The assessment of hemispheric lateralization in functional MRI—Robustness and reproducibility


Department of Neurology, University of Münster, Germany
Department of Radiology, Klinikum Bergmannsheil, Bochum, Germany
IZKF Münster, University of Münster, Germany
Department of Epileptology, University of Bonn, Germany
Department of Neuroradiology, University Medical Center Hamburg-Eppendorf, Hamburg, Germany
Department of NeuroCognition, Life and Brain Centre, Bonn, Germany

Received 28 November 2005; revised 8 June 2006; accepted 18 June 2006
Available online 14 August 2006

Various methods have been proposed to calculate a lateralization index (LI) on the basis of functional magnetic resonance imaging (fMRI) data. Most of them are either based on the extent of the activated brain region (i.e., the number of “active” voxels) or the magnitude of the fMRI signal change. The purpose of the present study was to investigate the characteristics of various variants of these approaches and to identify the one that yields the most robust and reproducible results. Robustness was assessed by evaluating the dependence on arbitrary external parameters, reproducibility was assessed by Pearson’s correlation coefficient.

LIs based on active voxels counts at one single fixed statistical threshold as well as LIs based on unthresholded signal intensity changes (i.e., based on all voxels in a region of interest) yielded neither robust nor reproducible laterality results. Instead, the lateralization of a cognitive function was best described by “thresholded” signal intensity changes where the activity measure was based on signal intensity changes in those voxels in a region of interest that exceeded a predefined activation level.

However, not all other approaches should be discarded completely since they have their own specific application fields. First, LIs based on active voxel counts in the form of p-value-dependent lateralization plots (LI = LI(p)) can be used as a straightforward measure to describe hemispheric dominance. Second, LIs based on active voxel counts at variable thresholds (standardized by the total number of active voxels) are a good alternative for big regions of interest since LIs based on signal intensity changes are restricted to small ROIs.

© 2006 Elsevier Inc. All rights reserved.

Keywords: Lateralization; Reproducibility; Reliability; fMRI

Introduction

Cognitive brain functions like language, memory and spatial attention are continuously distributed across the hemispheres, with a preponderance of one hemisphere in most individuals. The relative contribution of the hemispheres during a task is expressed by a lateralization index (LI), describing both hemispheric dominance (i.e., left, right or bilateral) and the degree of lateralization (i.e., how strongly left) (e.g., Hinke et al., 1993; Binder et al., 1995).

The lateralization of a cognitive function is of clinical and theoretical relevance. The degree of language lateralization, for instance, is a marker for the susceptibility to language deficits after unilateral brain lesions (Knecht et al., 2002). The determination of language lateralization is also important for the presurgical evaluation of patients with medically intractable epilepsy, because the risk of postsurgical language and memory deficits is related to the presurgical lateralization of language functions in the brain (Bell et al., 2000; Sabsevitz et al., 2003). Further, the lateralization of cognitive functions has become a subject of interest in psychiatry. Disorders like autism (Hier et al., 1979), attention deficit hyperactivity disorder (Mostofsky et al., 2002) and schizophrenia (Bertolino et al., 2004) have been related to decreased lateralization of brain functions. Abnormal lateralization has been linked to dyslexia (Heim and Keil, 2004). Another field of laterality research is the study of genetic cohorts. By providing...
information on the differences between individuals in brain activation, functional imaging can be applied in the validation of theoretical models about a genetic basis of the lateralization of higher cognitive functions (McManus, 1985; Annett and Alexander, 1996; Klar, 2003; Anneken et al., 2004; Weber et al., 2006).

For a long time, the reliable assessment of hemispheric dominance could only be achieved by invasive methods, for instance by the injection of amobarbital. By selectively narcotizing one hemisphere, the functional role of the other hemisphere for a particular cognitive function can be determined (Wada and Rasmussen, 1960). Since the technique carries the risk of fatal complications, it is only administered to neurosurgical candidates. Only with the advent of non-invasive imaging techniques such as functional magnetic resonance imaging (fMRI), functional transcranial Doppler sonography (Deppe et al., 2004), magnetoencephalography (Hirata et al., 2004) and infrared spectroscopy (Watson et al., 2004), it became possible to non-invasively investigate the lateralization of cognitive brain functions systematically in the healthy population.

Due to its high availability and good spatial resolution one of the most widely used imaging techniques is fMRI which can measure neuronal activity indirectly through changes in blood oxygenation level-dependent (BOLD) signal. Evaluating hemispheric lateralization on the basis of fMRI data, however, is problematic. Although several approaches have been established, the optimal method for evaluating the lateralization of a cognitive function remains controversial. To be a useful marker of interhemispheric brain organization, a suitable approach has to be robust and reproducible (Nagata et al., 2001).

Robustness

Laterality calculations depend on decisions concerning the brain activity measure and a precise region of interest (ROI). Both in turn depend on a number of external parameters, especially the statistical threshold used. These choices are partially constrained by the questions being addressed, but additional arbitrary decisions remain. A suitable method for evaluating the lateralization of a cognitive function has to be robust, that is, the LI must not vary strongly with external parameters.

Reproducibility

fMRI measurements are not only variable across subjects, but also within subjects when tested repeatedly. In the same subject, the magnitude and extent of brain activation can change considerably. Therefore, a suitable method for evaluating the lateralization of a cognitive function has to be reproducible, that is, yield similar results in repeated measurements.

Several studies investigated the reproducibility of fMRI laterality measures (e.g., Nagata et al., 2001; Rutten et al., 2002; Liegeois et al., 2002; Adcock et al., 2003; Knecht et al., 2003; Fernandez et al., 2003; Lohmann et al., 2004; Baciu et al., 2005), but none of them systematically compared all established approaches. Furthermore, it has been shown that some methods, although perfectly valid for some tasks, do not describe hemispheric lateralization correctly for others (Jansen et al., 2004a). As a consequence, one would have to apply different evaluation strategies to different types of cognitive tasks. In the present study, we therefore systematically evaluated different approaches to determine lateralization for a number of cognitive tasks. We attempted to identify the best approach to describe the lateralization of a cognitive function, that is, an approach yielding the most robust and reproducible results.

Methods

Subjects

Ten healthy volunteers (6 men), aged 24 to 31 years, participated in the study. All subjects had completed the equivalent of a high school degree (“Gymnasium”). None had a history of serious medical, neurological or psychiatric illnesses or any brain pathology or abnormal brain morphology on T1-weighted MR images. All subjects gave their written informed consent prior to participation, according to the declaration of Helsinki. Nine subjects were right-handed and one subject left-handed, as assessed by the Edinburgh handedness inventory (Oldfield, 1971). Each subject was investigated twice within 1 day, but in two independent sessions. The time interval between both measurements was between 2 and 2.5 h.

Experimental design

Subjects performed four different tasks, three language tasks (word generation, semantic decision, picture naming) and a spatial attention task (Landmark task). The sequence of tasks was randomized across subjects. Stimuli were presented visually, and for each of the two sessions, a different version was used. Before the actual experiment, subjects were trained (both outside and inside the scanner) in each task to ensure that they understood the instructions.

Word generation task (WGT)

Word generation tasks have been employed in numerous studies and are known to produce marked lateralization of brain activation (Cuenod et al., 1995; Hertz-Pannier et al., 1997; Pujol et al., 1999; Jansen et al., 2004b).

The paradigm consisted of an activation and a control task lasting 20 s each. Tasks were performed in an alternating sequence and repeated 8 times. Before the first and after the last block, subjects were presented with a blank screen for 10 s and 5 s, respectively. Activation task: 5 s after an exclamation mark (presented for 0.5 s to tell subjects that a block ended and another will start soon), subjects saw a letter for 2.5 s on a computer screen. They silently had to find as many words as possible starting with the displayed letter. All letters were displayed only once. After 15 s, another exclamation marked the beginning of a new block. Control task: the meaningless word “baba” was presented on the screen instead of a letter. Subjects had to silently repeat the phrase “baba”.

Semantic decision task (SDT)

The synonym detection task is based on a semantic decision task as described in Fernandez et al. (2001). This task is known to provide a robust identification of language areas as Broca’s and Wernicke’s.

The paradigm consisted of a semantic decision and a control task lasting 26 s each. Tasks were presented in an alternating sequence and repeated 7 times. Before the first and after the last block, subjects were presented with a blank screen for 10 s. Task description:
subjects were visually presented with either two words (semantic decision task) or letter strings (control task) above and below a fixation point every 4.125 s. The stimuli were presented for 4 s, and each stimulus was followed by a short interstimulus interval of 0.125 s (the last interstimulus interval in each block lasted 1.375 s). Each block consisted of 6 words/letter strings with a probability of 50% for synonym words/identical letters. Subjects were instructed to decide whether the two words represented a synonym (like car-auto) and whether the consonant strings were identical (like dqfgh–dqfgh), respectively, and report their decision by pressing corresponding buttons with their right hand on a response pad.

**Picture naming task (PNT)**

Picture naming is commonly used a language task in fMRI studies (e.g., Rutten et al., 2002).

The paradigm consisted of an activation and a control task lasting 20 s each. Tasks were performed in an alternating sequence and repeated 10 times, starting with the active condition. Before the first block, subjects were presented with a blank screen for 10 s. In the active condition, a picture was presented every 2 s, which subjects were instructed to name silently. Ten objects were presented in each block. Each stimulus was presented for 1.5 s, followed by an interstimulus interval of 0.5 s. The stimulus consisted of black and white drawings of common objects, taken from the standardized picture set by Snodgrass and Vanderwart (1980). In the control condition, subjects were instructed to fixate a fixation point in the middle of the screen.

**Landmark task (LT)**

For the purpose of functional imaging studies, hemispheric lateralization for attention may be assessed in a number of ways (for a review, see Cabeza and Nyberg, 2000). In the present study, we chose to examine lateralization for spatial attention by the Landmark task (Fink et al., 2000). Subjects had to decide whether prebisected horizontal lines are correctly bisected. The paradigm consisted of an activation and a control task lasting 30 s each. Tasks were performed in an alternating sequence and each task was repeated twice. Each block was preceded by a 10-s interstimulus interval. This “baseline” period included a 6-s presentation of the instruction for the respective task. Before the first and after the last block, subjects were presented with a blank screen for 10 s. In the activation task, subjects had to decide whether prebisected horizontal lines were correctly bisected. In the control task, subjects detected whether the horizontal lines contained a transection mark irrespective of the position of that mark. Subjects were instructed to report their decision by pressing corresponding buttons with their right hand on a response pad. Twenty stimuli were presented in each block. Each stimulus was presented for 0.45 s followed by an interstimulus interval of 1.05 s. Stimuli were presented in varied positions on the screen in order to prevent subjects from holding the representation of a single fixed point as the center of all lines.

**MRI data acquisition**

All MRI data were collected with a Siemens Vision 1.5 T whole body scanner equipped with a standard head coil. Functional images were acquired using a T2*-weighted gradient echo EPI sequence (TE = 60 ms, TR = 2 s, flip angle 80°, slice thickness 5 mm, gap 1 mm, FOV 200 mm, in-plane resolution 3.125 x 3.125 mm). Twenty axial slices were taken, oriented in parallel to the AC–PC line and covering the whole head. In addition, high-resolution anatomical images were acquired for each subject (TE = 3.88 ms, TR = 1790 ms, flip angle 30°, matrix 256 x 256, FOV 256 mm, 160 sagittal slices, slice thickness 1 mm).

**MRI data analysis**

Imaging data were analyzed with SPM99 (available online at http://www.fil.ion.ucl.ac.uk/spm). The first five images of each session, during which the MR signal reaches a steady state, were discarded. The remaining scans of each individual were realigned to the first image (corresponding to the sixth acquired image of the time series) to correct for movement artefacts. Sinc interpolation was used in the transformation. A mean functional image volume was constructed for each subject from the realigned images. The functional images were normalized to a template brain image created by the Montreal Neurological Institute (MNI) (Evans et al., 1994). The mean functional image was used to determine the parameters for the spatial normalization process (Friston et al., 1995). The resulting pixel size in standard stereotactic coordinates was 3 x 3 x 3 mm3. The normalized images were spatially smoothed using an isotropic Gaussian kernel (FWHM 10 mm).

Subsequently, data were bandpass filtered (hrf as lowpass filter, cut-off period of the highpass filter was set using standard settings) and statistically analyzed in the context of the general linear model, using boxcar regressors convolved with the canonical hemodynamic response to model responses during each experimental condition. The tasks and sessions were analyzed separately, i.e., a single design matrix was used for each task/session. Fitting the boxcar function to the time series at each voxel results in a parameter estimate image (a β-image). This image indicates how strongly the waveform fits to the fMRI data at each voxel. By applying appropriate linear contrasts to the parameter estimates, a weighted β-image (a “con-image”) is calculated, displaying signal changes between two conditions. The weighted parameter estimate is divided by its standard error, which is derived from the residual model after the complete model has been fit, and is thus converted to a t-statistic image. These t-statistics constitute a statistical parametric map (SPM) and can be interpreted by referring to the probabilistic behavior of Gaussian random fields.

The anatomical localization of activated brain regions was determined using automated anatomical labeling (AAL) (Tzourio-Mazoyer et al., 2002). AAL is a software package for the anatomical labeling of functional brain mapping experiments and is based on an anatomical parcellation of the spatially normalized single subject high resolution T1 volume provided by the MNI.

**Determination of a lateralization index**

A lateralization index was calculated by the formula

\[
LI = \frac{A_L - A_R}{A_L + A_R},
\]

where \( A_L \) and \( A_R \) refer to values of fMRI-measured activity for equal ROIs within the left (L) and right (R) hemisphere. A positive value of LI represents left-hemisphere dominance, a negative value right-hemisphere dominance. The intermediate values reflect varying degrees of laterality (Binder et al., 1996) (cf., Fig. 7).

Numerous factors can affect fMRI laterality results, for instance, acquisition parameters, subjects’ cooperation, and the
design of a paradigm. For data analysis, however, the choice of an activity measure and the definition of a precise ROI are the most important issues. Therefore, the present study will focus on these factors.

Choice of activity measure $A_L$ and $A_R$

In the last decade, several approaches have been established to describe the brain activity $A_{L,R}$ and to choose a suitable ROI. All approaches are to some degree dependent on external parameters (e.g., the statistical threshold at which voxels are declared as “active” (Binder et al., 1996; Deblaere et al., 2004; Seghier et al., 2004)). The most common measures of brain activity are either based on the extent of the activated brain region, i.e., the number of active voxels, or the magnitude of the fMRI signal change. In the following, an LI based on the extent of activated brain volume is labelled LIe, an LI based on the magnitude of signal change LIm.

LIs based on the extent of activated brain volume

A LI was calculated based on the number of active voxels in a given ROI using either fixed (approach 1, cf., Binder et al., 1996) or variable (approach 2, cf., Knecht et al., 2003) statistical thresholds across subjects. Using fixed thresholds, the LIs were calculated for a broad range of statistical thresholds ($p=10^{-2}$ to $p=10^{-5}$) in order to examine the dependence of LI on the statistical threshold. Using variable statistical thresholds, the individual significance level was set in each subject so that it corresponded to a fixed number of activated voxels ($3000\pm 100$ and $6000\pm 100$ active voxels, respectively for WGT, SDT, and LT; $2000\pm 100$ and $4000\pm 100$ active voxels for PNT).

LIs based on the magnitude of signal change

A LI was calculated based on the magnitude of signal change defined by the $t$ values. Either all voxels within a chosen ROI were considered (approach 3, signal intensity change, cf., Adcock et al., 2003) or only those that exceeded a predefined activation level (approach 4, “thresholded” signal intensity change, cf., Fernandez et al., 2001), using the following procedure: In a first step, for each subject and each ROI, a mean maximum activation value was calculated defined as the mean of those $5\%$ of voxels showing the highest level of activation in the respective ROI. In a second step, the threshold for inclusion in the calculation of the LI was set at $50\%$ of this mean maximum activation value. Only signal intensity changes of those voxels that exceeded the threshold were used to calculate the LI.

Choice of ROI

Cognitive functions are subserved by large-scale neurocognitive networks. Regions within this network show different extents of lateralization. By averaging over all activations, the LI is unlikely to reflect a systematic pattern of lateralization in the key participating brain regions. Thus, another important issue in laterality studies is how to determine a precise ROI in which the brain activity is measured.

ROIs can be defined either functionally (i.e., based on the pattern of activation) or anatomically (i.e., based on anatomic landmarks). Pure anatomical definitions often include areas that are not engaged by the task and due to the normalization procedure activations of interest may lie outside the anatomically chosen ROI (Fig. 1). Furthermore, macroscopic anatomic landmarks are rarely reliable indicators of cytoarchitectonic borders (Amunts et al., 1999). Functional ROIs are easier to handle. They might be derived, for instance, from data of a pilot study investigating the functional activation in another group of subjects or within the same cohort with similar paradigms (Fernandez et al., 2001; Adcock et al., 2003). However, functionally defined ROIs can include regions outside the actual area of interest. Thus, they often require additional masking (Fig. 2).

In the present study, we used a combination of functional and anatomical derived ROIs as described in Adcock et al. (2003) and Jansen et al. (2004a). In a first step, ROIs were derived from those voxels that were active at $p<0.01$ in at least $80\%$ of the subjects. (In the LT, the general activation was lower. Here, we had to choose the more liberal threshold of $p<0.05$) Those commonly activated brain regions were predominantly located in one hemisphere. Corresponding ROIs in the other hemisphere were generated by reflection through the midline. In a second step, we used prior neuroanatomical knowledge to differentiate between regions of interest and regions of no interest and to separate contiguous activated brain regions encompassing several different regions of interest (Fig. 2).

For the language tasks, regions of interest comprised the prefrontal cortex (subdivided into Broca’s area and the remaining regions) and for the SDT also the temporo-parietal cortex (cf., Fernandez et al., 2001). Regions of no interest comprised all remaining regions that are possibly activated during the language tasks (e.g., cerebellum, supplementary motor area, visual cortex; cf., Fig. 2). For the LT, we chose as regions of interest brain areas in the frontal and parietal cortex (cf., Jansen et al., 2004a). Again, other brain regions such as the visual cortex, the supplementary motor area or the cerebellum were considered as regions of no interest.

Additionally, we calculated a LI in a large, purely anatomically defined ROI including all cerebral regions and excluding the cerebellum. In this ROI brain activity related to a number of different cognitive operations is averaged. Therefore, we expected that the test–retest reproducibility with respect to the degree of lateralization for LIs calculated in large brain regions is lower than for smaller and functional more specific ROIs.

Choice of paradigms

LIs were calculated for each of the four individual tasks (WGT, SDT, PNT, LT). Additionally, we used a combined task analysis (CTA) approach proposed by Ramsey et al. (2001) to calculate LIs based upon the combination of the three language tasks (WGT, SDT, PNT). By combining several language tasks, this approach is geared toward identifying brain regions that are involved in generic language functions rather than regions that are involved in functions that are specific to a single task. Since the CTA is conceptually similar to a conjunction analysis implemented in SPM99 (Friston et al., 1999; Ramsey et al., 2001), we used this

---

2 The magnitude of signal change is in fMRI studies commonly described by weighted $t$ values instead of $t$ values since $t$ values are driven by the variance of the signal as well. However, with regard to robustness and reproducibility, both approaches yielded similar results (see Results and discussion).
approach to calculate a combined language activation pattern for each subject.

Reproducibility

For each approach, we first tested whether hemispheric dominance was determined reliably. We then assessed whether also the degree of lateralization can be calculated reproducibly (Fig. 3).

Hemispheric dominance: left, right or bilateral

In a first step, we analyzed whether an approach describes hemispheric dominance reproducibly. For this purpose, a subject was classified as left-hemisphere dominant for positive values of LI and as right-hemisphere dominant for negative values. Any approach that could not even reproducibly determine hemispheric dominance was discarded as inappropriate. Many authors introduce additionally a third category to describe hemispheric dominance. Subjects whose brain activation pattern is described by a LI with |LI| < 0.20 are classified as bilateral (e.g., Springer et al., 1999). Therefore, we also investigated whether a subject can reproducibly be categorized as bilateral.

Degree of lateralization: |LI|

In a second step, we investigated which of the approaches determine not only hemispheric dominance, but also the degree of lateralization reproducibly. Test–retest reproducibility was determined by the Pearson product moment correlation coefficient \( r \), which is based on the correlation between individual LIs from repeated examinations relative to the variance of the indices. \( r \) is defined as

\[
\hat{r} = \frac{\sum_{k=1}^{n} (LI_{k1} - \overline{LI}_1) \cdot (LI_{k2} - \overline{LI}_2)}{\sqrt{\sum_{k=1}^{n} (LI_{k1} - \overline{LI}_1)^2 \cdot \sum_{k=1}^{n} (LI_{k2} - \overline{LI}_2)^2}},
\]

where \( LI_{k1} \) represents the LI of the \( k \)th subject for the first examination and \( LI_{k2} \) the LI of the \( k \)th subject for the second examination.

\[
\overline{LI}_i = \frac{1}{n} \sum_{k=1}^{n} LI_{ki}
\]

is the corresponding average laterality index from \( n \) subjects during the \( i \)th examination. Only approaches with a Pearson correlation coefficient of at least 0.5 were considered as appropriate to describe the lateralization of a cognitive function.

To convey information about the implementation of cognitive brain functions, a LI must be able to discriminate lateralization differences also in the healthy population. If an approach yields only small variance across subjects, it does not fulfill this criterion. Therefore, any approach that indicated the degree of lateralization in more than 50% of the subjects as LI = 1 was considered as inappropriate to describe differences in lateralization across subjects, since it is merely a marker of hemispheric dominance.

All approaches were first tested on the WGT and the SDT. Both tasks are known to lead to stable and lateralized brain activation also in individual subjects and not only on a group level. “Stable” means that core brain regions are activated at typical statistical thresholds (\( p \sim 10^{-3} \)) on an individual level in most of the subjects. “Lateralized” means that in subjects with typical hemispheric dominance the brain activation is unevenly distributed over the hemispheres; if not, the best approach might not be able to produce a reliable LI. Those approaches that did not even determine hemispheric dominance reliably in these “convenient” tasks were disapproved.

Second, we tested whether the remaining approaches could also be applied to describe functional brain lateralization for the PNT and the LT. We considered these tasks as less “convenient” for the determination of brain lateralization. In the PNT, the individual brain activation pattern is typically less lateralized than in the WGT and the SDT (Rutten et al., 2002). Therefore, it was unclear whether a reproducible LI can be determined at all. In contrast, although the LT leads to a strongly lateralized brain activation pattern on a group level (Fink et al., 2000), on an individual basis brain activation is rather weak and in many subjects only observable at very liberal statistical thresholds.

Third, we investigated whether the CTA performs better than the single language tasks.

Results and discussion

Brain activation pattern and regions of interest

The brain regions that were activated during the cognitive tasks employed in the study are shown in Fig. 5 for a representative...
subject. Appendix 1 lists additionally the major activated brain regions for all tasks during the first session (as obtained by a random effect group analysis, $p < 0.001$ uncorrected). The three language tasks (WGT, SDT, PNT) activated a large-scale neuronal network. While some brain regions were active in all three language tasks (as indicated by the CTA), others were specific to a single task. The visuo-spatial attention task (LT) activated a cognitive network, in which the main activation centers were located in the anterior cingular cortex, the lateral parietal cortex and the frontal cortex.

Fig. 2. Brain activation during the word generation task. Regions of interest are defined functionally as those voxels in lateral prefrontal cortex that are activated in at least 80% of the subjects. Additional masking was needed to differentiate between regions of interest (prefrontal cortex) and regions of no interest (e.g., supplementary motor cortex (SMA) and cerebellum). A differentiation between different areas of interest is not always simple: the prefrontal activation encompasses Broca’s area as well as other prefrontal regions. The borderline between these regions remains to a certain degree arbitrary.

The functionally defined ROIs, derived from a combination of the individual activation patterns, are displayed in Fig. 4. Activated brain regions were predominantly located in the left hemisphere for the language tasks and the right hemisphere for the Landmark task. Corresponding ROIs on the other hemisphere were generated by mirror projection through the midline. Using prior neuroanatomical knowledge, we differentiated between regions of interest (e.g., activation in Broca’s area) and regions of no interest (e.g., activation in the visual cortex during the PNT).

In the PNT, the strongest activation was not found in language areas but in the visual cortex (Fig. 5), because activation and control condition had a different visual input. To calculate the LI for the whole hemisphere, we discarded the activation in the visual cortex (i.e., voxels with $y < -32$ were not considered). Otherwise, the strong bilateral visual activation would have always indicated a fairly bilateral lateralization, independent of the language organization. Since the beginning of fMRI, the most popular approach in practice has been to calculate $A_L$ and $A_R$ by the volume of significantly activated brain above a given statistical threshold $p$ in a given ROI (approach 1, Fig. 6). A drawback of this measure is its dependence on the statistical threshold $p$. At low thresholds (high $p$ values), the occurrence of false positive voxels in both hemispheres typically leads to a bilateral activation pattern, moving LI towards zero. At high thresholds (low $p$ values) the number of active voxels decreases, false positive voxels as well as “truly” activated voxels are excluded. $L_{Ie}$ approaches $\pm 1$. Thus, $L_{Ie}$ can express almost the complete continuum between bilaterality and fully lateralized brain activity (Fig. 7).

Because of the strong dependence of $L_{Ie}$ on the statistical threshold $p$, its absolute value and thus the degree of lateralization are as arbitrary as the choice of $p$. Furthermore, $L_{Ie}$ based on fixed statistical thresholds has in no way the capability of intersubject comparisons. Due to the high intersubject variability in activation levels, two subjects may have a similar activation pattern and thus a similar LI, but at completely different statistical thresholds (Fig. 8). Correspondingly several studies have shown that individually adapted thresholds are better then fixed thresholds for analysing single subject data to obtain comparable results for several individuals and increase intrasubject reliability (Moser et al., 1996; Specht et al., 2003; Klöppel and Büchel, 2005). Robustness and reproducibility of $L_{Ie}$ and $L_{Im}$

$L_{Ie}$ based on the extent of activated brain volume ($L_{Ie}$)

Robustness. Since the beginning of fMRI, the most popular approach in practice has been to calculate $A_L$ and $A_R$ by the volume of significantly activated brain above a given statistical threshold $p$ in a given ROI (approach 1, Fig. 6). A drawback of this measure is its dependence on the statistical threshold $p$. At low thresholds (high $p$ values), the occurrence of false positive voxels in both hemispheres typically leads to a bilateral activation pattern, moving LI towards zero. At high thresholds (low $p$ values) the number of active voxels decreases, false positive voxels as well as “truly” activated voxels are excluded. $L_{Ie}$ approaches $\pm 1$. Thus, $L_{Ie}$ can express almost the complete continuum between bilaterality and fully lateralized brain activity (Fig. 7).

Because of the strong dependence of $L_{Ie}$ on the statistical threshold $p$, its absolute value and thus the degree of lateralization are as arbitrary as the choice of $p$. Furthermore, $L_{Ie}$ based on fixed statistical thresholds has in no way the capability of intersubject comparisons. Due to the high intersubject variability in activation levels, two subjects may have a similar activation pattern and thus a similar LI, but at completely different statistical thresholds (Fig. 8). Correspondingly several studies have shown that individually adapted thresholds are better then fixed thresholds for analysing single subject data to obtain comparable results for several individuals and increase intrasubject reliability (Moser et al., 1996; Specht et al., 2003; Klöppel and Büchel, 2005). There have been several attempts to deal with the dependence of $L_{Ie}$ on arbitrary statistical thresholds and to cope with the high interindividual variability in general activation levels. In plotting $L_{Ie}$ as a function of the statistical threshold, Binder et al. observed that, in most cases, the $L_{Ie}$ curve reaches a plateau for a relatively broad range of thresholds (Binder et al., 1996) (cf., Fig 7 right). Calculating $L_{Ie}$ for $p$ values within such a plateau makes the value less dependent on external parameters, because it allows for some variation in statistical thresholding without changes in $L_{Ie}$. However, in many subjects, we did not observe such a plateau. This finding is in agreement with other studies, e.g., Deblaere et al. (2004). Therefore, this approach was not applicable in the present study.

---

4 It may have been more judicious to use a different control condition that allows removing low-level stimuli features (i.e., visual processing) from the main effect. For example, instead of fixation point, presenting meaningless pictures (e.g., random patterns or scrambled line drawings; see Vingerhoets et al., 2003) during the control condition might have avoided detecting huge and bilateral low-level visual areas.
In the present study, we used an approach proposed by Knecht et al. (2003). The authors set the level of significance in each subject so that it corresponded to a fixed total number of activated voxels (approach 2, Fig. 6). On the assumption that the subjects should engage the same amount of brain tissue—they all performed the same task at more or less the same level—they postulated that the extent of activated brain volume is similar for all subjects. The number of activated voxels was chosen "reasonably," that is, in accordance with the results of lesion studies as well as with other functional imaging studies. In principle, this approach is more robust and, from a conceptual point of view, also more reasonable than the application of fixed statistical thresholds. However, an arbitrary decision remains, that is the choice of the total number of activated voxels. LI threshold dependency in approach 1 is replaced by LI extent dependency in approach 2. Furthermore, this approach is problematic in subjects with a low activation level. In these subjects, the LI is based on "active" voxels that are subject to high type I errors.

Reproducibility

Hemispheric dominance. LI, based either on fixed or variable thresholds, determined hemispheric dominance reliably.

It is noteworthy that for two subjects a change of the statistical threshold reversed the sign of the lateralization index and consequently hemispheric dominance (WGT, Broca's area). At low thresholds LI indicated left-, at high thresholds right-hemisphere dominance in one of the runs. This discrepancy is explained by a more detailed analysis of their respective activation pattern: the activation cluster in the left hemisphere is spatially more extended, while the activation cluster in the right hemisphere is more significantly (in terms of t values) activated. Therefore, at

---

**Fig. 3.** Schematic overview describing the procedure chosen to assess the reproducibility of a LI. All approaches were first tested on the WGT and the SDT. Those approaches that did not even determine hemispheric dominance reliably in these "convenient" tasks were disapproved. The remaining approaches were then tested on the PNT, LT and the CTA.

---

**Fig. 4.** Graphical depiction of the functionally defined ROIs, overlaid onto the left hemisphere of the standard SPM99 template. Abbreviations used: WGT—word generation task, SDT—synonym detection task, PNT—picture naming task, CTA—combined task analysis, LT—Landmark task.
low thresholds, these subjects have more activated brain volume in the left hemisphere. At high thresholds, only the most significant activated cluster survives, now indicating right hemispheric dominance.

Since in these cases the change of the statistical threshold can reverse the sign of the lateralization index, lateralization curves (LI = LI(p)) instead of a single lateralization index are a more comprehensive estimate of hemispheric dominance. Hemispheric dominance should therefore not be determined at one single statistical threshold.

**Degree of lateralization.** Pearson’s correlation coefficient $r$, describing the reproducibility of $|LI|$ between two sessions, ranged for fixed statistical thresholds from −0.34 to 0.19 for the SDT and from 0.20 to 0.92 for the WGT (Fig. 9). For variable statistical thresholds Pearson’s $r$ ranged from −0.23 to 0.68 for the SDT and from 0.48 to 0.93 for the WGT. Unexpectedly, the reproducibility was higher for the whole hemisphere than for the functionally defined language ROIs. Especially in the functionally defined ROIs, LIe indicated fully lateralized brain activity (LI=1) in many subjects if fixed statistical thresholds were applied. For example, at $p=0.001$ more than half of the subjects did not show any subdominant activation in either of the functionally defined ROIs during the SDT. The same applies for Broca’s area for the WGT (Fig. 9). Therefore, LIe based on fixed statistical thresholds does not discriminate lateralization differences within healthy subjects with typical hemispheric dominance, but is rather a marker for hemispheric dominance alone. The situation is somewhat different.

Fig. 5. Brain activation patterns for a representative subject during the cognitive tasks employed in the study. Activations are rendered on the surface of the standard SPM99 template. Abbreviations used: WGT—word generation task, SDT—synonym detection task, PNT—picture naming task, CTA—combined task analysis, LT—Landmark task.
Determination of a LI

**Definition:** \[ LI = \frac{A_L - A_R}{A_L + A_R} \]

**Choice of activity measure \( A_{L/R} \)**
- #1: active voxels, fixed statistical thresholds for all subjects/measurements
- #2: active voxels, variable thresholds – normalized by a fixed number of active voxels for all subjects/measurements
- #3: signal intensity change, based on all voxels in ROI
- #4: signal intensity change, based only on those voxels with highest activation in each ROI

\( LI_e \)
\( LI_m \)

**Choice of ROI**
- #1: defined by those voxels that are active at \( p=0.01 \) in at least 80% of subjects
- #2: whole hemisphere

for \( LI_e \) based on variable thresholds. However, one has to be aware that this is only made possible because many “active” voxels are used for the calculation of \( LI \) that are subject to a high type I error.

**LIs based on the magnitude of signal change (\( LI_m \))**

Robustness. \( LI_m \) based on the magnitude of signal intensity change between two conditions (approach 3, Fig. 4), is in principle a more robust measure of brain activation than \( LI_e \) because no choice of threshold is involved in calculating these indices. However, this measure has its own set of problems. Its robustness is strongly depending on the chosen ROI. The implicit assumption for the calculation of \( LI_m \) is that one deals with activations, i.e., stronger activation in the activation task than in the control task. If there is a deactivation in the ROI of the subdominant hemisphere, i.e., stronger activation in the control task than in the activation task, \( LI_m \) can become greater than \( \pm 1 \) and loses a meaningful metric. This is by no means an isolated case. In the present study in more than half of the subjects the \( LI \) is greater than \( \pm 1 \) in at least one measurement.

A solution to this problem has been proposed by Fernandez et al. (2001). They did not account for all voxels in a chosen ROI, but only those with the strongest activation (approach 4, Fig. 6). In a first step, they calculated for each subject and each ROI a mean maximum activation value defined as the mean of those 5% of voxels showing the highest level of activation in the respective ROI. In a second step, they set the threshold for inclusion in the calculation of the \( LI \) at 50% of this mean maximum activation value. Only signal intensity changes of those voxels that exceeded the threshold were used to calculate \( LI_m \). Weakly activated voxels, possibly showing a deactivation, were excluded.

**Reproducibility**

Hemispheric dominance. \( LI_{sw} \) based on all voxels within a ROI, did not determine hemispheric dominance reproducibly in about 20% of the subjects. This is explained by a more detailed analysis of their respective activation pattern: if deactivation in the subdominant hemisphere is greater than the activation in the dominant hemisphere, the sign of the lateralization index is

---

Fig. 6. Overview: definition of \( LI \), choice of activity measures \( A_e \) and \( A_r \), choice of ROIs. LIs are either based on the extent of the activated brain region (\( LI_e \)) or the magnitude of signal intensity change (\( LI_m \)).

Fig. 7. Lateralization index \( LI \) as a function of the statistical threshold \( p \) for three subjects. The \( LI(p) \) curves of both measurements are shown. \( LI_e \) is calculated by the volume of significantly activated brain volume during word generation in the left and right cerebral hemisphere, respectively. Depending on the threshold, \( LI_e \) expresses almost the complete continuum between bilaterality and fully lateralized brain activity. The curve progression \( LI(p) \) is reproducible in some subjects, but not in others.
reversed, i.e., LI_m denotes the subdominant hemisphere as dominant (Fig. 10). This problem is solved if only those voxels with the strongest activation in a chosen ROI are taken into account. LI_m based on those voxels with the highest activation in each ROI, always determines hemispheric dominance reliably.

Degree of lateralization. Since no dichotomous thresholding (‘active or not active’) is involved in calculating LI_m, the activity measures in the left as well as in the right hemisphere were dissimilar from zero. LIs indicating fully lateralized brain activity (LI =1) were therefore the exception. Unlike LI_e, LI_m was therefore in all cases appropriate to describe differences in the degree of lateralization across subjects. Pearson’s r was only calculated for approach 4 since approach 3 did not even determine hemispheric dominance reliably. It ranged from 0.61 to 0.66 for the WGT and from −0.32 to 0.49 for the SDT (Fig. 9).

Degree of lateralization. Since no dichotomous thresholding (‘active or not active’) is involved in calculating LI_m, the activity measures in the left as well as in the right hemisphere were dissimilar from zero. LIs indicating fully lateralized brain activity (LI =1) were therefore the exception. Unlike LI_e, LI_m was therefore in all cases appropriate to describe differences in the degree of lateralization across subjects. Pearson’s r was only calculated for approach 4 since approach 3 did not even determine hemispheric dominance reliably. It ranged from 0.61 to 0.66 for the WGT and from −0.32 to 0.49 for the SDT (Fig. 9).

Fig. 8. Activated brain regions associated with a word generation task presented as through projection onto a transversal representation of stereotactic space. Both subjects show similar activation patterns, but at different statistical thresholds. Given the high intersubject variability in activation levels, for an intersubject comparison no generally predefined threshold should be applied. Instead, statistical thresholds have to be individually adjusted. Abbreviations used: L—left, R—right.

Fig. 9. Pearson’s correlation coefficient r describing the reproducibility of the degree of lateralization between two sessions for (a) WGT and (b) SDT using different approaches (cf., Fig. 6): (1) LI based on active voxels, fixed thresholds, (2) LI based on active voxels, variable thresholds, (4) LI based on “thresholded” signal intensity change, since approach 3 (LI based on “unthresholded” signal intensity change) did not even determine hemispheric dominance reliably in about 20% of the cases, we refrained from determining the reliability of the degree of lateralization. All LIs were calculated in different ROIs: Broca’s area, remaining prefrontal area (rPF), temporo-parietal cortex (TP), and whole hemisphere (WH). Please note that approach 4 was not applicable to the whole hemisphere. Any approach that indicated fully lateralized brain activity (LI=1) in more than half of the subjects was considered as inappropriate to describe differences in the degree of lateralization across subjects (cf., Fig. 3). In these cases Pearson’s r can be calculated, but does not yield useful information. This is indicated by an ‘X’.

Fig. 10. Activation pattern for a selected subject during the semantic decision task presented as through projection onto a transversal representation of MNI space. Different LIs are calculated for a ROI in Broca’s area: based on the extent of activation, the subject is clearly left-hemisphere dominant for language (LI_e (Broca))=1.0, p<10^{-3}) with no activation in Broca’s homologue on the right hemisphere (Syn>Con, Synonym detection>Control task). If the laterality index is calculated by the magnitude of the mean signal intensity change, the subject displays right-hemisphere dominance (LI_m (Broca) =−1.61). A more detailed analysis of its activation pattern reveals that this does not describe the “true” hemispheric dominance. The implicit assumption was that one deals with activation in both the left and the right ROI. As one can see in the reverse contrast (Con>Sync, Control task>Synonym detection), there is a strong “deactivation” in the right ROI, i.e., stronger activation in the control task than in the actual activation task. The deactivation in the subdominant hemisphere is greater than the activation in the dominant hemisphere. This leads to a reversal of laterality. This problem can be solved if only those voxels with the strongest activation in a chosen ROI are taken into account. Applying this approach to the example above, the LI based on magnitude of signal change changes from LI_m =−1.61 (all voxels in ROI, indicating hemispheric dominance wrongly) to LI_m=0.41.

Abbreviations used: L—left, R—right.
Bilaterality

So far we classified subjects as left-hemisphere dominant for LIs greater than zero and as right-hemisphere dominant for LIs less than zero. Many authors introduce a third category and classify subjects as bilateral for LIs around zero (e.g., \( |LI| < 0.2 \) (Springer et al., 1999)). To challenge this use, we investigated in a further step whether a subject can be reproducibly categorized as bilateral. In over 90% of the cases, subjects were not reproducibly categorized as bilateral, independent of ROI and paradigm. Furthermore, even if a subject was classified as bilateral by one method of calculation, it was not by the other methods (not even in the same measurement).

Some subjects are likely to have a more bilateral brain organization (especially where functional reorganization occurred, e.g., in epileptic patients). However, as was shown by the present data, a healthy subject might be easily classified as bilateral by one method of calculation although this classification is neither reproducible in a second measurement nor supported by other methods of calculation. Therefore, bilaterality must be dealt with caution. In our view, a subject should only be classified as “bilateral” if there is support from different methods of calculation of hemispheric dominance, preferably by repeated measurements. Furthermore, it has to be kept in mind that ultimately the categorization into left, right, or bilateral remains to some degree arbitrary unless there is an independent validation by other measures.

Reproducibility of LIs calculated for PNT, LT and CTA

Picture naming task

The individual brain activation pattern in the PNT is typically less lateralized than in the WGT or the SDT. For about one third of the subjects hemispheric dominance could not be determined reproducibly. This was independent of ROI and method of calculation. Furthermore, even in one measurement, the results were ambiguous. Often subjects were classified as being left-hemisphere dominant by one method of calculation and right-hemisphere dominant by other methods of calculation. This was distinct from the results obtained during WGT and SDT. For these tasks, all approaches yielded unambiguous results. On the basis of fMRI data, it was therefore difficult to determine hemispheric dominance at all with the PNT.

Landmark task

The LT led to strongly lateralized brain activation on the group level. However, the individual brain activation patterns of many subjects were not stable, i.e., activation was only observed at rather liberal statistical thresholds. This resulted in decreased reliability of hemispheric dominance. Consequently, hemispheric dominance could not be determined reproducibly for about one fourth of the subjects. Generally, reproducibility of hemispheric dominance was better for those subjects who showed more extended activation. For “unstable” activation tasks such as the LT, it is therefore necessary to introduce a “criterion of stability”. Hemispheric dominance should only be determined if a subject shows “sufficient” activity (e.g., 1000 activated voxels) at a typical statistical threshold (e.g., \( p = 10^{-7} \)).

Combined task analysis

The CTA aims at selectively detecting activity that is common to all tasks. It favors detection of those regions that are active in all tasks over those that are not (Ramsey et al., 2001; Rutten et al., 2002). As expected, the severalfold increase in the number of scans led to a clear increase in sensitivity and consequently in robustness. Stable activity could still be observed at much more conservative statistical thresholds than for the individual tasks. However, the combination of the three language tasks led to an emphasis on the left-hemispheric language areas at the cost of homologous right-hemispheric brain regions. Generally, the LI was higher for the CTA than for the individual tasks. In the language ROIs the LI indicated fully lateralized brain activity (LI=1) for most of the subjects.

Hemispheric dominance could always be determined reproducibly with LIe (based either on fixed or variable statistical thresholds). LIm (even if accounting only for those voxels with the strongest activation in a ROI) yielded values greater than \( \pm 1 \) in at least one measurement in 3 of 10 cases for the “remaining prefrontal ROI” and in 1 of 10 cases for Broca’s area. In both ROIs hemispheric dominance was not determined reproducibly in one case. Again, this finding is explained by a high “deactivation” in the subdominant hemisphere. However, in contrast to the single language paradigms, the CTA reduces the activation in the subdominant hemisphere so strongly that deactivation can also be found in those voxels that are strongest activated. Again, the classification “bilateral” could not be determined reproducibly.

For LI, Pearson’s \( r \) ranged for from 0.34 to 0.58 for fixed statistical thresholds and from 0.36 to 0.70 for variable statistical thresholds. It has to be noted that for fixed thresholds in the functionally defined ROIs the degree of lateralization was denoted as LI=1 in more than half of the subjects. Overall the reproducibility for the CTA is comparable to the reproducibility for the WGT alone.

Comparison of LIe and LIm

We systematically evaluated the characteristics of different methods that describe the lateralization of a cognitive function in fMRI. We demonstrated that laterality calculations depend strongly on decisions concerning the activity measure and the ROI. At any rate, choices have to be made clear and it has to be shown how they influence the overall findings. Only then fMRI can provide meaningful laterality results.

The choice of the optimal method to calculate a LI depends on the specific question asked. In general, LIs based on the signal intensity change (LIe) are a more robust measure than LIs based on active voxels counts. However, the robustness of this approach strongly depends on the chosen ROI. ROIs have to capture activity related to the task component of interest and exclude noisy voxels. Our data show that the choice of such a ROI is a major problem that may become even more prominent in patients or subjects with gross morphological changes. A more suitable approach is therefore to compare BOLD intensity changes of only those voxels that are most strongly activated in each respective ROI. This approach can reproducibly determine hemispheric dominance as well as the degree of lateralization.

However, the other approaches should not be discarded completely since they have their own specific application fields. LIs based on active voxel counts at fixed statistical thresholds can be used as the most straightforward measure to describe hemispheric dominance. Since the degree of lateralization (and in some cases also the hemispheric dominance) depends on the significance threshold selected for the activation map, one should avoid to
calculate LI, at one single threshold. Instead the use of \( p \)-value-dependent lateralization plots (\( LI = LI(p) \)) are more suitable—especially from a clinical perspective.

The reproducibility of LIs based on active voxel counts at variable thresholds is comparable to the reproducibility of LIs based on thresholded signal intensity changes, but from an overall perspective the approach is less robust and — from a conceptual point of view — also more circuitous. However, LIs based on signal intensity changes are always restricted to small ROIs.\(^5\) If a ROI cannot be defined a priori or if one is interested in big ROIs (e.g., the whole hemisphere) LIs based on active voxel counts at variable thresholds are a good alternative if one is not only interested in hemispheric dominance but also in the degree of lateralization.

In most functional MRI, the “magnitude of signal change” is characterized by weighted \( \beta \) values. In contrast, Fernandez et al. also proposed to apply statistical \( t \) values (calculated by dividing the \( \beta \) values by the standard error) as a measure of signal strength (Fernandez et al., 2001). Since one of our approaches was designed similar to the one proposed by Fernandez et al., we used this measure as well. One has to be aware that the statistical \( t \) value is driven by the variance of the signal as well. We therefore additionally calculated LIs using weighted \( \beta \) values as activity measure. With regard to robustness and reproducibility, both approaches showed identical characteristics and yielded similar results. On the basis of our data, it was therefore not possible to decide which approach is more suitable.

However, from a conceptual point of view the use of weighted \( \beta \) values as activity measure might often be preferred over the use of \( t \) values. In our study, all subjects were investigated within the same day. Therefore, it is unlikely that the variance between sessions has changed much. If the time interval between the two sessions is large, this might be different, e.g., by changes of the scanner characteristics. If the variance changes differently for different brain regions, the reproducibility of the LI might therefore decrease if it is based on \( t \) values as activity measure. Therefore, the use of weighted \( \beta \) values might be the better choice. Furthermore, it might be also useful to incorporate repeated sessions in one single design matrix. This yields the assessment of a common (i.e., single) estimated residual variance and therefore more specificity to the weighted beta values as a marker of the between-sessions differences.

**Comparison with other studies**

Recently, several studies have investigated the characteristics of fMRI-based laterality measures (e.g., Nagata et al., 2001; Rutten et al., 2002; Liegeois et al., 2002; Adcock et al., 2003; Knecht et al., 2003; Fernandez et al., 2003; Lohmann et al., 2004; Jansen et al., 2004a; Baciu et al., 2005). It has been acknowledged by most authors that LIs based on “active” voxels counts at one single fixed statistical threshold yield neither robust nor reproducible results. Therefore, many studies advocate LIs based on the BOLD signal intensity change as a more robust approach. However, it has not yet been generally acknowledged how severe the robustness of this approach depends on the chosen ROI. Since offhand it is difficult to guarantee that a predefined ROI contains more “true” activation than noise (especially in the subdominant hemisphere), we would advocate to rely on thresholded signal intensity changes, i.e., to restrict the activity measure to those voxels that show the highest activation.

In functional neuroimaging, several techniques have been developed to evaluate the reproducibility of activation maps (Specht et al., 2003). Most studies use a simple linear correlation of LIs derived from the first and second investigation to describe the test–retest reliability. For a better comparison with these studies, we applied this approach as well. Generally, the reported correlation coefficients that describe the reproducibility between repeated measurements have to be interpreted with care. In many studies, the statistical values of the correlations are inflated by the inclusion of subjects with atypical hemispheric dominance. If and only if the correlations between the lateralization indices continue to remain stable after omitting these subjects from the analysis, one can be sure that high reproducibility values are not only caused by outliers. This fact might explain some discrepancies in laterality research. Often the reported agreement between language lateralization measures obtained by fMRI and other imaging techniques (e.g., the intracarotid amobarbital procedure) is much higher than the reproducibility of the fMRI lateralization index itself.

Another important method to directly assess language lateralization with fMRI was proposed by Liegeois et al. (2002). The authors base lateralization on the direct statistical comparison of the magnitude of task-induced activation in homotopic brain regions. The main advantage of this method is that it identifies regions in which left- and right task-induced activations are statistically different. However, this approach denotes hemispheric dominance in the majority of cases as LI = \( \pm 1 \) and is therefore not a suitable marker to describe lateralization differences between healthy individuals with the typical hemispheric dominance of their cognitive functions (Baciu et al., 2005).

Nagata et al. proposed another elegant way to minimize the threshold influence on the assessment of LI, (Nagata et al., 2001). The authors investigated the correlation between \( z \) scores and the number of activated voxels in a ROI and searched for monomial regression function that best described this relation. They found that the number of activated voxels was proportional to (\( 1/z \) score)\(^4\), with each hemisphere showing its own proportional constant. A LI was calculated from these 2 values. In the present study, we refrained from further following up this approach, because it did not become widely accepted throughout the functional imaging community and lacks a solid theoretical foundation (as acknowledged by the authors).

**Clinical implications**

To assess the clinical usefulness of a lateralization index, intrasubject reproducibility is one of the most important issues. For instance, to investigate language lateralization shifts after stroke, it has first to be shown that the lateralization of a cognitive function as assessed by fMRI is a reliable marker for the neural brain organization in healthy subjects. Otherwise, it is difficult to distinguish lateralization shifts due to reorganization from pseudo-shifts of lateralization caused by habituation (Lohmann et al., 2004).
However, the demonstration of robustness and reproducibility alone is not sufficient to evaluate the utility of a paradigm for clinical use. fMRI is a correlate technique, that is, an activation does not necessarily imply that the activated brain region is necessary for the task. It is well conceivable that an area is co-activated without direct involvement in the task. Therefore, the functional relevance of the LI has to be established. Patient studies are needed to precisely assess the predictive value of fMRI based LIs as a clinical marker, for example, in the presurgical evaluation of patients with intractable epilepsy (Klöppel and Büchel, 2005). However, robustness as well as reproducibility are necessary prerequisites for a LI to be of use in the clinical field. To assess the potential of fMRI based testing for clinical applications, it is essential to define how much of the variation in LI is a result of test–retest error.

One has to be aware that for the work with patients the reproducibility of a paradigm is not the only important criterion. Although the WGT showed a somewhat higher reproducibility than the SDT, it is has several disadvantages. First, with covert word generation tasks, it is difficult to measure behavioural responses. Therefore, the cognitive state and the cooperation of the patient is difficult to monitor. Second, activation is often restricted to the frontal part of the neural language network and does not assess temporoparietal regions.

Limitations of the study

In the present study, we did only investigate healthy subjects. It has to be noted that the overall test–retest reliability might be lower in clinical populations. However, the main characteristics of each approach should remain the same in both diseased and healthy subjects. Other factors may also complicate the determination of hemispheric dominance, e.g., the selection of ROIs in the presence of brain lesions, the signal change variability due to hemodynamic perturbations (D’Esposito et al., 2003).

The inclusion of 10 subjects is sufficient to evaluate the main characteristics of each approach (e.g., the dependence of LI on the statistical threshold or the problems arising by deactivation when using unthresholded signal intensity changes as activity measure). However, we have to acknowledge that 10 subjects might not be sufficient to differentiate properly between those approaches that fulfill our criterion of reproducibility ($r>0.5$).

Conclusions

In conclusion, we investigated the characteristics of several approaches that describe laterality in fMRI in terms of robustness and reproducibility. We showed that laterality calculations depend strongly on decisions concerning the activity measure. To provide meaningful results, this choice has to made clear and it has to be shown how it influences the overall results, since none of the tested approaches is fully free of the influence of external parameters (such as the statistical threshold for approach 1, extent threshold for approach 2, signal magnitude threshold for approach 4, ROI selection for all approaches).

Overall, the lateralization of a cognitive function can be best described by “thresholded” signal intensity changes where the activity measure is based on signal intensity changes, but restricted to those voxels in a region of interest that exceeded a predefined activation level.

Acknowledgments

This work was supported by the NRW-Nachwuchsguppe Kn2000 of the Nordrhein-Westfalen Ministry of Education and Research (Fö.1KS9604/0), the Interdisciplinary Center of Clinical Research Münster (IZKF Projects FG2 and Knec3/074/04), the Innovative Medizinische Forschung Münster (KN520301), the Deutsche Forschungsgemeinschaft (Kn 285/6-1 and 6-3), the Volkswagen Stiftung (Az.: I/80 708), as well as Marie Curie Research and Training Network: Language and Brain (RTN:LAB) funded by the European Commission (MRTN-CT-2004-512141) as part of its Sixth Framework Programme. We thank Sören Krach for helpful comments on an earlier manuscript version.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.neuroimage.2006.06.019.

References


