CORTEX 94 (2017) 182–192

Available online at www.sciencedirect.com

ScienceDirect

Journal homepage: www.elsevier.com/locate/cortex

Research report

Mechanisms of hemispheric lateralization: A replication study

Kim C. Wende [a,**], Catherine Thiel [a], Jens Sommer [b], Frieder M. Paulus [c], Sören Krach [c] and Andreas Jansen [a,b,*]

[a] Laboratory for Multimodal Neuroimaging (LMN), Department of Psychiatry and Psychotherapy, University of Marburg, Germany

[b] Core-Unit Brainimaging, Faculty of Medicine, University of Marburg, Germany

[c] Social Neuroscience Lab, Department of Psychiatry and Psychotherapy, University of Lübeck, Germany

ARTICLE INFO

Article history:
Received 13 September 2016
Reviewed 2 November 2016
Revised 4 January 2017
Accepted 13 April 2017
Action editor Pia Rotshtein
Published online 22 April 2017

Keywords:
Hemispheric lateralization
Laterality
Reliability
fMRI
ICC
Connectivity

ABSTRACT

It has been shown, using functional magnetic resonance imaging (fMRI), that hemispheric lateralization of brain activity depends on the requirements of the cognitive task performed during the processing of a sensory stimulus rather than on the intrinsic characteristics of that stimulus [Stephan et al., 2003, Science 301 (5631): 384–6]. Task-dependent increase in the coupling of the anterior cingulate cortex (ACC), a region involved in cognitive control, and brain areas in the left prefrontal and right parietal cortex, respectively, regions involved in task execution, was proposed as the mechanism underlying this task-dependency of hemispheric lateralization. The aim of the present study was two-fold: First, we aimed for a conceptual replication of these findings in an independent sample of subjects. Second, we investigated the test–retest reliability of the imaging paradigm to assess whether the task can be used to capture reliable measures of inter-individual differences in hemispheric lateralization. We were able to confirm previous findings showing that hemispheric lateralization depends on the nature of the cognitive task rather than on the nature of the processed stimuli. The task-related brain activation patterns were highly reliable across sessions (as indicated by intra-class correlation coefficients – ICCs, $\geq$.51). We could, however, not replicate previous results proposing task-dependent changes in the coupling between ACC and brain regions for task execution as the mechanism underlying hemispheric lateralization. This re-opens the question which mechanisms could determine the task-dependent functional asymmetries that were observed previously and replicated in this study.

* *Corresponding author*. Department of Psychiatry, University of Marburg, Rudolf-Bultmann-Str. 8, 35039 Marburg, Germany.
** *Corresponding author*. Department of Psychiatry, University of Marburg, Rudolf-Bultmann-Str. 8, 35039 Marburg, Germany.
E-mail addresses: wendek@staff.uni-marburg.de (K.C. Wende), jansena2@staff.uni-marburg.de (A. Jansen).

http://dx.doi.org/10.1016/j.cortex.2017.04.013
0010-9452/

1. Introduction

Functional asymmetries between the hemispheres have been known since the middle of the 19th century. The underlying mechanisms however are still not completely understood. Imaging techniques such as functional magnetic resonance imaging (fMRI) can help to develop and test theories on these mechanisms. Newer approaches emphasize on the one hand that hemispheric lateralization should not only be operationalized by hemispheric asymmetries in local structure or function, but also in terms of asymmetries of intra- and interhemispheric brain connectivity (Frässle, Paulus, Krach, & Jansen, 2016; Frässle, Paulus, Krach, Schweinberger, et al., 2016; Seghier, Josse, Leff, & Price, 2011; Stephan, Penny, Marshall, Fink, Friston, 2005; Stephan, Fink, Marshall, 2007; Stephan, Marshall, Penny, Friston, Fink, 2007). On the other hand, comprehensive models of hemispheric lateralization must also consider the interaction of several lateralized brain functions such as language, spatial attention or memory, since this interaction can lead to tremendous heterogeneity across participants with regard to hemispheric lateralization (Axmacher et al., 2009; Jansen et al., 2006; Jansen, Flöel, Menke, Kanowski, & Knecht, 2005; Weber, Fliessbach, Lange, Kugler, & Elger, 2007; Willems, der Haegen, Fisher, Francks, 2014; Willems, Peelen, & Hagoort, 2010). We therefore need a battery of imaging tasks that (1) are able to assess the lateralization of several cognitive functions, (2) allow the calculation of intra- as well as interhemispheric connectivity measures, (3) yield reliable results at the individual subject level.

In the present study, we explored the utility of an imaging task previously described by Stephan et al. (2003). In this paradigm, subjects processed stimuli containing both verbal and spatial information; critically, task instructions were varied (subjects had to perform either a verbal or a spatial task) while stimuli were identical across the tasks. In a first analysis that directly contrasted the two tasks, the authors showed that processing of verbal information led to left-lateralized brain activity, processing of spatial information to right-hemispheric activity. Since the same stimuli were used throughout the experiment, the authors could show that hemispheric lateralization depends on a top-down regulated mechanism, i.e., that it is task-dependent rather than stimulus-dependent. In a second analysis, Stephan et al. further investigated potential mechanisms that were responsible for the differences in the lateralization of brain activation. They showed that the anterior cingulate cortex (ACC) selectively increased its coupling with either left prefrontal brain regions or right parietal brain areas depending on the task (for a more comprehensive summary of the study, see Supplementary material S1).

The purpose of the present study was two-fold: The first aim was to replicate the results of Stephan and colleagues in an independent sample of subjects ["constructive type replication", (Lykken 1968)]. Replication of previous results is an important, but often neglected issue in neuroscientific studies. As stated by Bennett and Miller (2010) in a recent review on the reliability of fMRI results, "if results do not generalize from one set of subjects to another (…), then the findings are of little value scientifically". The second aim was to assess the test–retest reliability of the paradigm, as a crucial prerequisite to decide whether this paradigm can be used to determine stable brain imaging markers characterizing hemispheric lateralization differences (both with regard to brain activation and brain connectivity) of individual subjects ["internal replication", (Lykken 1968)].

2. Methods

2.1. Subjects

Twenty male subjects participated in the study (mean age 25.1 ± 3.9 years, range 20–33 years). Inclusion criteria were right-handedness according to the Edinburgh inventory (Oldfield, 1971), German as native language, and normal or corrected-to-normal vision. Exclusion criteria were alcohol or drug abuse, past or present psychiatric or neurological disorders according to ICD-10, reading or spelling disorders and disturbances of color vision. All subjects gave written informed consent before participation in the study. The study was approved by the local ethics committee of the medical faculty of the University of Marburg. To investigate test–retest reliability, subjects were scanned during two sessions separated by 18.1 ± 6.7 days (range 7–42 days). Four participants were not available for a second measurement and thus were excluded from analyses of test–retest reliability.

2.2. Experimental design

The experimental design was based on the 2 × 2 × 2 factorial paradigm previously described by Stephan et al. (2003). The central idea of the paradigm was to apply two tasks requiring predominantly language processes (letter decision task – LT) and visuospatial processes (visuospatial decision task – ST), respectively, to stimuli containing both language and spatial features (factor "task"). In addition, stimuli were either presented in the left or right visual field (factor "hemifield") and subjects gave their responses with either the left or right hand (factor "response hand").

All stimuli were concrete, high-frequency German nouns, each four letters in length. They were originally drawn from a linguistic database (http://wortschatz.uni-leipzig.de) and matched for frequency between words with and without the target letter A, respectively. Words were displayed in upper case, using a nonproportional font without serifs (LettrGoth12BT-Bold from Corel Draw V9). The stimuli had a width of 4.69° and a height of 1.25°. The medial border of the stimulus was 3.52° lateral to the fixation cross. [Of note, the presentation differed from the original study of Stephan and colleagues (width 10°, height 2.3°, medial border 6°). The reason for that is that the authors of the present study used the original stimuli, but did not adjust stimulus size in relation to the geometric properties of the visual projection. Potential implications are discussed in the discussion section.] In each stimulus either the second or third letter was red whereas all

other letters were black. In the LT, the subjects had to ignore the position of the red letter and had to indicate whether or not the displayed word contained the target letter "A". In the ST, subjects had to ignore the language related properties of the words and had to indicate whether the red letter was located left or right of the center of the word. In an additional baseline task (BT), subjects were asked to respond as quickly as possible to the onset of each stimulus.

During the task periods a fixation cross was presented in the middle of the screen. Subjects were instructed to maintain fixation of the cross throughout the experiment, except for the instruction periods. Fixation was controlled by monitoring eye movements for all subjects. The stimuli were presented for 150 msec at the same vertical angle as the fixation cross but either in the left or right of the visual hemifield. The short display time was chosen to ensure that subjects could not saccade to the stimuli during presentation. The stimulus onset asynchrony (SOA) was 2000 ± 500 msec (randomly jittered). All responses were given by button presses on a response pad using the index and the middle finger of either the left or right hand.

Data acquisition was performed in a single run, but subdivided into four parts which were separated by short breaks of 1 min. Each part consisted of eight task blocks of either letter or visuospatial decisions that alternated with eight baseline blocks. In each block there were equal numbers of words with and without the target letter "A" and words with red letters at the 2nd and 3rd position. Each block consisted of 12 trials (together 24s) and was preceded by instructions (6s). The instructions informed the subjects which task they were going to perform in the following block and which response hand had to be used. For each combination of the three experimental factors (letter *vs* visuospatial decisions, left *vs* right visual field, and left *vs* right response hand) there were equal numbers of blocks and each of the four parts contained all eight different conditions. The condition order was pseudorandomized and counterbalanced across subjects. For the second measurement we changed the order of presented words, response hand, and hemifield, but used the same order of the LT or ST. Subjects who started for example with the LT in the first measurement also started with the LT in the second run. All subjects were given standardized written instructions before the experiment. To avoid training effects all participants were trained before the experiment with a different set of stimuli.

2.3. Data acquisition

Magnetic resonance imaging (MRI) data was acquired on a 3T Tim Trio MR scanner (Siemens Medical Systems, Erlangen) at the University of Marburg. Functional images were collected with a T2*-weighted echo planar imaging (EPI) sequence sensitive to BOLD contrast (64 × 64 matrix, FOV 230 mm, in plane resolution 3.6 mm, 42 slices, slice thickness 3.6 mm, TR = 3.0 sec, TE = 30 msec, flip angle 90°). Slices covered the whole brain and were positioned transaxially parallel to the anterior–posterior commissural line (AC–PC). In total, 708 functional images were collected. The initial five images were excluded from further analyses to remove the influence of T1 stabilization effects.

2.4. Data analysis

Behavioral data (i.e., error rate and reaction time) were analyzed separately for each run using a 2 × 2 × 2 analysis of variance (ANOVA) design with the factors task (LT, ST), hemifield field (left, right) and response hand (left, right) as independent variables.

Imaging data were analyzed with SPM8 (www.fil.ion.ucl.ac.uk/spm) using standard routines and templates, analogous to the procedure described by Stephan and colleagues. After slice-timing, functional images were realigned, normalized (resulting voxel size 2 × 2 × 2 mm^3), smoothed (applying a 8 mm full-width-at-half-maximum – FWHM, isotropic Gaussian filter), and high-pass filtered (cut-off period 128 sec). Subsequently, we performed three statistical analyses, trying to replicate the findings of Stephan and colleagues (analysis 1 and 2) and assessing the test–retest reliability of the findings (analysis 3). Activated brain regions were described using the MNI coordinate system defined by the standard SPM EPI template (Brett, Johnsrude, & Owen, 2002).

Additionally, after obtaining the results, we repeated all statistical analyses (i.e., analysis 1 and 2 described below) using the software version SPM99 that was used in the original study, in order to ensure that the failure to replicate some effects was not due to the newer software version. A detailed description of the SPM99 results can be found in the Supplementary material S3.

2.4.1. Analysis 1: Task dependence of hemispheric lateralization

The preprocessed images were statistically analyzed in a two-level, mixed-effects procedure. At the subject level, a fixed-effect general linear model (GLM) included all three experimental factors (task, response hand, hemifield), the BT, the instruction periods, and six regressors modeling head movement parameters. Parameter estimate (β–) and contrast images were calculated for each subject contrasting LT > BT, ST > BT, and LT > ST. At the group level, the contrast images were entered separately for each run into one sample *t*-tests. According to the results reported by Stephan and colleagues, for the contrast LT > ST we expected significant activation differences in left-hemispheric brain regions, in particular in the lateral prefrontal cortex ("Broca's area"). For the contrast ST > LT, we expected activation differences in right-hemispheric parietal regions. Therefore, we conducted region-of-interest (ROI) analyses based on the Wake Forest University PickAtlas software (www.fmri.wfubmc.edu) using masks of the left and right frontal cortex (LT > ST) and the left and right parietal cortex (ST > LT), respectively.

2.4.2. Analysis 2: Mechanisms of cognitive control

Task-dependent contributions of the left and right ACC to brain activation were assessed by a psychophysiological interaction (PPI) analysis. Two PPI analyses were calculated separately for the left and the right ACC. In both analyses, the preprocessed images were statistically analyzed in a two-level, mixed-effects procedure. At the subject level, a fixed-effect GLM included either the left or right ACC time series, a corresponding PPI regressor, a regressor modeling the main

effect of task, the BT, the instruction periods, and six regressors modeling head movement parameters.

The ACC time series was extracted as follows: first, analysis 1 (see above) was repeated with preprocessed images that were smoothed with 4 mm FWHM. This was done to minimize potential midline overlap of ACC signal across hemispheres. Second, we determined for both sessions the local maxima of left and right ACC activation for the contrast "(LT + ST) > 2BT" on the group level. Third, for each subject and each session, we aimed to identify the local ACC maxima that was closest to the group activation maximum for the "(LT + ST) > 2BT" contrast at $p < .05$ (uncorrected) located within a mask of the left and right ACC, respectively. Fourth, seed time series were extracted as the first eigenvariate in a sphere around the individual local maximum voxel with a 4 mm radius as implemented in SPM8. The time series were mean-corrected and adjusted for movement variance by applying an effects-of-interest correction using the appropriate F-contrast.

The PPI regressors were created using the PPI toolbox of SPM8. First, the ACC time series was deconvolved. Second, it was multiplied by a vector coding for the main effect of task (1 for letter decisions, −1 for visuospatial decisions, 0 elsewhere). Third, the resulting vector was reconvolved with the hemodynamic response function employed by SPM8 and used as regressors of interest in two new first level analysis for the left and right ACC, respectively. Parameter estimate (β−) and *t*-statistic images were calculated for each subject using the preprocessed images that were smoothed with 8 mm FWHM. Brain regions that were stronger correlated with the ACC time series during LT than during ST were determined by applying a *t*-contrast on the PPI regressor, 0 elsewhere. Accordingly, areas that were stronger correlated with the ACC time series during visuospatial than during letter decision were determined by applying a t-contrast that was −1 for the PPI regressor, 0 elsewhere. At the group level, the contrast images were then tested separately for the left and for the right ACC time series and separately for each run using one-sample *t*-tests. According to the results reported by Stephan and colleagues, we expected (i) stronger coupling of the left ACC with regions in the left prefrontal cortex during letter decisions compared to visuospatial decisions, and (ii) stronger coupling of the right ACC with areas in the right parietal cortex during visuospatial compared to letter decisions. Again, we conducted ROI analyses using the same masks as described above.

2.4.3. *Analysis 3: Test–retest reliability*

As a measure of test–retest reliability, we used the intra-class correlation coefficient (ICC). The ICC describes the stability of inter-individual differences in the brain activation over time. Mathematically, this coefficient sets within-subject variance (σ^2_{within}) in relation to between-subject variance ($\sigma^2_{between}$). We used the ICC(3,1)-type (Shrout & Fleiss, 1979) computed as

$$ICC = (\sigma^2_{between} - \sigma^2_{within})/(\sigma^2_{between} + \sigma^2_{within})$$

Variance components were calculated by the individual contrast images for the following contrasts: LT > BT, ST > BT, LT > ST, ST > LT.

ICC values range from −1 to 1. According to established standards, reliability is often classified as "poor" for ICC ≤ .40, as "fair" for .40 < ICC ≤ .6, as "good" for .60 < ICC ≤ .80, and as "excellent" for ICC > .80 [see (Brandt et al., 2013; Caceres, Hall, Zelaya, Williams, & Mehta, 2009; Fliessbach et al., 2010) for studies on the application of ICCs on fMRI data].

ICCs can be calculated for each voxel ("Voxel-ICCs") and for predefined ROIs ("ROI-ICCs") [see (Brandt et al., 2013) for a discussion]. In a first approach (applied to the contrasts LT > BT, ST > BT, LT > ST, ST > LT), we calculated ICCs for each voxel using a toolbox provided by Caceres et al. (2009). The calculation of ICCs for each voxel allows testing for test–retest reliabilities in the whole brain and further enables to relate the reliability of brain activity (expressed by the ICC) to the strength of brain activity (expressed by the *t*-value). Also the analyses of the voxel-wise ICCs can be constrained to those voxels only, which have absolute *t*-values above a certain threshold, supporting the notion that voxels which have a greater signal-to-noise ratio also achieve greater test–retest reliability ("activation network"). In a second approach (applied to the contrasts LT > ST, ST > LT), we also calculated ICCs by first averaging activation values from the individual contrast images within specific ROIs. Although this procedure is less flexible than the calculation of ICC maps, it is supposed to decrease random noise due to the averaging of activation values across voxels. Analogous to the approach described by Caceres et al. (2009), we calculated an average contrast value either by the mean value of all voxels in a ROI or by the median value of all voxels. The ROIs were defined functionally by determining the overlap of activity of session 1 and 2 either in the left prefrontal cortex (LT > ST; $p = .05$ uncorrected, cluster size 30) or the right parietal cortex (ST > LT; $p = .01$ uncorrected, cluster size 100).

3. Results

3.1. *Behavioral data*

Error rate: All subjects were able to perform all tasks with only few errors. In session 1, the average error rate was 6.9% ± 8.9% for the LT, 8.5% ± 9.2% for the ST, and 3.8% ± 16.9% for the BT. In session 2, the average error rate was 8.6% ± 3.6% for the LT, 10.1% ± 6.5% for the ST, and 3.8% ± 15.8% for the BT. For the error rate, we did not find main effects or interactions that were significant in both sessions.

Reaction time: For the reaction time, we found a significant main effect of task in both sessions (session 1: $F = 22.94$, $p < .001$; session 2: $F = 54.51$, $p < .001$) as well as an interaction of task and visual field (session 1: $F = 16.83$, $p = .001$; session 2: $F = 7.60$, $p = .015$). Subjects needed more time for the LT (session 1: 521.7 ± 23.3 msec; session 2: 546.9 ± 17 msec) than for the ST (session 1: 438.6 ± 13.3 msec; session 2: 466.5 ± 14.5 msec). During the LT, subjects' responses were faster when the stimulus was presented in the right hemifield (RHF) in comparison to the left hemifield (LHF, session 1: LHF: 539.8 ± 17.5 msec, RHF: 503.5 ± 1.7 msec; session 2: LHF: 558.7 ± 1.5 msec, RHF: 533.3 ± 14.6 msec). During the ST, we found the opposite effect. Subjects' responses were faster when the stimulus was presented in the LHF in comparison to RHF (session 1: LHF: 434.2 ± 7.6 msec, RHF: 443.0 ± 19.8 msec; session 2: LHF: 457.1 ± 7.1 msec, RHF: 475.9 ± 15.1 msec).

3.2. Functional imaging data

(a) Replication of previous results.

Analysis 1 (Task dependence of hemispheric lateralization): For the baseline contrasts (LT > BT, ST > BT), a bilateral network was activated in both sessions, encompassing the lateral prefrontal cortex, the bilateral dorsocaudal ACC, parietal regions and the cerebellum (see Supplementary results S2). For the difference contrasts (LT > ST, ST > LT), we replicated the results of Stephan and colleagues, however with considerably lower effect sizes (as expressed by lower *t*-values). In the following, we will give a summary of our main findings.[1] We also uploaded the *t*-maps describing our results to NeuroVault (http://www.neurovault.org/collections/1516/) so that the reader is able to visualize and aggregate the results across studies more easily.

Contrast ST > LT. In a first step, we analyzed brain activity for the contrast ST > LT in the parietal ROI (in which Stephan and colleagues reported activity differences between both conditions). For the first session, we found at $p = .001$ (uncorrected) one activated cluster ("cluster 1", MNI coordinates 18, −70, 58; cluster size $k = 219$, $t_{max} = 5.02$). At $p = .01$ (uncorrected), we found an additional cluster located more laterally in the right parietal ROI ("cluster 2", MNI coordinates 38, −40, 36; $k = 669$, $t_{max} = 3.67$). Both clusters were significant at $p < .05$, corrected for multiple comparisons at the cluster level. In the second session, we did not find brain activity at $p = .001$ (uncorrected, cluster size threshold 100 voxels). At $p = .01$ (uncorrected), we found two clusters located in the right precuneus (MNI coordinates 6, −48, 44, $k = 204$, $t_{max} = 4.90$, $p < .05$ corrected for multiple comparisons at the cluster level) and in the right superior parietal cortex (MNI coordinates 30, −46, 48, $k = 103$, $t_{max} = 5.45$, $p = .548$ corrected for multiple comparisons at the cluster level). In both sessions, the same clusters were activated, but the activation was more extended in the first session. The first cluster of the second session corresponds to "cluster 1" of the first session, the second cluster to "cluster 2" (Fig. 1). In a second step, we also analyzed brain activity outside the parietal cortex. We found in both sessions an activated cluster in the right parietal-occipital cortex (partially outside of the parietal ROI, "cluster 3", see Fig. 1). In the first session, the maximum activation of the cluster is located at MNI coordinates 44, −76, 20 (281 voxel at $p = .001$, $t_{max} = 6.31$, $p < .05$ corrected for multiple comparisons at the cluster level). In the second session, we found the same cluster, located at MNI coordinates 38, −74, 30 (252 voxels at $p = .01$, $t_{max} = 5.85$, $p < .05$ corrected for multiple comparisons at the cluster level).

Comparison to Stephan et al. (2003): Stephan and colleagues reported brain activity in two clusters in the anterior (MNI coordinates 58, −24, 44) and posterior part of the right inferior parietal lobule (MNI coordinates 46, −76, 34). The first cluster corresponds to cluster 2 of the present study, the second cluster to cluster 3. Cluster 1 of the present study has no clear analog in the Stephan et al. study.

Contrast LT > ST. In a first step, we analyzed brain activity for the contrast LT > ST in the lateral prefrontal ROI (in which Stephan and colleagues reported main activity differences between both conditions). Brain activity is illustrated for both sessions in Fig. 2. In the first session, we only found activation at $p = .05$, uncorrected ("cluster 1", MNI coordinates −42, 24, 18, $k = 233$, $t_{max} = 2.88$). In the second session, the contrast resulted in two activated clusters at $p = .01$, uncorrected ("cluster 2": MNI coordinates −52, 12, 8, $k = 118$, $t_{max} = 5.78$; "cluster 3": MNI coordinates −44, 26, 2, $k = 92$, $t_{max} = 4.38$). The first cluster was significant at $p < .05$, corrected for multiple comparisons on the cluster level. The cluster of the first session is located somewhere between the clusters of the second session.

In a second step, we also analyzed brain activity outside the prefrontal cortex. Surprisingly, whole-brain activation patterns of both sessions clearly differed. While the activation pattern of the second session was similar to the pattern described by Stephan and colleagues (e.g., with additional activations in the left ACC and in the left cerebellum, extending into the fusiform gyrus – FG), in the first session the most strongly activated areas (at $p = .01$) were located in the cerebellar vermis (MNI coordinates −2, −60, −36, $t_{max} = 3.63$, $k = 177$) and bilaterally in the thalamus and basal ganglia [MNI coordinates: (i) −18, −6, 6, $t_{max} = 3.67$, $k = 134$; (ii) 32, 20, −14, $t_{max} = 3.75$, $k = 93$].

Comparison to Stephan et al. (2003): Stephan and colleagues reported brain activity in two clusters in the prefrontal cortex (MNI coordinates −42, 32, 0 and −48, 12, 22). The first activation maximum corresponds to cluster 3 of the present study, the second cluster to cluster 2 (Fig. 2).

Analysis 2 (Mechanisms of cognitive control). As seed region for the PPI analysis, we chose the left and right ACC, respectively. The group coordinates of the ACC activation in the present study were comparable to the coordinates reported by Stephan and colleagues [Stephan: left ACC −6, 8, 50, right ACC 8, 14, 48; present study: left ACC −4, 24, 54 (session 1), −4/20/46 (session 2); right ACC 12, 20, 48 (session 1), 6/16/46 (session 2)]. However, in none of the sessions we were able to replicate the findings of Stephan and colleagues for the PPI analysis. We did not find significantly increased correlations of the ACC time series with the left prefrontal cortex during LT or the right parietal cortex during ST. A possibility to interpret the null findings is the application of Bayes factors (BF) (Dienes, 2014). BF are most informative in cases of failure to replicate previous findings when null results are observed, but the literature reports positive results. We therefore computed BF for the PPI analyses (see Supplementary material S4). BF ranged between 3 and 1/3, indicating data insensitivity in distinguishing H0 (no connectivity changes) and H1 (connectivity changes).

Imaging data was analyzed with SPM8, the newest version that was available when the data was analyzed for the first time. As pointed out by one of the reviewers, the original study however used another software version (SPM99), which differed in some key aspects of fMRI analysis (e.g., the treatment of temporal autocorrelation and in particular the implementation of the PPI analysis; see discussion). To ensure that differences in software versions were not responsible for

[1] We present results at different statistical thresholds set at the voxel level ($p = .001$, $p = .01$, $p = .05$; uncorrected). We additionally denote whether the presented clusters were significant at the cluster level ($p < .05$, corrected for multiple comparisons).

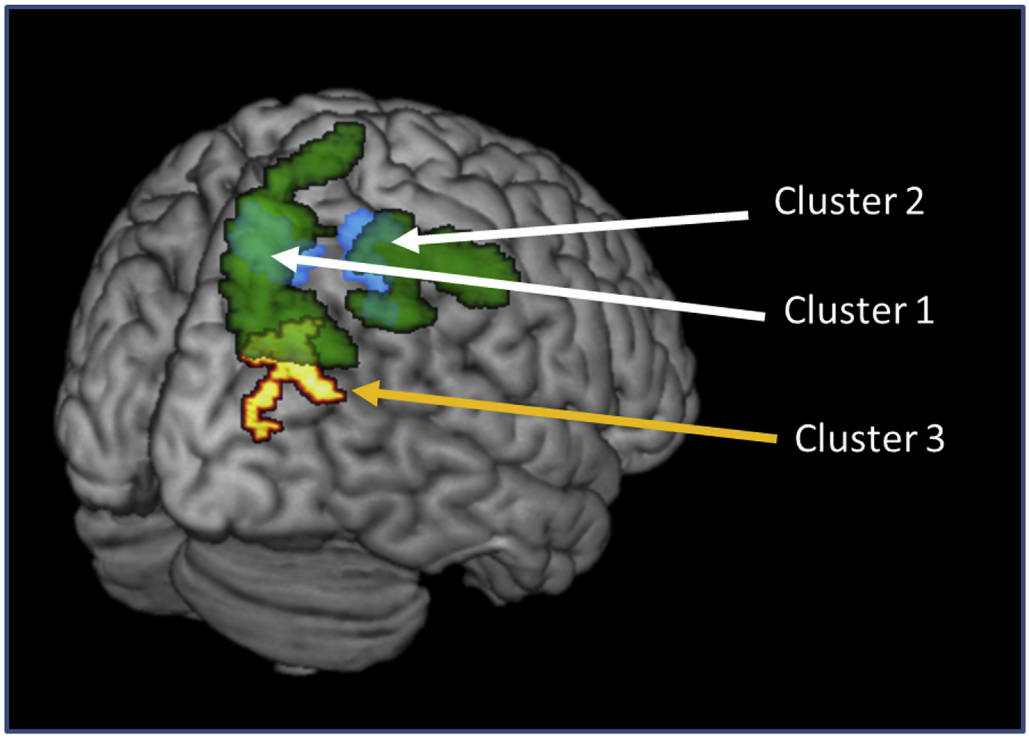

Fig. 1 – Spatial task > language task (ST > LT): Left: Group activation pattern (parietal ROI) for the contrast ST > LT for session 1 (green) and session 2 (blue). For illustrational purposes, we used a statistical threshold of $p = .01$, uncorrected, and a cluster size threshold of 100 voxels. In both sessions the same regions were activated (cluster 1 and 2, see text), although the activation was more extended in the first session. We additionally found activity in a cluster partially outside the parietal cortex (cluster 3, depicted in orange as overlay of session 1 and 2) replicating brain activity reported in the Stephan et al. study (see text).

the divergent findings between both studies, in particular with regard to the non-replication of the PPI results, we re-analyzed the data with SPM99. The results however did not change (see Supplementary material s3).

(b) Test–retest reliability.

Baseline contrasts. Test–retest reliability of the baseline contrasts was calculated for each voxel. In Fig. 3 (top), we present whole-brain joint probability distributions of *t*-values and ICCs. These distributions showed, as expected, an association between *t*-values and ICCs. ICCs were generally higher in brain areas showing strong effects in absolute terms for the contrast LT versus BT and ST versus BT. In Fig. 3 (bottom), we present the ICC distribution as a histogram for the whole brain (red) as well as for the activated and deactivated network (blue/green). The networks were defined based on the t-maps either from the first or second session. Voxels were classified as (de)activated if they had *t*-values $|t| > 3.73$ (corresponding to $p < .001$ uncorrected). For the contrast LT > BT, the median ICC for the activated network was .65 and .59, respectively, depending on whether the active network was defined by the first or the second measurement. For the deactivated network, the median ICC was .52 and .49, respectively. For the contrast ST > BT, the median ICC of the activated network was .51 and .52, respectively. For the deactivated network, the median ICC was .39 and .40, respectively.

Difference contrasts. In a first step, test–retest reliability of the difference contrasts was analyzed voxel-by-voxel ("Voxel-ICCs"). For the contrast ST > LT, the median ICC for the activated network (at $p = .05$) was .16 and .24, respectively, depending on whether the active network was defined by the first or the second measurement. For the contrast LT > ST, the median ICC for the activated network (at $p = .05$) was .15 and .18, respectively, depending on whether the active network was defined by the first or the second measurement. Overall, the reliability values for the difference contrasts were considerably smaller than for the baseline contrasts.

In a second step, test–retest reliability was analyzed for specific ROIs ("ROI-ICCs"). Activation values were first averaged within a ROI, then the ICC was calculated. As ROIs we chose those regions that were most strongly activated on the group level. The ROIs were defined functionally by determining the overlap of activity of session 1 and 2 either in the left prefrontal cortex (LT > ST; $p = .05$ uncorrected, cluster size 30) or the right parietal cortex (ST > LT; $p = .05$ uncorrected, cluster size 30). For the contrast ST > LT, we defined three ROIs: the two clusters in the right parietal cortex (ROI "medial parietal", corresponding to cluster 1, and ROI "lateral parietal", corresponding to clusters 2) and the adjacent cluster in the right parietal-occipital cortex (ROI "parieto-occipital", corresponding to cluster 3). For the contrast LT > ST, we chose one ROI in the left lateral prefrontal cortex (ROI "prefrontal")

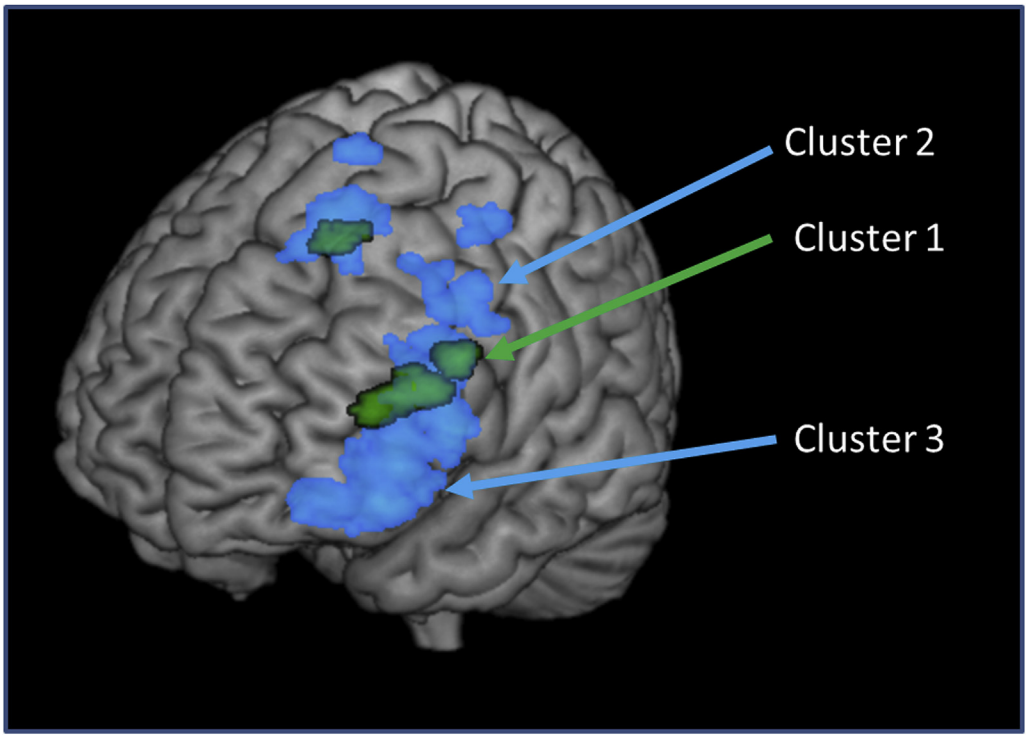

Fig. 2 – Language task > spatial task (LT > ST): Group activation pattern (left prefrontal ROI) for the contrast LT > ST for session 1 (green) and session 2 (blue). For illustrational purposes, we used a statistical threshold of $p = .05$, uncorrected, and a cluster size threshold of 30 voxels. Cluster 1 belongs to session 1, cluster 2 and 3 to session 2 (see text).

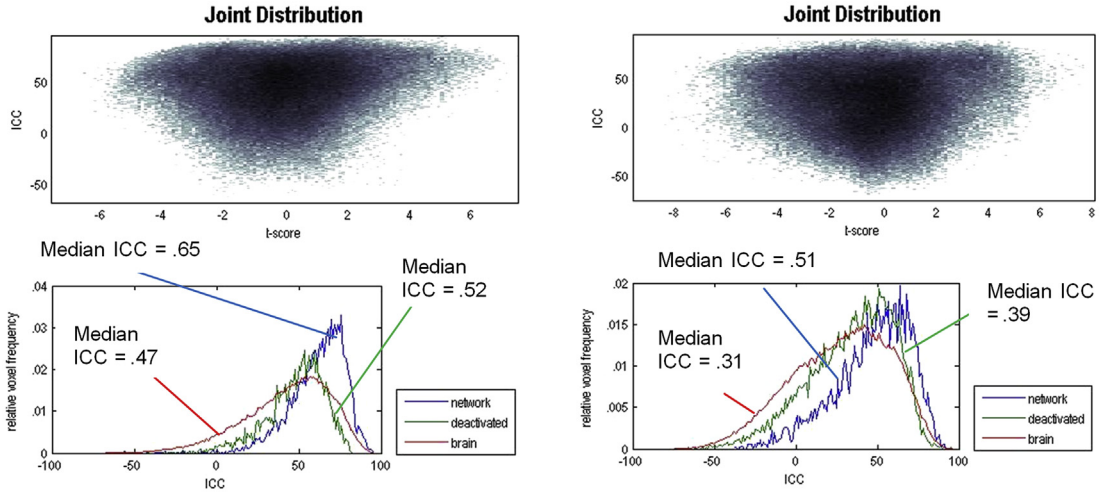

Fig. 3 – Test–retest results: Top: Joint probability distribution of voxel-wise t-values and associated ICC values for the contrast LT > BT (left) and ST > BT (right) (session 1). Bottom: ICC frequency distributions for the whole-brain (red), and for the voxels in the activated network (blue) and "deactivated" network (green). The "activated" and "deactivated" networks were defined based on the results from the first measurement. Voxels were classified as (de)activated if they had t-values $|t| > 3.73$ (corresponding to $p < .001$). Frequency distributions are presented for the contrasts LT > BT (left) and ST > BT (right).

where we found maximal overlap between both sessions. All ICCs calculated on a ROI basis are presented in Table 1.

PPI-results: Despite the non-significant results, we determined the voxel-wise test–retest reliability for the PPI analysis. For the left ACC PPI (LT > ST), the median ICC for the activated network (at $p = .05$ uncorrected) was .00 and −.02, respectively, depending on whether the active network was defined by the first or the second measurement. For the right ACC PPI (ST > LT), the median ICC for the activated network (at $p = .05$ uncorrected) was .00 and .07, respectively, depending on whether the active network was defined by the first or the second measurement. Overall, the reliability values for the PPI

Table 1 – ROI-based ICCs for the contrasts LT > ST and ST > LT.

Contrast	ROI	ICC-ROI (mean)	ICC-ROI (median)
LT > ST	ROI "prefrontal"	.44	.48
ST > LT	ROI "medial parietal"	.59	.59
	ROI "lateral parietal"	.54	.56
	ROI "parieto-occipital"	.46	.56

results were even considerably smaller than for the differential contrasts and partly negative, indicating that the within-group variance was exceeding between-group variance.

4. Discussion

The purpose of the current study was two-fold. First, we aimed to replicate the results of a previous study that described basic mechanisms underlying hemispheric lateralization using both activation and connectivity metrics (Stephan et al., 2003) in an independent sample of subjects ["constructive type replication", (Lykken 1968)]. Second, we assessed the test–retest reliability of the paradigm ["internal replication", (Lykken 1968)]. With regard to the first aim, we were able to confirm previous findings showing that hemispheric lateralization depends on the nature of the cognitive task rather than on the nature of the processed stimuli.[2] We could, however, not replicate previous results proposing task-dependent changes in the coupling between ACC and brain regions for task execution as the mechanism underlying hemispheric lateralization. With regard to the second aim, we showed that brain activation patterns were highly reliable across sessions (as indicated by ICCs $\geq$ .51), both for the baseline contrasts (LT > BT, ST > BT) and the difference contrasts (LT *vs* ST). Altogether, the measures of brain activation for the baseline and the difference contrasts suggest a positive evaluation of this task. The missing robustness of the paradigm with regard to aspects of brain connectivity however, together with the long measurement time and the relatively high difficulty level, might be an obstacle when one is interested in using this task in a task battery assessing the lateralization of brain connectivity at the single subject level. In the following, we will discuss both the results from behavioral data and imaging data.

Behavioral data. The analysis of behavioral data consistently yielded a significant interaction of task and visual field for the reaction times. During the letter task, subjects' responses were significantly faster when the stimulus was presented in the RHF in comparison to LHF. During the spatial task, we found the opposite effect. Subjects' responses were faster when the stimulus was presented in the LHF in comparison to RHF. This interaction effect reflects the different functional properties of the hemispheres, that is, left-hemispheric dominance for language and right-hemispheric dominance for spatial attention. If a stimulus is presented in the RHF, it will be first processed in the left (i.e., language dominant) hemisphere. During the letter task, the stimulus can be processed without interhemispheric transfer so that reaction times are relatively short. If the stimulus however is presented in the LHF, it will be first processed in the right hemisphere. For language processing, information must thus be transferred to the left hemisphere and this additional processing contributes to the higher reaction times. During the spatial task this interaction reverses due to the right-hemispheric dominance for spatial attention. Overall, the interaction effect demonstrates the different lateralization properties of the hemispheres already on the behavioral level and is an indicator for the validity of the paradigm.

Imaging data. It might be argued that the results of the present study are of limited relevance to the broader scientific community when one primarily considers their novelty or their implications for future studies. Only a few experts in the field are interested in the fact that a specific imaging paradigm (that has already convincingly yielded important findings about the mechanisms of hemispheric lateralization) might not be a suitable tool for the assessment of brain metrics in single subjects. We nevertheless decided to publish the present results since we believe that both the replication of the results of previous studies and the assessment of the test–retest reliability of imaging paradigms are essential for further developments in neuroscience. The last few years have shown that the interpretation of fMRI results is often hampered by the low reliability of the data (Bennett & Miller, 2010). Initially promising findings have often not been replicated by different labs, with different equipment and in independent samples (Ioannidis, 2005; Pashler & Harris, 2012). While most of the debate on the "replication crisis" has focused on psychology – and social psychology in particular (Open Science Collaboration, 2015) – a similar line of reasoning applies for studies in the neurosciences as well (Button et al., 2013). Here, part of the problems in replicating earlier results is obviously related to the lack of statistical power, which is specifically relevant for the imaging community in the neurosciences (Weinberger & Radulescu, 2016). Regardless of the causes for a specific non-replication to occur, reproducibility of previously reported results however is a defining feature of science in any domain. In the present study, we thus aimed towards a replication of the main findings of Stephan and colleagues and thus used, where applicable, the same parameters as in the original study. We additionally investigated the test–retest reliability of the paradigm, in order to test if it is reliable in capturing individual differences in hemispheric lateralization.

To answer the question to which extent we replicated earlier findings, we first want to define what exactly constitutes a successful replication of fMRI results. In fMRI studies, typically only a small amount of the overall data is reported. Most researchers present statistical values of peak voxels of those clusters that were classified as statistically significantly

[2] Of note, as pointed out by one of the reviewers: Although both the original and the current study found significant differences between tasks in the left (LT > ST) and right hemisphere (ST > LT), respectively, this does – strictly speaking – only show, e.g., that the left prefrontal cortex is more activated than its right-hemispheric homolog for the LT > ST contrast, but not that these differences are significantly different. To show that, one would have to find a significant hemisphere × task interaction. We therefore re-analyzed the data. Results however were not significant.

activated. Many researchers therefore demand that for a successful replication, follow-up studies must also find statistically significant effects in regions previously reported as significant. Following this view, the present results represent a partial replication since results had the same directionality of the findings from the original study and several of the resulting activation clusters were significant at corrected statistical thresholds.[3] Another possibility to interpret the present findings is application of BF. Unlike null-hypothesis significance testing, BF allow to interpret non-significant findings either as evidence for H0 or not much evidence at all (Dienes, 2014). Using the script provided by Zoltan Dienes (www.lifesci.sussex.ac.uk/home/Zoltan_Dienes/inference/Bayes.htm), we calculated BF to estimate whether the current study provided support for existing or non-existing effects in all activated regions for both the differential contrasts (LT > ST, ST > LT) and the PPI analyses. These calculations showed that we had strong evidence (B >> 3) for activity in two clusters (LT > ST, cluster 3, session 2, −44/26/2; ST > LT, cluster 1, session 1, 18/−70/58), while the other BFs, ranging between 3 and 1/3, indicate data insensitivity in distinguishing H0 (no activation/connectivity changes) and H1 (activation/connectivity changes) (see Supplementary material S4 for a detailed description of the BF calculation).

We have also to discuss whether the partial non-replication can be attributed to methodological differences between both studies. In the present study, we used the same stimulus material, applied an analogous data analysis strategy, and included a comparable number of subjects (20 subjects in the present study, 16 subjects in the original study) with the same sample characteristics (~25 years, native German speakers, right-handed, no psychiatric or neurological history, all male). It is also unlikely that simple technical problems can explain the differences between the studies. On the one hand, the baseline contrasts showed highly significant and reliable brain activation patterns. On the other hand, the interaction effect found in the behavioral data shows that the implementation of the task was successful and yielded highly plausible results.

Of note however, the current study differed from the previous study in two points. On the one hand, we used another software version (SPM8 instead of SPM99). On the other hand, the stimuli were presented in a slightly different way. With regard to the analysis software, we decided to use the newest version of SPM that was available at that time, i.e., SPM8, not SPM99 as in the original study. As pointed out by one of the reviewers, both software package versions however differ in some key aspects of fMRI analysis, such as the treatment of temporal autocorrelation or the implementation of the PPI analysis (the inbuilt SPM8 routine for PPI uses for instance a deconvolution procedure which was not readily available in SPM99). These differences might be responsible for the divergent findings of both studies, in particular with regard to the non-replication of the PPI results. We therefore re-analyzed the data also with SPM99 (see Supplementary material S3). The results however remained qualitatively similar.

With regard to the stimulus parameters, the situation is unfortunately more complicated. We used the same stimuli as Stephan and colleagues, but did not adjust stimulus size in relation to the geometric properties of our visual projection system (see Methods). Therefore, the presented stimuli were smaller (width 4.69°, height 1.25°) compared to the Stephan et al. study (width 10°, height 2.3°). Furthermore, the medial border of the stimulus was only 3.52° lateral to the fixation cross (Stephan et al.: medial border 6°).

A sufficiently high visual angle, however, is necessary to safely provide non-foveal vision to ensure that the visual information of each stimulus will be initially received by the contralateral hemisphere only. One has therefore to discuss whether we achieved, as planned, non-foveal stimulation in the present study. On the one hand, there are reasons to believe that the stimulus was initially stronger processed by one hemisphere. First, we found, at the behavioral level, a significant task × hemisphere interaction which would not have been expected if the information entered both hemispheres to the same degree. Second, an analysis of the fMRI data contrasting stimuli presented in the left versus the RHF showed significant activity only in the left and right primary visual cortex, respectively. On the other hand however, the known anatomy of the visual system speaks against a completely non-foveal processing. Foveal vision covers about 5° of visual angle. Presenting a horizontally extended stimulus with a medial border at 3.5°, as in the present experiment, means that the medial part of the stimulus was located in central vision, while the lateral part was in peripheral vision. The stimuli used in the present study therefore most likely activated also the ipsilateral visual cortex (albeit to a lesser degree).[4]

In summary, in the present experiment we presented stimuli most likely not completely non-foveal, but in a mixed central-peripheral vision. This might constitute a substantial difference between the studies and may have led to two somewhat different cognitive processing. Additionally, covert attention becomes considerably more demanding the farther away from fixation the attended object is located. It might therefore be argued, as convincingly pointed out by one of the reviewers, that the difference in visual angle between both studies may have led to significant differences in attentional load. This might not only explain behavioral differences between the two studies, as reflected for instance by reaction time differences of 160 msec between both studies,[5] but could

[3] In a follow-up study, Stephan, Fink et al. (2007), Stephan, Marshall et al. (2007) also investigated mechanism of interhemispheric integration using Dynamic Causal Modeling (DCM). Here, they used a four-region ventral stream model comprising the lingual gyrus (LG) and the fusiform gyrus (FG) in both hemispheres. We also tried to replicate these findings. However, it was not possible to extract from the present data set time series in all four ROIs that were sufficiently correlated with the experimental design (see Supplementary material S5 for a more detailed description).

[4] This conclusion is not in contradiction to the imaging results that show significant activity only in one hemisphere, when comparing both half fields, as pointed out before. At a lower statistical threshold, we also see activity in the ipsilateral visual cortex (albeit less strong).

[5] Stephan et al. reported reaction times of 686 ± 21 msec for the LT and 612 ± 28 msec for the ST, being much larger than in the present study. In addition, they did not find at the behavioral level an interaction effect for hemifield × task.

also have led to differential brain activation. We have to acknowledge that in particular the substantial reaction time differences caution against the assumption that the tasks were truly comparable across both studies."

We believe that for follow-up studies the criterion of finding significant effects in brain regions previously report as significant is not necessarily constructive since the power of fMRI studies is difficult to estimate. Recent debates have highlighted how difficult it is to estimate population effect sizes from the analyses of imaging data (Button et al., 2013). This problem has manifold reasons in part being related to the practice of only publishing maximum effect sizes in stringently thresholded effect size maps (Paulus, Krach, Albrecht, & Jansen, 2013), publication bias in the sense of only publishing positive findings (Ioannidis, 2011) and also the phenomenon that first reports of a specific effect often overestimate the underlying population effect by several magnitudes (Munafò & Flint, 2010). The overestimated population effect sizes together with the required lowering of the alpha error due to corrections for multiple comparison has established a situation in the literature, in which many studies are fundamentally under-powered, making it rather unlikely to find significant results, even if there is a true population effect (Button et al., 2013). We therefore believe that the demand for replication studies to reproduce significant results in the areas of interests is a too strict criterion. Thus we here rather argue for adopting a non-dichotomous, effect size focused idea of replication, where evidence is accumulated across independent observations and can be aggregated in e.g., meta-analyses of effect sizes. In our opinion one study can or cannot ultimately replicate a specific finding so that we rather include all these effects in the support of the initial hypotheses if they show the same directionality, potentially leading to greater posterior evidence. This view suggests that replication is not an issue of a single study to find similar results but an ongoing process of collecting and publishing effects that at some point might lead to overwhelming evidence for a specific hypothesis. To avoid a potential bias in the aggregated evidence, we should by all means also publish those findings that fail to achieve statistical significance or are contradictory to original results (Paulus, et al., 2014; Paulus, Krach, Bedenbender, et al., 2013).

This view is becoming more popular in the imaging community. The development of web-based data repositories such as NeuroVault (Gorgolewski et al., 2015; Gorgolewski et al., 2016) enables researchers to share unthresholded statistical maps describing all the results of fMRI studies, not only those at the peak within a thresholded statistical map. This provides the possibility to share statistical maps with other researchers and enables them to visualize the full data. More importantly, it provides the means to meta-analytically combine effect sizes from different studies. To further accumulate evidence across different studies, one might envision a standardized practice of publishing unthresholded data. With the present study and our sharing of the results, we believe that we are supporting an important endeavor increasing the reliability of previously published data. We were thus able to qualitatively support previous findings showing that hemispheric lateralization depends on the nature of the cognitive task rather than on the nature of the processed stimuli, while leaving open the question if there is a single, itself task-independent mechanism in the brain determining such functional asymmetries across tasks and task contexts.

Acknowledgements

This work was supported by research grants from the Else Kröner-Fresenius-Stiftung (2012_A219) and the CRC/Transregio 135 (Cardinal mechanisms of perception: prediction, valuation, categorization). We thank Klaas Stephan for providing the stimulus material for the present study and especially for helpful feedback on earlier versions of the manuscript. We also thank Olaf Steinsträter to support us with the implementation of the outdated SPM99 software version on our IT system. Last, we would like to thank the two reviewers and the editor for constructive comments, in particular with regard to the application of Bayes factors.

Supplementary data

Supplementary data related to this article can be found at http://dx.doi.org/10.1016/j.cortex.2017.04.013.

REFERENCES

Axmacher, N., Bialleck, K. A., Weber, B., Helmstaedter, C., Elger, C. E., & Fell, J. (2009). Working memory representation in atypical language dominance. *Human Brain Mapping, 30*, 2032–2043.

Bennett, C. M., & Miller, M. B. (2010). How reliable are the results from functional magnetic resonance imaging? *Annals of the New York Academy of Sciences, 1191*, 133–155.

Button, K. S., Ioannidis, J. P. A., Mokrysz, C., Nosek, B. A., Flint, J., Robinson, E. S. J., et al. (2013). Power failure: Why small sample size undermines the reliability of neuroscience. *Nature Reviews. Neuroscience, 14*, 365–376.

Frässle, S., Paulus, F. M., Krach, S., & Jansen, A. (2016). Test-retest reliability of effective connectivity in the face perception network. *Human Brain Mapping, 37*, 730–744.

Frässle, S., Paulus, F. M., Krach, S., Schweinberger, S. R., Stephan, K. E., & Jansen, A. (2016). Mechanisms of hemispheric lateralization: Asymmetric interhemispheric recruitment in the face perception network. *NeuroImage, 124*, 977–988.

Gorgolewski, K. J., Varoquaux, G., Rivera, G., Schwarz, Y., Ghosh, S. S., Maumet, C., et al. (2015). NeuroVault.org: A web-based repository for collecting and sharing unthresholded statistical maps of the human brain. *Frontiers in Neuroinformatics, 9*, 8.

Gorgolewski, K. J., Varoquaux, G., Rivera, G., Schwartz, Y., Sochat, V. V., Ghosh, S. S., et al. (2016). NeuroVault.org: A repository for sharing unthresholded statistical maps, parcellations, and atlases of the human brain. *NeuroImage, 124*, 1242–1244.

Ioannidis, J. P. A. (2005). Why most published research findings are false. *PLoS Medicine, 2*, e124.

Ioannidis, J. P. A. (2011). Excess significance bias in the literature on brain volume abnormalities. *Archives of General Psychiatry, 68*, 773–780.

Jansen, A., Deppe, M., Schwindt, W., Mohammadi, S., Sehlmeyer, C., & Knecht, S. (2006). Interhemispheric dissociation of language regions in a healthy subject. *Archives of Neurology, 63*, 1344–1346.

Jansen, A., Flöel, A., Menke, R., Kanowski, M., & Knecht, S. (2005). Dominance for language and spatial processing: Limited capacity of a single hemisphere. *NeuroReport, 16*, 1017–1021.

Munafò, M. R., & Flint, J. (2010). How reliable are scientific studies? *The British Journal of Psychiatry, 197*, 257–258.

Oldfield, R. C. (1971). The assessment and analysis of handedness: The Edinburgh inventory. *Neuropsychologia, 9*, 97–113.

Open Science Collaboration. (2015). PSYCHOLOGY. Estimating the reproducibility of psychological science. *Science, 349*, aac4716.

Pashler, H., & Harris, C. R. (2012). Is the replicability crisis overblown? Three arguments examined. *Perspectives on Psychological Science, 7*, 531–536.

Paulus, F. M., Bedenbender, J., Krach, S., Pyka, M., Krug, A., Sommer, J., et al. (2014 Apr). Association of rs1006737 in CACNA1C with alterations in prefrontal activation and fronto-hippocampal connectivity. *Human Brain Mapping, 35*(4), 1190–1200.

Paulus, F. M., Krach, S., Albrecht, A.-G., & Jansen, A. (2013a). Potential bias in meta- analyses of effect sizes in imaging genetics. *Schizophrenia Bulletin, 39*, 501–503.

Paulus, F. M., Krach, S., Bedenbender, J., Pyka, M., Sommer, J., Krug, A., et al. (2013b). Partial support for ZNF804A genotype-dependent alterations in prefrontal connectivity. *Human Brain Mapping, 34*, 304–313.

Seghier, M. L., Josse, G., Leff, A. P., & Price, C. J. (2011). Lateralization is predicted by reduced coupling from the left to right prefrontal cortex during semantic decisions on written words. *Cerebral Cortex, 21*, 1519–1531.

Stephan, K. E., Fink, G. R., & Marshall, J. C. (2007). Mechanisms of hemispheric specialization: Insights from analyses of connectivity. *Neuropsychologia, 45*, 209–228.

Stephan, K. E., Marshall, J. C., Friston, K. J., Rowe, J. B., Ritzl, A., Zilles, K., et al. (2003). Lateralized cognitive processes and lateralized task control in the human brain. *Science, 301*, 384–386.

Stephan, K. E., Marshall, J. C., Penny, W. D., Friston, K. J., & Fink, G. R. (2007). Interhemispheric integration of visual processing during task-driven lateralization. *The Journal of Neuroscience, 27*, 3512–3522.

Stephan, K. E., Penny, W. D., Marshall, J. C., Fink, G. R., & Friston, K. J. (2005). Investigating the functional role of callosal connections with dynamic causal models. *Annals of the New York Academy of Sciences, 1064*, 16–36.

Weber, B., Fliessbach, K., Lange, N., Kugler, F., & Elger, C. E. (2007). Material-specific memory processing is related to language dominance. *NeuroImage, 37*, 611–617.

Weinberger, D. R., & Radulescu, E. (2016). Finding the elusive psychiatric "Lesion" with 21st-century neuroanatomy: A note of caution. *The American Journal of Psychiatry, 173*, 27–33.

Willems, R. M., der Haegen, L. Van, Fisher, S. E., & Francks, C. (2014 Mar). On the other hand: Including left-handers in cognitive neuroscience and neurogenetics. *Nature Reviews Neuroscience, 15*(3), 193–201.

Willems, R. M., Peelen, M. V., & Hagoort, P. (2010). Cerebral lateralization of face-selective and body-selective visual areas depends on handedness. *Cerebral Cortex, 20*, 1719–1725.