
- Yuan, P., Raz, N., 2014. Prefrontal cortex and executive functions in healthy adults: a meta-analysis of structural neuroimaging studies. *Neurosci. Biobehav. Rev.* 42, 180–192.
- Vonk, J.M.J., Rizvi, B., Lao, P.J., Budge, M., Manly, J.J., Mayeux R, et al. 2019. Letter and category fluency performance correlates with distinct patterns of cortical thickness in older adults. *Cereb. Cortex* 29(6), 2694–700.
- Hoagey, D.A., Lazarus, L.T.T., Rodrigue, K.M., Kennedy, K.M., 2021. The effect of vascular health factors on white matter microstructure mediates age-related differences in executive function performance. *Cortex*. 1 (141), 403–420.
- Garcin, B., Urbanski, M., Thiebaut de Schotten, M., Levy, R., Volle, E., 2018. Anterior temporal lobe morphometry predicts categorization ability. *Front. Hum. Neurosci.* 7 (12), 36.
- Li, M., Zhang, Y., Song, L., Huang, R., Ding, J., Fang, Y., Xu, Y., Han, Z., 2017. Structural connectivity subserving verbal fluency revealed by lesion-behavior mapping in stroke patients. *Neuropsychologia*. 101, 85–96.
- Quilty, L.C., Robinson, J.J., Rolland, J.-P., Fruyt, F.D., Rouillon, F., Bagby, R.M., 2013. The structure of the Montgomery-Åsberg depression rating scale over the course of treatment for depression. *Int. J. Methods Psychiatr. Res.* 22 (3), 175–184.
- Tombaugh, T., 1999. Normative data stratified by age and education for two measures of verbal fluency FAS and animal naming. *Arch. Clin. Neuropsychol.* 14 (2), 167–177.
- Grave Edouard, Bojanowski et al. 2018. Learning Word Vectors for 157 Languages. In: Proceedings of the International Conference on Language Resources and Evaluation (LREC 2018).
- Ashburner, J., Friston, K.J., 2000. Voxel-based morphometry—The methods. *NeuroImage* 11 (6), 805–821.
- Gaser, C., Dahnke, R., 2016. CAT - A Computational Anatomy Toolbox for the Analysis of Structural MRI Data. HBM 2016 [Internet]. Available from: <http://www.neuro.uni-jena.de/hbm2016/GaserHBM2016.pdf>.
- Dahnke, R., Yotter, R.A., Gaser, C., 2013. Cortical thickness and central surface estimation. *Neuroimage* 15 (65), 336–348.
- Desikan, R.S., Ségonne, F., Fischl, B., Quinn, B.T., Dickerson, B.C., Blacker, D., Buckner, R.L., Dale, A.M., Maguire, R.P., Hyman, B.T., Albert, M.S., Killiany, R.J., 2006. An automated labeling system for subdividing the human cerebral cortex on MRI scans into gyral based regions of interest. *NeuroImage* 31 (3), 968–980.
- Yotter, R.A., Dahnke, R., Thompson, P.M., Gaser, C., 2011. Topological correction of brain surface meshes using spherical harmonics. *Hum. Brain Mapp.* 32 (7), 1109–1124.
- Yotter, R.A., Thompson, P.M., Gaser, C., 2011. Algorithms to improve the reparameterization of spherical mappings of brain surface meshes. *J. Neuroimaging*. 21 (2), e134–e147.
- Pardoe, H.R., Abbott, D.F., Jackson, G.D., 2013. Alzheimer's Disease Neuroimaging Initiative. Sample size estimates for well-powered cross-sectional cortical thickness studies. *Hum. Brain Mapp.* 34 (11), 3000–3009.
- Liem, F., Méritat, S., Bezzola, L., Hirsiger, S., Philipp, M., Madhyastha, T., Jäncke, L., 2015. Reliability and statistical power analysis of cortical and subcortical FreeSurfer metrics in a large sample of healthy elderly. *Neuroimage* 108, 95–109.
- Leemans, A., Jones, D.K., 2009. The B-matrix must be rotated when correcting for subject motion in DTI data. *Magn. Reson. Med.* 61 (6), 1336–1349.
- Tax, C.M.W., Otte, W.M., Viergever, M.A., Dijkhuizen, R.M., Leemans, A., 2015. REKINDLE: robust extraction of kurtosis INDices with linear estimation. *Magn. Reson. Med.* 73 (2), 794–808.
- Basser, P.J., Pierpaoli, C., 1996. Microstructural and physiological features of tissues elucidated by quantitative-diffusion-tensor MRI. *J. Magn. Reson. B.* 111 (3), 209–219.
- Smith, S.M., Johansen-Berg, H., Jenkinson, M., Rueckert, D., Nichols, T.E., Miller, K.L., Robson, M.D., Jones, D.K., Klein, J.C., Bartsch, A.J., Behrens, T.E.J., 2007. Acquisition and voxelwise analysis of multi-subject diffusion data with tract-based spatial statistics. *Nat. Protoc.* 2 (3), 499–503.

- Jenkinson, M., Beckmann, C.F., Behrens, T.E.J., Woolrich, M.W., Smith, S.M. 2012. FSL. *Neuroimage* 62(2), 782–790.
- Benjamini, Y., Hochberg, Y., 1995. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J. Roy. Stat. Soc.: Ser. B (Methodol.)* 57 (1), 289–300.
- Ashburner, J., Csernansky, J.G., Davatzikos, C., Fox, N.C., Frisoni, G.B., Thompson, P.M., 2003. Computer-assisted imaging to assess brain structure in healthy and diseased brains. *Lancet Neurol.* 2 (2), 79–88.
- Nichols, T.E., Holmes, A.P., 2002. Nonparametric permutation tests for functional neuroimaging: a primer with examples. *Hum. Brain Mapp.* 15 (1), 1–25.
- Smith, S., Nichols, T., 2009. Threshold-free cluster enhancement: addressing problems of smoothing, threshold dependence and localisation in cluster inference. *NeuroImage*. 44 (1), 83–98.
- Eklund, A., Nichols, T.E., Knutsson, H., 2016. Cluster failure: why fMRI inferences for spatial extent have inflated false-positive rates. *Proc. Natl. Acad. Sci. USA* 113 (28), 7900–7905.
- Hua, K., Zhang, J., Wakana, S., Jiang, H., Li, X., Reich, D.S., Calabresi, P.A., Pekar, J.J., van Zijl, P.C.M., Mori, S., 2008. Tract probability maps in stereotaxic spaces: analyses of white matter anatomy and tract-specific quantification. *NeuroImage* 39 (1), 336–347.
- Hirshorn, E.A., Thompson-Schill, S.L., 2006. Role of the left inferior frontal gyrus in covert word retrieval: neural correlates of switching during verbal fluency. *Neuropsychologia* 44 (12), 2547–2557.
- Smith, A.B., Taylor, E., Brammer, M., Rubia, K., 2004. Neural correlates of switching set as measured in fast, event-related functional magnetic resonance imaging. *Hum. Brain Mapp.* 21 (4), 247–256.
- Ralph, M.A.L., Jefferies, E., Patterson, K., Rogers, T.T., 2017. The neural and computational bases of semantic cognition. *Nat. Rev. Neurosci.* 18 (1), 42–55.
- Patterson, K., Nestor, P.J., Rogers, T.T., 2007. Where do you know what you know? The representation of semantic knowledge in the human brain. *Nat. Rev. Neurosci.* 8 (12), 976–987.
- Pobric, G., Jefferies, E., Lambon Ralph, M.A., 2010. Amodal semantic representations depend on both anterior temporal lobes: evidence from repetitive transcranial magnetic stimulation. *Neuropsychologia* 48 (5), 1336–1342.
- Chouinard, P.A., Goodale, M.A., 2010. Category-specific neural processing for naming pictures of animals and naming pictures of tools: An ALE meta-analysis. *Neuropsychologia*. 48 (2), 409–418.
- Badre, D., Poldrack, R.A., Paré-Blagoev, E.J., Inslar, R.Z., Wagner, A.D., 2005. Dissociable controlled retrieval and generalized selection mechanisms in ventrolateral prefrontal cortex. *Neuron* 47 (6), 907–918.
- Cristofori, I., Zhong, W., Chau, A., Solomon, J., Krueger, F., Grafman, J., 2015. White and gray matter contributions to executive function recovery after traumatic brain injury. *Neurology* 84 (14), 1394–1401.
- Almairac, F., Herbet, G., Moritz-Gasser, S., de Champfleure, N.M., Duffau, H., 2015. The left inferior fronto-occipital fasciculus subserves language semantics: a multilevel lesion study. *Brain Struct. Funct.* 220 (4), 1983–1995.
- Kerchner, G.A., Racine, C.A., Hale, S., Wilhelm, R., Laluz, V., Miller, B.L., et al. 2012. Cognitive Processing Speed in Older Adults: Relationship with White Matter Integrity. Stamatakis EA, editor. *PLoS One* 7(11):e50425.
- Peters, B.D., Szeszko, P.R., Radua, J., Ikuta, T., Gruner, P., DeRosse, P., Zhang, J.-P., Giorgio, A., Qiu, D., Tapert, S.F., Brauer, J., Asato, M.R., Khong, P.L., James, A.C., Gallego, J.A., Malhotra, A.K., 2012. White matter development in adolescence: diffusion tensor imaging and meta-analytic results. *Schizophr. Bull.* 38 (6), 1308–1317.
- Costafreda, S.G., Fu, C.H.Y., Lee, L., Everitt, B., Brammer, M.J., David, A.S., 2006. A systematic review and quantitative appraisal of fMRI studies of verbal fluency: role of the left inferior frontal gyrus. *Hum. Brain Mapp.* 27 (10), 799–810.
- Wagner, S., Sebastian, A., Lieb, K., Tüscher, O., Tadić, A., 2014. A coordinate-based ALE functional MRI meta-analysis of brain activation during verbal fluency tasks in healthy control subjects. *BMC Neurosci.* 15 (1), 19.
- Tariot, P.N., 1986. A psychobiologic analysis of cognitive failures: structure and mechanisms. *Arch. Gen. Psychiatry* 43 (12), 1183. <https://doi.org/10.1001/archpsyc.1986.01800120069014>.
- Kuehner, C., 2017. Why is depression more common among women than among men? *Lancet Psychiatry* 4 (2), 146–158.
- Sowell, E.R., Peterson, B.S., Kan, E., Woods, R.P., Yoshii, J., Bansal, R., Xu, D., Zhu, H., Thompson, P.M., Toga, A.W., 2007. Sex differences in cortical thickness mapped in 176 healthy individuals between 7 and 87 years of age. *Cereb. Cortex* 17 (7), 1550–1560.
- van Hemmen, J., Saris, I.M.J., Cohen-Kettenis, P.T., Veltman, D.J., Pouwels, P.J.W., Bakker, J., 2016. Sex differences in white matter microstructure in the human brain predominantly reflect differences in sex hormone exposure. *Cereb. Cortex* bhw156.
- Scheuringer, A., Wittig, R., Pletzer, B., 2017. Sex differences in verbal fluency: the role of strategies and instructions. *Cogn. Process.* 18 (4), 407–417.
- Killian, G.A., Holzman, P.S., Davis, J.M., Gibbons, R., 1984. Effects of psychotropic medication on selected cognitive and perceptual measures. *J. Abnorm. Psychol.* 93 (1), 58–70.
- Orzechowska, A., Filip, M., Gatecki, P., 2015. Influence of pharmacotherapy on cognitive functions in depression: a review of the literature. *Med. Sci. Monit.* 24 (21), 3643–3651.
- Chouiter, L., Holmberg, J., Manuel, A.L., Colombo, F., Clarke, S., Annoni, J.-M., Spierer, L., 2016. Partly segregated cortico-subcortical pathways support phonologic and semantic verbal fluency: a lesion study. *Neuroscience* 329, 275–283.
- De Santis, S., Drakesmith, M., Bells, S., Assaf, Y., Jones, D.K., 2014. Why diffusion tensor MRI does well only some of the time: variance and covariance of white matter tissue microstructure attributes in the living human brain. *Neuroimage* 1 (89), 35–44.