Jane Neumann: Beyond Activation Detection: Advancing Computational Techniques for the Analysis of Functional MRI Data. Leipzig: Max Planck Institute for Human Cognitive and Brain Sciences, 2012 (MPI Series in Human Cognitive and Brain Sciences; 141)

Beyond activation detection:
Advancing computational techniques for
the analysis of functional MRI data

Beyond activation detection:
Advancing computational techniques for
the analysis of functional MRI data

Der Fakultät für Mathematik und Informatik
der Universität Leipzig

eingereichte

Habilitationsschrift

zur Erlangung des akademischen Grades

Doctor habilitatus

(Dr. habil.)

vorgelegt

von

Dr. Jane Neumann

geboren am 21.08.1970 in Rochlitz

Die Annahme der Habilitationsschrift haben empfohlen:

1. Prof. Dr. Robert Turner, Max-Planck-Institut für Kognitions- und Neurowissenschaften Leipzig
2. Prof. Dr. Fred Hamker, Technische Universität Chemnitz
3. Prof. Dr. Gerik Scheuermann, Universität Leipzig

Beschluß über die Verleihung des akademischen Grades vom 16.7.2012

Acknowledgements

"One of the great joys of science lies in the moment of shared discovery."

James L. McClelland
David E. Rumelhart

For the past years, I was given the chance to work in a truly interdisciplinary, inspiring, and creative environment. Without the support of my colleagues and collaborators who created this unique atmosphere, this work would not have been possible.

Science...

First of all, I would like to thank Gabriele Lohmann and Karsten Müller who introduced me to functional neuroimaging and shared with me their invaluable knowledge and insights into the depths of fMRI data analysis. I wish to thank D. Yves von Cramon, Robert Turner and Arno Villringer for giving me the freedom to pursue my own research interest while, at the same time, providing generous support in all matters regarding my scientific life. I wish to thank the former and current scientific, administrative and technical staff as well as the students at the Max Planck Institute for Human Cognitive and Brain Sciences for their support over the years and the Max Planck Society (together with the tax payer) for providing outstanding research facilities.

A special thank goes to my colleagues and collaborators Birte Forstmann, Andreja Bubic, Tilmann Klein, Matthias Schroeter, Annette Horstmann, Gerhard Jocham, Markus Ullsperger, Jan Derrfuss, Jöran Lepsien, Andreas Schäfer, Stefan Zysset, Jens Brauer, Evelyn Ferstl and Marcel (Panda) Weiss for their invaluable input and the many serious and not so serious scientific discussions. I also wish to thank Birgit Mittag, Ingrid Schmude, Aline Peter and Christine Becker for their support in all

In memory of Matthias Müller, who once put a computer in front
of me saying: Look this could be interesting...

Contents

Chapter 1

Introduction

The wish to understand the brain is as old as modern human civilization. More than two thousand years ago, Aristotle believed that the brain is an organ of only minor importance, a compound of earth and water, "...perhaps necessary to cool the blood" (De partibus animalum, ca. 350 BC). Several hundred years later, Greek anatomist Galen (ca. 129 to 200 AD) attributed cognitive processing to the brain's ventricular system, a scientific view that remained influential until well into the Renaissance. Ever since then, our knowledge has greatly advanced and the brain's significance has been reassessed, yet the full functionality of this complex organ is still only partly understood.

Today, *in vivo* brain imaging techniques facilitate the investigation of the living and fully functioning brain. For example, *magnetic resonance imaging* (MRI) provides detailed information about the brain's anatomy while *functional magnetic resonance imaging* (fMRI) allows us to watch the brain at work. Such techniques provide invaluable research and diagnostic tools for cognitive, medical, and biological sciences.

However, the processing and analysis of neuroimaging data also poses huge challenges to mathematics and computer science. Functional imaging experiments often result in several gigabytes of data that require fast and efficient processing. In addition, the high noise level typically observed in neuroimaging data calls for sophisticated mathematical models and statistical analysis methods. Their development and implementation constitutes an important part of modern neuroscience.

In this work we present a series of new methods for the analysis of fMRI data. They comprise three levels of analysis, reaching from the analysis of imaging data obtained for individual test subjects or patients to the conjoint assessment of data representing several independent imaging experiments. The methods are presented in the three main

chapters of this work, each chapter covering a different analysis level. Throughout this work, we use the term *first-level analysis* to refer to the analysis of functional imaging data from individual subjects (Chapter 3). The term *second-level analysis* refers to the analysis of data obtained from groups of subjects performing the same functional imaging experiment (Chapter 4). With *meta-analysis* we refer to the evaluation of data obtained in several independent imaging experiments (Chapter 5).

The majority of the presented methods is *data-driven*, that is, they investigate the inherent structure of the imaging data in an exploratory fashion. This facilitates the search for information in the data that is often not captured by model-based approaches. The methods can be used to address research questions, such as the search for functional networks and the investigation of temporal aspects in cognitive processing, that go beyond the mere detection of brain activation related to a specific cognitive task. One exception is the method presented in Chapter 4 where Bayesian statistics is applied to parameters derived from modelling fMRI measurements. This model-based technique is designed to detect brain activation with particularly high sensitivity.

Throughout this work, special emphasis is placed on the high variability of fMRI data across brain regions, subjects, and imaging experiments. With the application of our new analysis techniques to a number of different imaging data sets we demonstrate how some of the problems arising from this condition can be alleviated.

Outline of the work

Magnetic resonance imaging exploits the properties of hydrogen nuclei when placed in a high magnetic field. As these properties are tissue dependent, structural MRI enables us to distinguish between different brain tissue such as grey and white matter, and hence between different anatomical structures in the brain. Functional MRI in contrast exploits changes in the amount of oxygen available in the blood at any particular time and location in the brain. While a test subject is presented with a stimulus or performing a particular task in the MRI scanner, changes in blood oxygenation are recorded, which are assumed to be linked to neuronal activation. Results of these measurements can then be mapped onto structural brain images, showing where in the brain the activation has occurred during task performance.

The signal obtained from fMRI measurements is referred to as *blood oxygenation level dependent* (BOLD) signal or response. It can be measured for brain volumes as small as a few cubic millimeters. These volumes are called *voxels*. A single func-

tional image or measurement of the BOLD signal, covering the entire or substantial parts of the brain, can be obtained within a few hundred milliseconds. During an fMRI experiment, several hundred or thousand such measurements of the BOLD signal are obtained. For an individual voxel, these consecutive measurements can be viewed as a time series that captures the changes in blood oxygenation over the entire time of an experiment. These time series, typically referred to as *fMRI time series* or *fMRI time course*, can in turn be related to the experimental task performed during the measurements.

While we can observe the BOLD signal with very high spatial accuracy, it is only an indirect marker of neuronal activity and the exact link between increased neuronal activity and changes in the level of blood oxygenation is still not fully understood. Moreover, BOLD signal changes are very small and hence require long measurements followed by sophisticated mathematical modelling and statistical analysis. In Chapter 2 of this work we present the basic physiological and physical principles of structural and functional MRI measurements, together with the standard approach to the analysis of fMRI time series based on the application of the *general linear model* (GLM) and the subsequent construction of *statistical parametric maps* (SPM). We further introduce a number of data sets that are used as test cases for our newly developed analysis methods. They comprise data from individual subjects, measurements from groups of subjects, and results from several imaging experiments investigating comparable experimental paradigms.

Since the advent of functional neuroimaging, the number of experimental studies published each year has grown exponentially, with a total of approximately 10,900 fMRI studies published by September 2010 in English language journals alone. The majority of these experiments is aimed at investigating the parcellation of the brain into regions with different functionality. This is most commonly done by designing an imaging experiment such that, across the entire brain, BOLD signals elicited by different stimuli or cognitive tasks can be measured and compared. GLMs and SPMs are used for modelling and statistical evaluation of the data. If a brain region shows a statistically significant difference between BOLD signals elicited by two different stimuli or tasks, this region is assumed to exhibit a functionality that is necessary to process one of the stimuli or perform one of the tasks but not, or only to a much lesser degree, the other. This model-based approach thus facilitates the functional dissociation of brain regions by directly manipulating the underlying cognitive processes.

In recent years, exploratory analysis techniques such as principal component analysis, independent component analysis, clustering, and pattern recognition have provided a viable alternative to this model-based approach. Such techniques explore the inherent structure of the data independently of the experimental design or any pre-defined model of the BOLD response. They are based on the assumption that the specific functionality of brain regions and their interplay are reflected in the pattern of cortical activation. For example, clustering can be employed for the parcellation of the brain into distinct functional regions, if we assume that fMRI time series of voxels within the same functional brain region are similar to each other, but different from time series of voxels in other functional regions. The most commonly used clustering methods like hard or fuzzy k-means or hierarchical clustering perform the grouping of voxels directly on the observed fMRI time series. Other techniques such as cross-correlation or spectral clustering operate on some form of similarity matrix, typically derived from the correlation or covariance of fMRI time series.

In Chapter 3 of this work we present a new exploratory method for the parcellation of brain areas into distinct functional regions based on the similarity of fMRI time series. The method employs replicator dynamics and canonical correlation, two mathematical concepts developed in the fields of theoretical biology and multivariate data analysis, respectively. Our method operates on the first analysis level, i.e. it facilitates the search for functional regions in brain data from individual subjects. Like cross-correlation or spectral clustering, it exploits the properties of a similarity matrix for fMRI time series. However, unlike usual clustering techniques the method selects the most informative voxels for the clustering process. The application of our method to two different test data sets shows that it is very robust against the variability of fMRI time series, i.e. against differences in multiple measurements of the same subject as well as against the variability across several subjects.

The second analysis technique presented in Chapter 3 is aimed at investigating temporal aspects of the BOLD signal. Exploring this issue is of great relevance, given the high interest of cognitive neuroscientists in characterizing the precise dynamics of the cognitive processes underlying the BOLD response. Unfortunately, the temporal resolution of the BOLD signal is rather low with a typical sampling rate of one or two seconds. Moreover, changes in the blood oxygenation level appear only several seconds after the underlying neuronal activity. The dynamics of cognitive processes are thus only poorly captured by the BOLD signal.

Our new method facilitates the estimation of specific time points of interest along fMRI time series. They include, for example, the time point of the observed signal maximum and the onset of the signal increase after stimulation. Despite the poor temporal resolution of the BOLD signal, some of these points can be estimated with high accuracy. This enables us to compare the temporal dynamics of brain functions across different experimental stimuli or tasks. Moreover, our method aids the identification and interpretation of the sources for the high variability of the BOLD dynamics observed across brain regions, multiple measurements, and subjects.

In Chapter 4 of this work we focus on the second-level analysis of fMRI data, i.e. the analysis of groups of subjects. This form of analysis is of particular importance to the statistical evaluation of fMRI data. A large body of work has shown that results from individual imaging sessions can vary considerably from subject to subject and even between different measurements of the same subject. Results from individual imaging sessions thus only represent a sample of one subject's brain activation. Consequently, meaningful generalizations of imaging results require the conjoint analysis of fMRI data obtained from groups of subjects. This necessitates statistical inference methods on the second analysis level that reflect the commonalities in the neuronal activity of different subjects while, at the same time, being robust against differences between subjects caused by their neuroanatomical or physiological variability.

In recent years, Bayesian statistics has been suggested as an alternative to the statistical hypothesis test most commonly employed in functional imaging. With Bayesian techniques we can alleviate some of the problems inherent in this approach, such as the need to correct for multiple statistical testing or the inability to assess any alternative to the pre-formulated null hypothesis about the effect of interest. However, Bayesian techniques are often computationally costly and mathematically complex, in particular when applied on both the first and the second analysis level. In Chapter 4 of this work, we propose a new method for the analysis of functional MRI data based on Bayesian statistics. In contrast to previous Bayesian approaches, our method does not require a computationally costly Bayesian model on the first level. Rather, modeling for single subjects can still be realized by means of the computationally efficient general linear model, as introduced in Chapter 2. Model parameters obtained from the GLM are then further processed on the second level within a Bayesian framework. The comparison of our method with conventionally used statistical parametric mapping shows that it

is more robust against outliers, mainly on the grounds of a better account for both the within-subject and the between-subject variability of the fMRI measurements.

Despite the use of increasingly sophisticated analysis techniques on the first and second analysis level, the wealth of imaging data obtained so far still conveys a rather variable picture. A conclusive interpretation of experimental results is often limited by their inconsistency with results from related experimental investigations. Such inconsistencies demand consolidation, which calls for image analysis techniques on the meta-analysis level. Moreover, functional neuroimaging is currently advancing from the simple detection and localization of cortical activation to the investigation of complex cognitive processes and associated functional relationships between brain regions. Such research questions can no longer be addressed by the isolated analysis of single experiments alone, but necessitate the integration of results across different cognitive tasks and experimental paradigms. This again makes meta-analyses an increasingly important part in the evaluation of functional imaging experiments. To date, however, only very few computational approaches to the meta-analysis of functional imaging data exist.

In Chapter 5 of this work we present two new methods for the meta-analysis of fMRI data. Our first method is aimed again at the detection of distinct functional regions in the brain. This time, however, the search for functional regions is based on activation coordinates obtained across a number of independently performed imaging experiments. Such activation coordinates reflect the spatial localization of maximum brain activation when a particular stimulus is processed or a specific cognitive task is performed in the course of an experiment. The use of standardized brain coordinate systems facilitates the comparability of these coordinates across experiments.

Our new method employs Gaussian mixture modelling and Bayesian model selection for clustering activation coordinates. The method can once more be viewed as an exploratory analysis technique, as clustering is based solely on the spatial distribution of the coordinates and does not take into account any information arising from the experimental setup or data analysis on previous levels. The method is suited, in particular, for data sets where other meta-analysis techniques yield brain areas that are too large to be meaningfully interpreted as distinct functional regions.

Our second new meta-analysis method aids the search for functional brain regions that frequently show a joint activation across imaging experiments. Such regions can be thought of as belonging to a functional network, i.e. a group of brain regions which

is, as a whole, involved in processing a certain stimulus or task. In the search for such functional brain networks we again employ the concept of replicator dynamics, as introduced in Chapter 3. However, on the meta-analysis level replicator dynamics is applied to a co-activation rather than a similarity matrix.

The use of activation coordinates rather than fMRI time series in both meta-analysis techniques aids a fast and efficient processing of the data. Moreover, it facilitates the direct use of data sets that are provided in *neuroimaging databases*. Such databases contain activation coordinates together with detailed information about experimental procedures and imaging modalities. Combining intelligent search mechanisms for neuroimaging databases with our newly developed meta-analysis techniques promotes fast and efficient meta-analyses that can comprise several hundred imaging experiments.

In Chapter 6 we point to a number of recent applications of the presented methods and conclude our work with a discussion of the remaining problems and limitations of fMRI data analysis. While our methods address some of these problems, we also discuss possible directions of their further improvement as well as more general research questions for the development of neuroimaging data analysis techniques in the future.

Chapter 2

The fMRI Signal

2.1 Physical and physiological background

In this section we provide a brief introduction to the physical and physiological underpinnings of magnetic resonance imaging and functional magnetic resonance imaging. This introduction is by no means intended to be comprehensive, but should provide the reader with the basic concepts necessary to understand the origin of the data presented in this work and the measurement methods used to obtain them. For a detailed exposition of the principles underlying MRI and fMRI see, for example, the books by Buxton (2002) and Jezzard et al. (2002).

The physical basis for both MRI and fMRI is provided by the nuclear magnetic resonance phenomenon. Protons and neutrons possess an angular momentum property which can be viewed as arising from a rotation or spin. In atomic nuclei with an uneven number of protons or neutrons, this property results in a magnetic moment. Hydrogen nuclei, only consisting of a single proton, are of particular interest, as they possess a relatively large magnetic moment and constitute large parts of biological tissue. Placing a hydrogen nucleus into an externally applied magnetic field B_0 causes its magnetic moment to align or counter-align with the direction of the magnetic field. Looking at a large ensemble of hydrogen nuclei, the net magnetic vector, which is the sum of all individual moments, is pointing in the direction of B_0, since slightly more nuclei do align rather than counter-align with the magnetic field direction. Moreover, applying the external magnetic field forces the magnetic moment to precess around the field direction of B_0. The frequency of this precession is referred to as Lamor frequency ω_0 and is proportional to the strength of the magnetic field.

When an oscillating radio frequency (RF) pulse with frequency ω_0 is applied, some of the nuclei are transfered from an aligned to a counter-aligned state, causing the net magnetic vector to be tilted away towards the plane perpendicular to the field direction. In a receiver coil, a current is induced by the precession of the spins in this transversal plane. Over time, the transfered nuclei assume their original state and the net magnetic vector regains its original strength and direction. This process is referred to as T_1-relaxation. The rate with which the net magnetic vector is tilted back towards its original direction is measured by the time constant T_1.

The application of the RF pulse has a second effect on the nuclei in the magnetic field. The phase of their precession is synchronized, leading to a phase coherence of all nuclei. Due to random interactions between the nuclei, this phase coherence decays again over time at a constant rate T_2. This process is referred to as T_2-relaxation. Additionally, the phase coherence between nuclei also decays due to slight magnetic field inhomogeneities which cause nuclei to precess with slightly different frequencies. The combined effect of random interactions between nuclei and magnetic field inhomogeneities on the decay of the phase coherence is measured by the time constant T_2^*, whereby $T_2 > T_2^*$.

The spatial resolution of MR measurements depends on the strength of B_0, typically between 1.5 and 7 Tesla, and the specific imaging sequence used. Voxel sizes for fMRI measurements are usually in the range of $2 \times 2 \times 2 \, \text{mm}^3$ and $3 \times 3 \times 3 \, \text{mm}^3$. At high field strengths, an even higher resolution is possible. For anatomical MRI measurements voxel sizes below $1 \, \text{mm}^3$ are achievable. The spatially selective measurement for the MR signal is realized by means of magnetic field gradients which again cause differences in the precession frequency of the nuclei. Using these gradients, nuclei in a single slice of the brain and, within a slice, in a single row or column can be spatially encoded.

As the time constants T_1 and T_2 are tissue dependent, exploiting their differences facilitates the distinction between different types of tissue and hence the generation of structural images that reflect the anatomical organization of the brain. Two examples of anatomical images resulting from T_1- and T_2-weighted sequences are shown in Figure 2.1. In the context of functional imaging, the acquisition of high quality anatomical brain images is necessary for a precise localization of cortical and subcortical activation detected by the fMRI measurements.
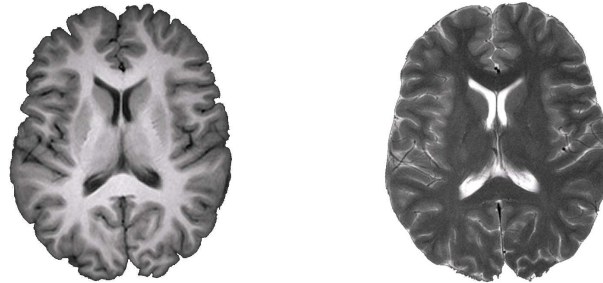
Figure 2.1: Structural magnetic resonance images resulting from a T_1-weighted (left) and T_2-weighted (right) imaging sequences.

Functional MRI relies on imaging sequences that are sensitive to differences in T_2^* which arise from different levels of oxygen concentration in the blood (Ogawa et al., 1990, 1993). In the blood, oxygen is attached to hemoglobin molecules. Oxygenated hemoglobin molecules are slightly diamagnetic. In contrast, without oxygen attached to the hemoglobin, the molecules become paramagnetic, causing slight local inhomogeneities in the magnetic field B_0. These in turn cause hydrogen nuclei in the blood to quickly loose any phase coherence that was assumed after an RF pulse. In contrast, phase coherence decays slower, if the hydrogen nuclei are surrounded by oxygenated hemoglobin molecules causing smaller local field inhomogeneities. Consequently, T_2^* is shorter in areas with low oxygen concentration in the blood and longer in areas with high oxygen concentration. Imaging sequences sensitive to these differences in T_2^* can thus measure changes in the oxygenation level of the blood over time. These changes are supposed to reflect changes in the level of neuronal activity. The resulting signal is referred to as blood oxygenation level dependent signal or BOLD response.

The precise relationship between the BOLD signal and the underlying neuronal activity is still not fully understood (Buxton, 2010). In the current understanding, the BOLD signal reflects the hemodynamic response[1] to neuronal activity which is a complex interplay of cerebral blood flow (CBF), cerebral blood volume (CBV) and local oxygen uptake (cerebral metabolic rate for oxygen, $CMRO_2$) in the vicinity of active neurons. More specifically, neuronal activity following stimulation is accompanied by an increase in $CMRO_2$, a large increase in CBF and a small increase in CBV. CBF in-

---

[1] In the literature, the terms 'BOLD signal' or 'BOLD response' and 'hemodynamic response' are often used interchangeably.

crease is unproportionally high compared to the oxygen uptake, causing an increase of the blood oxygenation level and corresponding changes in T_2^*-sensitive signals. The characteristic shape of the hemodynamic response results from the dynamics of these processes which are illustrated schematically in Figure 2.2 together with the corresponding changes in T_2^*.
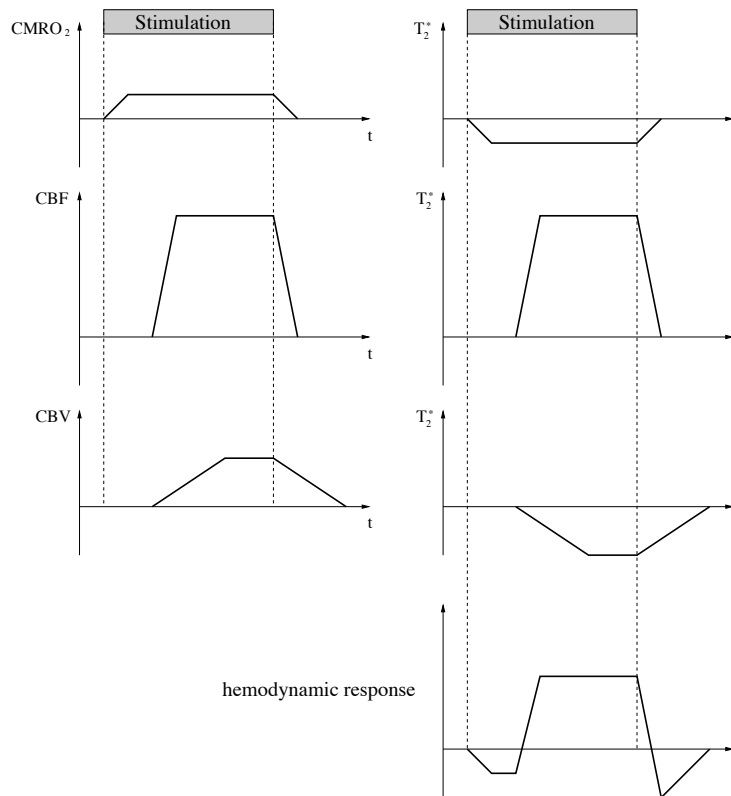
Figure 2.2: Relationship between the physiological effects of neuronal activity and T_2^*. Neuronal activity is followed by an increase in CMRO$_2$ and CBV, causing the T_2^*-sensitive signal to decrease, while an increase in CBF yields an increase in the T_2^*-sensitive signal. The general shape of the resulting hemodynamic response is depicted schematically at the bottom.

The BOLD signal provides only an indirect marker for neuronal activity, but it can be measured by fMRI in a *non-invasive* manner. This is a big advantage compared to invasive imaging techniques like Positron Emission Tomography (PET). Moreover, functional MRI sequences allow changes in the BOLD signal to be detected with very

high local precision compared to imaging techniques such as Electroencephalography (EEG) and Magnetoencephalography (MEG). However, given the slow reaction of the hemodynamic system to neuronal activity, temporal information about neuronal activity is hard to obtain from the BOLD signal. Moreover, the low signal-to-noise ratio (SNR) of the BOLD signal requires repeated measurements of comparable events together with a sophisticated statistical analysis of the signal in order to reliably detect neural correlates of the investigated cognitive processes.

2.2 Modelling and statistical analysis of fMRI data

A functional MRI experiment results in a series of several hundred digital images which describe the blood oxygenation level across the brain. The main objective of the subsequent data analysis is to detect changes in the BOLD signal over time and identify brain regions where these changes can be related systematically to the experimental paradigm. This requires a series of processing steps which can broadly be divided into *preprocessing*, *modelling*, and *statistical evaluation*. While preprocessing and data modelling are usually performed for every subject individually, statistical evaluation can be performed for individual subjects (first-level analysis) and groups of subjects (second-level analysis). In the following, the typical processing steps of an fMRI data analysis are described in detail. We hereby focus on the processing steps that were performed for our test data sets, which are described at the end of this chapter.

2.2.1 Preprocessing

Prior to data modelling and statistical analysis, a number of procedures are performed on the obtained fMRI data in order to remove artifacts, improve the SNR, and make measurements comparable across time and space. This preprocessing chain typically consists of the following steps:

1. **Slicetime correction:** correction for temporal offsets between slices acquired in one scan

2. **Motion correction:** geometrical alignment of measurements within one experiment

3. **Spatial and temporal filtering:** removal of signal drifts in the course of the experiment, improvement of SNR, and reduction of inter-individual variability

An individual measurement of the entire brain, typically referred to as *scan*, consists of a number of slices. These slices are not acquired simultaneously but consecutively within an interval of several ten or hundred milliseconds. This temporal offset between slices has to be corrected for, in order to facilitate a comparable statistical analysis across all voxels in the brain. Slice time correction is performed by interpolating the value of each time course at time point t_0, the time point where the first slice was acquired. This is most commonly realized by means of linear, sinc, or b-spline interpolation.

Motion correction becomes necessary, if the subject moves during image acquisition. Even small movements such as the slight head movement caused by swallowing can result in a misalignment of images acquired at different time points during the experiment. If not accounted for, this misalignment can result in substantial errors in the subsequent modelling and statistical analysis steps (Oakes et al., 2005). There are several approaches to motion correction which differ in cost functions, optimization methods, and interpolation techniques. Most methods are based on rigid-body co-registration, assuming that the shape of the head is constant between scans. Motion correction can be performed in 2D space, i.e. for each slice of the brain independently, or in 3D space for all slices in parallel. The latter is mathematically more complex with three instead of one rotational parameter, but it can also account for movements across neighboring slices which are neglected by 2D approaches. Moreover, motion correction can precede or follow slicetime correction, or both steps can be performed simultaneously.

Spatial and temporal filtering of fMRI measurements serve different purposes. Temporal filtering is primarily aimed at eliminating slow signal drifts, i.e., different overall signal intensities at the beginning and the end of an experimental sessions. Such drifts can have both technical and physiological reasons, but are generally assumed to comprise low frequency components in the acquired time series. In order to eliminate these drifts, a high pass filter is applied to every BOLD time series. The cutoff frequency of the filter has to be chosen according to the experimental design in order to ensure that low frequency drifts are eliminated while frequency components relating to the experimental stimulus remain part of the corrected signal.

Spatial filtering is used to improve the SNR of the data and to reduce the inter-subject variability of the measurements. Although the reduction of random noise in the data generally improves the ability to detect brain activation in the subsequent processing steps, spatial smoothing also reduces the resolution of the data. While this makes brain images more comparable across subjects, it can also hamper the precise anatomical localization of brain activation. A Gaussian kernel with between 4 and 12 mm full width at half maximum (FWHM)[2] has been established as a workable spatial filter. This provides a good balance between improving SNR and maintaining the excellent spatial resolution of functional MR images.

In order to determine the exact anatomical location of detected brain activity and to perform statistical analyses of groups of subjects, the acquired functional data sets have to be geometrically aligned with anatomical data and placed into a standardized coordinate system. This requires two further preprocessing steps, implementing spatial transformations prior to modelling and statistical analysis:

1. **Registration:** alignment of the functional data with a high-resolution anatomical reference data set

2. **Normalization:** linear or non-linear scaling of the data to standard size and placement into a standardized coordinate system

Registration is performed by shifting and rotating the acquired functional data into correspondence with a high-resolution anatomical data set. This data set is usually acquired prior to functional scanning. Three rotational and three translational parameters required for the alignment between reference and functional images are determined by optimizing the match between the anatomical data and the functional data acquired at some pre-defined time step. Voxel-wise correlation or mutual information are typically used as matching metric. Note that the application of the registration parameters to the entire functional data set requires resampling, which is typically realized by trilinear interpolation. Note further that registration can also be performed between anatomical reference data sets and statistical parametric maps that result from subsequent statistical analyses.

---

[2]Within the neuroscience community, both σ and FWHM are used to refer to the width of a Gaussian. Throughout this work, we will use both notions.

For comparability across subjects and experiments, anatomical and functional data sets are placed within the Talairach coordinate system (Talairach and Tournoux, 1988). This standardized stereotactic coordinate system is defined by the position of the anterior (AC) and posterior commissure (PC), two well-defined anatomical structures that can be detected automatically or have to be detected by hand within each anatomical reference data set. Coordinate (0,0,0) is located at the position of AC. Starting from this point, the brain's right hemisphere has positive x-coordinates, the left hemisphere has negative x-coordinates. The anterior part of the brain has positive y-coordinates, the posterior part of the brain has negative y-coordinates. The superior part of the brain has positive and the inferior part has negative z-coordinates.

Finally data sets need to be scaled to match in size. In this second normalization step the image can be scaled linearly to standard size, which in the Talairach atlas (Talairach and Tournoux, 1988) is assumed to be $135 \times 175 \times 120\,\mathrm{mm}^3$. Other available brain templates such as the standard brains from the Montreal Neurological Institute (MNI) are slightly larger. Note that nonlinear deformation can be also be applied either alternatively or in addition to linear scaling.

The specific algorithms and parameters used for preprocessing the test data sets in this work are listed in Section 2.3 and described in more detail by Lohmann et al. (2001).

2.2.2 Modelling

After preprocessing, an fMRI data set for a single subject contains for every voxel in the brain a time series of several hundred time steps. These need to be further analyzed in search for changes in the blood oxygenation level that can be linked to the presented experimental stimuli[3]. This analysis is most commonly done by means of a general linear model (Friston et al., 1994; Worsley and Friston, 1995; Zarahn et al., 1997). In a GLM, it is assumed that the observed data can be expressed as a linear combination of explanatory variables and an error term. Specifically, for every time step $i, i = 1, \ldots, n$, in the measured BOLD signal of an individual voxel

$$y_i = \beta_0 + x_{i1} \cdot \beta_1 + x_{i2} \cdot \beta_2 + \ldots + x_{ip} \cdot \beta_p + \varepsilon_i \qquad (2.1)$$

---

[3]The presentation of an individual stimulus within an experiment, usually covering several time steps, is referred to as 'trial'. The collection of trials where the same stimulus is presented is commonly called 'condition'. Often, the terms 'stimulus' and 'condition' are used interchangeably. Throughout this work we will follow this convention.
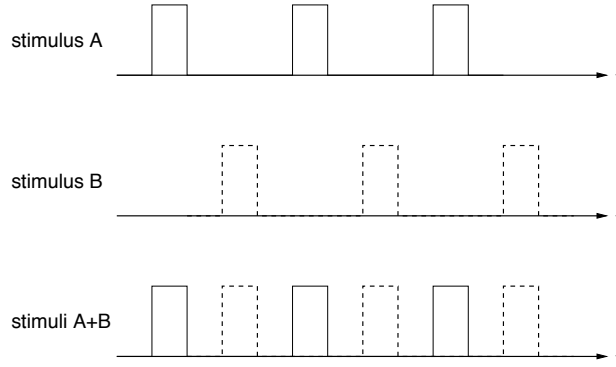
Figure 2.3: The hemodynamic response to two experimental stimuli A and B modelled simply by a square wave function (top and middle) and the sum of the modelled response functions (bottom).

where y_i is the value of the BOLD signal at time point i and x_{ik} represents the value of the modelled hemodynamic response function for stimulus k at time point i. In the most simple case, the hemodynamic response is modelled as a square wave function as shown in Figure 2.3. That is, x_{ik} equals one, if experimental condition k was presented at time step i and zero otherwise. In this case, the parameter β_k represents the signal change caused by condition k. β_0 denotes the mean of the entire time series, and ε_i is the measurement error at time step i.

The GLM can be written more conveniently in matrix notation for the entire time series as

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon} \qquad (2.2)$$

with data vector $\mathbf{y}$, the so-called design matrix $\mathbf{X}$, parameter vector $\boldsymbol{\beta}$, and error vector $\boldsymbol{\varepsilon}$. The error terms ε_i are assumed to be independent and identically normally distributed with $\boldsymbol{\varepsilon} \sim N(0, \sigma^2 \mathbf{I})$, and unbiased.

With $\mathbf{y}$ given and $\mathbf{X}$ determined by the experimental design, the least-squares estimator $\hat{\boldsymbol{\beta}}$ that minimizes $\sum_i \varepsilon_i^2$ with respect to $\boldsymbol{\beta}$ is obtained as

$$\hat{\boldsymbol{\beta}} = \mathbf{X}^+ \mathbf{y} \,, \qquad (2.3)$$

where $\mathbf{X}^+$ is the Moore-Penrose inverse of $\mathbf{X}$. Note that $\hat{\boldsymbol{\beta}}$ is the best linear unbiased estimate (BLUE) of $\boldsymbol{\beta}$ (Seber, 1977; Worsley, 2001). It can further be shown that the

sampling distribution of the least-squares estimates $\hat{\beta}_i$ is normal with $E(\hat{\beta}_i) = \beta_i$ and $\text{Var}(\hat{\beta}_i) = \sigma^2 c_{ii}$, where c_{ii} are the diagonal elements of $(\mathbf{X^T X})^{-1}$ (Seber, 1977). This normality of the sampling distribution is an important prerequisite for the Bayesian second-level analysis described in Chapter 4 of this work.

An extended GLM for serially autocorrelated observation data is

$$\mathbf{Ky} = \mathbf{G}\boldsymbol{\beta} + \mathbf{K}\boldsymbol{\varepsilon}, \tag{2.4}$$

where $\mathbf{K}$ is a convolution matrix using a Gaussian kernel[4] and $\mathbf{G} = \mathbf{KX}$ is the convolved design matrix. Here the variance of $\hat{\boldsymbol{\beta}}$ extends to

$$\text{Var}(\hat{\boldsymbol{\beta}}) = \hat{\sigma}^2 \mathbf{G^+ V (G^+)^T}, \tag{2.5}$$

where $\mathbf{V} = \mathbf{KK^+}$, $\mathbf{G^+}$ is again the Moore-Penrose inverse of $\mathbf{G}$, and $\hat{\sigma}^2$ is an unbiased estimator for the variance σ^2.

After fitting the linear model to the observed data, an effect of interest can be expressed by means of a so-called *contrast* $\mathbf{c}\hat{\boldsymbol{\beta}}$ which is a linear combination of the parameter estimates and, in our context, describes the difference between experimental conditions. The row vector $\mathbf{c}$ is a set of weights that usually sum to zero. For example, given parameter estimates $\hat{\beta}_0, \hat{\beta}_1, \hat{\beta}_2$, the contrast vector $\mathbf{c} = (0 \quad 1 \quad -1)$ can be used to express the activation difference between condition 1 and condition 2 corresponding to $\hat{\beta}_1 - \hat{\beta}_2$. The estimated variance of a contrast is

$$\begin{aligned} \text{Var}(\mathbf{c}\hat{\boldsymbol{\beta}}) &= \mathbf{c}\,\text{Var}(\hat{\boldsymbol{\beta}})\,\mathbf{c^T} \\ &= \hat{\sigma}^2 \mathbf{c G^+ V (c G^+)^T}. \end{aligned} \tag{2.6}$$

It is important to note that the calculation of contrasts up to this point is performed for every voxel independently but using identical model functions and parameters. After modelling, contrast values corresponding to the same contrast for all voxels are typically placed within a so-called *contrast image*. Note further that the calculation of contrasts does not involve any statistical evaluation of the data. A contrast image merely contains information about how one or more explanatory variables correspond to the experimental design.

---

[4]The idea of accounting for serial autocorrelation by temporal smoothing is usually referred to as 'pre-coloring'. An alternative, but more computationally costly approach is 'pre-whitening' which requires estimating the autocorrelation directly from the data, see e.g. Worsley et al. (2002).

2.2.3 Statistical analysis

In order to assess the significance of an observed contrast, contrast values can be subjected to further statistical analysis, both on the first and on higher analysis levels. This is most commonly realized by means of statistical *null hypothesis significance tests* (NHST) based on frequentist t- and F-statistics (Friston, 1994; Worsley and Friston, 1995; Ardekani and Kanno, 1998). For example, for an individual subject a statistical parametric map SPM$\{t\}$ can be constructed that contains for every voxel a significance value related to the contrast of interest. Specifically, an SPM$\{t\}$ is obtained by conducting a one-sample t-test that assesses the null hypothesis of zero activation for the contrast of interest. Given a contrast $\mathbf{c}\hat{\boldsymbol{\beta}}$,

$$t = \frac{\mathbf{c}\hat{\boldsymbol{\beta}}}{\sqrt{\mathrm{Var}(\mathbf{c}\hat{\boldsymbol{\beta}})}} \tag{2.7}$$

has an approximate Student's t-distribution with ν degrees of freedom (Worsley, 2001) from which the corresponding p-value can be derived. The null hypothesis is rejected for voxels whose p-value exceeds a commonly accepted threshold, typically $p = 0.001$ or $p = 0.05$ (corrected for multiple tests, see below). Note that ν, which for uncorrelated data equals the number of scans minus the number of explanatory variables, needs to be further reduced to account for the autocorrelation of the observed measurements and any applied temporal smoothing (Lohmann et al., 2001). Obtained t-values are typically transformed into z-values, giving a SPM$\{z\}$ which is then thresholded at a level corresponding to the desired p-value, e.g. at $z = 3.09$ corresponding to $p = 0.001$.

The construction of SPMs typically entails the application of several thousand statistical tests. For example, if our measurements contain 1,000 voxels, the same statistical test is repeated 1,000 times. If we then apply a significance threshold of 0.05, the results are expected to contain about 50 false positive voxels, i.e. voxels where brain activation was wrongly detected. Thus, any region containing less than 50 voxels could theoretically consist of only false positive voxels and might be wrongly interpreted as activated region. This problem needs to be accounted for by correcting the obtained results for multiple comparisons. The most common correction methods are based on Gaussian Random Fields, the False Discovery Rate, or Monte Carlo simulations (Worsley et al., 2004; Genovese et al., 2002; Lohmann et al., 2008, for example).

A second-level statistical analysis is aimed at combining results from a group of subjects performing the same experiment. The majority of early fMRI studies employed a so-called *fixed-effect* analysis, treating the group data as if coming from a single subject. This approach is mathematically simple but neglects the variability of the observed contrast across subjects. A more realistic approach is provided by *random-effects* analyses based on a summary statistic (Holmes and Friston, 1998), where again a one-sample t-test is employed, now assessing the null hypothesis of zero activation for the contrast of interest across subjects. Note that in contrast to a full random-effects analysis, which requires a mathematically and computationally more demanding hierarchical model of the data, the summary statistic does not fully account for the within-subject variability of the obtained parameter estimates. This drawback is addressed by our Bayesian second-level analysis method proposed in Chapter 4.

2.2.4 Hemodynamic modelling

In Section 2.2.2 we have assumed a very simple model for the hemodynamic response to an experimental stimulus in form of a square wave function as depicted in Figure 2.3. A more realistic model of the hemodynamic response can be achieved using a Gamma function[5] as shown in Figure 2.4 (left). Functions of this general form are believed to capture the delay of the hemodynamic process in response to a short stimulus reasonably well (Friston et al., 1994; Boynton et al., 1996; Lange and Zeger, 1997). Friston et al. (1998) proposed a sum of two Gamma functions

$$h(t) = \left(\frac{t}{d_1}\right)^{a_1} e^{-\frac{t-d_1}{b_1}} - c \left(\frac{t}{d_2}\right)^{a_2} e^{-\frac{t-d_2}{b_2}} , \qquad (2.8)$$

where $a_1 = 6$, $a_2 = 12$, $b_i = 0.9$, $c = 0.35$, and $i = \{1,2\}$ was suggested a suitable set of parameters (Glover, 1999; Worsley, 2001). Here, $d_i = a_i b_i$ reflects the time of maximum activation, often referred to as *time-to-peak*. This is an important parameter used to determine temporal differences in the hemodynamic response across conditions or different regions of the brain. As seen in Figure 2.4 (right), this model function also accounts for the 'post-stimulus undershoot', a signal decrease below baseline that is often observed after the end of stimulus presentation.

---

[5]This is not to be confused with Γ. The name 'Gamma function' for the functions presented in Figure 2.4 and Equation 2.8 was coined in the late nineties, probably due to a confusion with earlier model functions that contained Γ. However, as it is frequently used within the neuroscience community, we will follow this convention throughout this work.
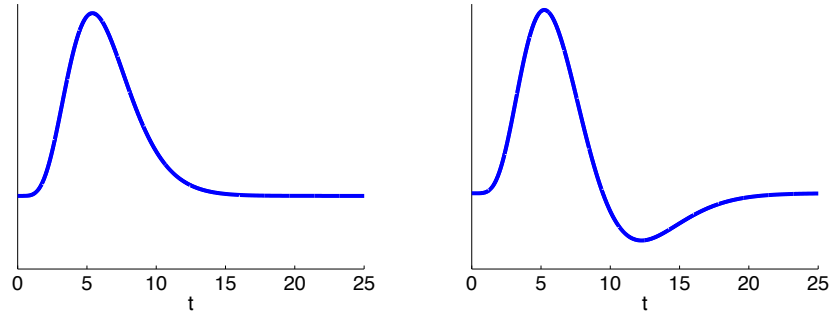
Figure 2.4: The hemodynamic response modelled as a single Gamma function (left) and the sum of two Gamma functions (right).

The entire BOLD response to an experimental stimulus can be modelled as the convolution of the hemodynamic response function with a square wave function $s(t)$, now describing the presence or absence of the stimulus at time point t:

$$x(t) = \int_0^\infty h(u)\, s(t-u)\, du\,. \tag{2.9}$$

Sampling this function at the n image acquisition time points provides an n-dimensional vector which replaces in the design matrix $\mathbf{X}$ the earlier described column of zeros and ones for this experimental stimulus. Note that Gamma functions are particularly suitable models of the hemodynamic response, if the corresponding stimulus is presented for a short time, typically in the range between a few hundred milliseconds and a couple of seconds. Experimental designs consisting of such stimuli are referred to as *event-related designs*, since every stimulus presentation can be viewed as a short event within the time course of the experiment. In contrast, experimental designs containing the presentation of stimuli over several seconds are called *epoch-related* or *block designs*. The hemodynamic response to such stimuli is sometimes modelled differently, for example, by means of Gaussians rather than Gamma functions.

For yet greater flexibility, the model of the hemodynamic response can be extended to a set of so-called *basis functions*, for example, a Gamma function and its temporal derivatives in an event-related design (Friston et al., 1998). As shown in Figure 2.5, adding or subtracting the first temporal derivative to or from a single Gamma function shifts the peak of the function along the time axis. This allows for a more flexible temporal modelling of the hemodynamic response.

Note that this extended hemodynamic model results in two explanatory variables and, consequently, in two parameter estimates $\hat{\beta}$ for every experimental condition. While the first parameter primarily reflects activation changes in relation to the experimental stimulus, the second one reflects the temporal delay with respect to time-to-peak as defined by the Gamma function. Contrasting this second parameter for two experimental conditions can be utilized in tests for relative temporal differences between conditions. Note that the actual amount of the temporal delay between conditions can not be determined from these parameter estimates. However, in Chapter 3 we present a method for determining and comparing characteristic time points of the hemodynamic response, such as time-to-peak, from trial-averaged fMRI time series.
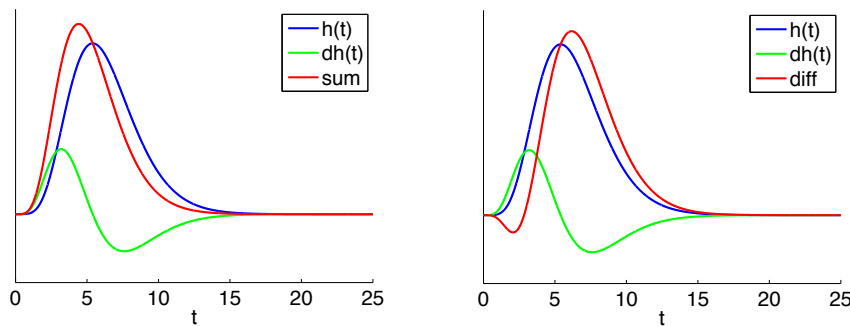
Figure 2.5: Sum (left) and difference (right) of a single Gamma function and its first temporal derivative.

Including the second temporal derivative of a Gamma function allows for a flexible model of the hemodynamic response's dispersion. It is important to note, however, that including more basis functions reduces the degrees of freedom and thus the statistical power of the subsequent analysis. This in turn restricts the number of different conditions or stimuli that can be reliably modelled and tested within an experimental setup. Such mathematical constraints have to be taken into account when designing a functional imaging experiment in addition to considerations related to the investigated neuropsychological phenomena.

2.3   Experimental data sets

The new analysis techniques presented in the following chapters can be applied to different data types on different levels of analysis, reaching from single-subject fMRI time courses to activation coordinates from several imaging experiments. Testing our techniques thus requires data sets comprising all of these levels. In the following sections, five data sets are described that were used as test cases for our methods. The first three data sets are comprised of fMRI time series obtained from individual subjects. These data sets were used for testing the first-level and second-level analysis techniques presented in Chapter 3 and Chapter 4, respectively. Prior to or in conjunction with the application of the newly developed methods, these data sets were preprocessed, modelled and statistically analyzed as described in the previous sections using the software package LIPSIA (Lohmann et al., 2001). This software package comprises tools for preprocessing, modelling, statistical evaluation, and visualization of fMRI data.

Data sets 4 and 5 contain activation coordinates that reflect results from several imaging experiments. They were used to test the meta-analysis techniques presented in Chapter 5. The majority of the included data was originally analyzed using the software package SPM (Friston, 1994; Worsley and Friston, 1995) which, like LIPSIA and most other fMRI data analysis packages, is based on the general linear model for serially autocorrelated data as presented in Section 2.2.2. The activation coordinates resulting from the individual analyses and included in our test data sets were extracted from the neuroimaging data base BrainMap (Fox and Lancaster, 2002; Laird et al., 2005b). This data base not only stores activation coordinates, corresponding statistical information, and descriptions of the performed experiments, it also facilitates the search within the stored data for particular experimental paradigms, imaging procedures, and brain regions of interest.

In the following, the experimental paradigms, imaging techniques and analysis parameters for the five test data sets that are essential for the understanding of the data are described in detail.

2.3.1   Data set 1: Stroop paradigm

For several of our analyses we employed the well-known *Stroop* paradigm or Stroop interference task (Stroop, 1935). This task is designed to investigate interference effects in the processing of a stimulus while a competing stimulus has to be suppressed.

For example, subjects are asked to name a color word, say 'red', which is presented on a screen in the color it stands for (congruent condition) or in a different color (incongruent condition). Other variants of the Stroop paradigm include the spatial word Stroop task (the word 'above' is written below a horizontal line), the counting Stroop task (the word 'two' appears 3 times on the screen) and the object-color Stroop task (an object is presented in an atypical color, e.g. a blue lemon).

For our data set, subjects performed an event-related version of the color-word matching Stroop task (Zysset et al., 2001; Neumann et al., 2003). Four subjects were examined each performing nine experimental sessions taking place within a time range of nine weeks. There were three experimental conditions (neutral, congruent, and incongruent) which are exemplified in Figure 2.6 (a). During neutral trials, letters presented in the top row of the screen were 'XXXX' printed in red, green, blue, or yellow. The bottom row consisted of the color words 'RED', 'GREEN', 'BLUE' and 'YELLOW' printed in black. For congruent trials, the top row consisted of the color words 'RED', 'GREEN', 'BLUE' and 'YELLOW' printed in the congruent color. The incongruent condition was identical to the congruent one, except that the color word was printed in an incongruent color (e.g. 'GREEN' printed in red), in order to produce an interference between color word and color name. The conditions were presented in a randomized order. Stimuli were presented until the subjects responded by tapping the index or the middle finger of the right hand, indicating whether the color of the letters in the top row corresponded to the meaning of the color word in the bottom row. This led to the presentation of a new stimulus every 6 seconds on average. Images were acquired at a sampling rate of 400 ms.

The experiment was carried out on a 3 Tesla scanner (Medspec 30/100, Bruker, Ettlingen). At 648 time points, 16 slices where acquired which covered substantial parts of the brain. Functional data were first corrected for motion using a matching metric based on linear correlation. For slicetime correction, a sinc-interpolation was used. A temporal high-pass filter with a cut-off frequency of 1/84 Hz and a spatial Gaussian filter with 4.24 mm FWHM were applied. Normalization and registration to previously acquired high-resolution anatomical data sets were performed as described in Section 2.2.1 using voxel-wise correlation and linear scaling.

In order to detect significant activation, modelling and statistical evaluation was performed as described in Section 2.2. The design matrix was generated using a Gamma function and its first and second derivative as model for the hemodynamic
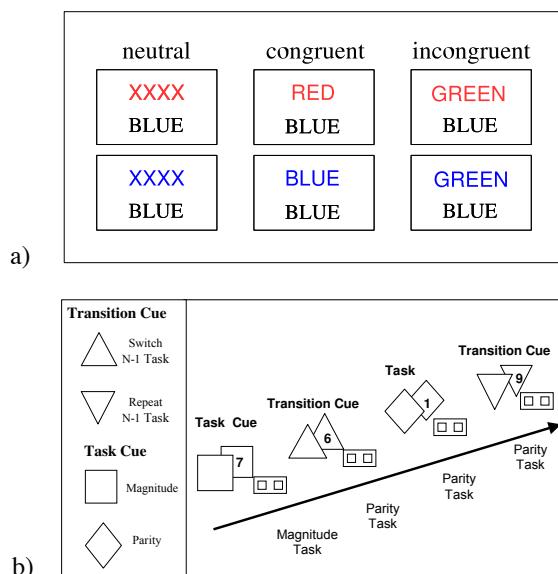
Figure 2.6: The two experimental paradigms for data sets 1 and 2. (a) Examples of the three conditions in the color-word matching Stroop task. Subjects were asked to indicate, whether the color in the top row corresponded to the color word in the bottom row. For the upper three examples, the correct answer is 'no', for the lower ones 'yes'. (b) The two different cue types (left) and an example of the cue and target presentation in the task switching paradigm.

response. This way, the amplitude as well as temporal aspects were taken into account for the detection of activated voxels. The model equation, including observation data, design matrix, and error term, was convolved with a Gaussian kernel with a dispersion of 4 s FWHM to account for serial autocorrelation. Contrast maps were generated for each subject and experimental session and one-sample t-tests were performed assessing the null hypothesis of zero activation within single subjects as well as across sessions and subjects. Statistical parametric maps SPM$\{t\}$ were constructed indicating the significance of the activation. Obtained t-values were subsequently transformed into z-values, giving a SPM$\{z\}$ for each subject and condition, for all experimental sessions of a single subject, and for the entire group of subjects. Since the contrast between the neutral and the incongruent conditions represents the main Stroop interference, only these conditions were considered in the subsequent tests of the analysis techniques described in Chapter 3.

2.3.2   Data set 2: Task switching paradigm

In our second experiment, data were analyzed from 14 subjects performing a *task switching* paradigm (Forstmann et al., 2005; Neumann et al., 2006). Subjects were presented with digits between 1 and 9 (except 5) on the screen and had to perform two different tasks, judging whether the presented digit was smaller or larger than 5 (magnitude task), or judging whether it was even or odd (parity task). Prior to the digits, subjects saw a so-called *cue*, a sign on the screen indicating which of the two tasks they had to perform next. In one half of trials, a *task cue* indicated the required task directly. In the other half, a *transition cue* told subjects to stick to the previous task or to switch to the other. Examples of the cue and digit presentation are shown in Figure 2.6 (b). Trials were presented to each participant in a different pseudo-randomized order, with equal transition probabilities between conditions and cue types.

The experiment was again carried out on a 3 Tesla scanner (Medspec 30/100, Bruker, Ettlingen). Functional slices covering the whole brain were acquired with a temporal resolution of 500 ms. Motion correction was performed using a matching metric based on linear correlation. For slicetime correction a sinc-interpolation was applied. A temporal high-pass filter with a cut-off frequency of 1/100 Hz and a spatial Gaussian filter with 4.24 mm FWHM were applied. Normalization and registration to previously acquired high-resolution anatomical data sets were performed as described in Section 2.2.1 using voxel-wise correlation and linear scaling.

Statistical evaluation was again performed based on the GLM for serially autocorrelated data as described in Section 2.2. Contrast images for individual subjects and second-level SPM(z) for the entire group of subjects were obtained for the contrast between the two different cue types. Statistically significant voxels were subsequently used to define so-called *regions of interest* (ROI), i.e. brain regions covering a particular anatomical structure or showing consistent activation in a connected component of voxels. These regions were required for the application of our exploratory first-level analysis technique described in Section 3.1.

2.3.3   Data set 3: Spatial cueing paradigm

Data set 3 was obtained for testing the Bayesian second-level analysis described in Chapter 4. It contains data from an experiment addressing differences between the left and right occipital cortex in response to *spatial cueing* (Pollmann and Morillo,

2003; Neumann and Lohmann, 2003). Twelve subjects were first presented with a
small or big visual cue (a small or big rectangle) on the left or the right side of a
screen. Cue presentation was followed by the presentation of a target (the letter 'C'
or 'O') either in the cue location, i.e. in the same visual hemifield as the cue (valid
trials), or in the contralateral visual hemifield (invalid trials). Subjects were instructed
to focus their attention to the cued area while fixating a marker in the center of the
screen. After the presentation of the target, subjects had to perform a simple target
discrimination task. The order of the 304 trials was randomized for each individual
subject.

Image acquisition and preprocessing was comparable to the procedures described
for data sets 1 and 2. Modelling and statistical evaluation as described in Section 2.2
resulted in parameter estimates for the valid-left, valid-right, valid-small, and valid-
big conditions and corresponding significance values for the contrasts between valid-
left and valid-right trials and between valid-small and valid-big trials. The Bayesian
second-level analysis described in Chapter 4 was performed on the parameter estimates
for these four conditions. Invalid trials were not further considered for our analysis.

2.3.4 Data set 4: 239 Stroop activation coordinates

Data sets 4 and 5 were obtained to test the meta-analysis techniques presented in Chap-
ter 5. They again included experimental results from the Stroop interference task as
described in Section 2.3.1. However, other variants of the task, such as the counting
and object-color Stroop task were now also considered.

For both data sets, activation coordinates were automatically extracted from
BrainMap which provides Talairach coordinates, i.e. coordinates in the Talairach coor-
dinate system, of activation maxima from functional neuroimaging experiments cover-
ing a variety of experimental paradigms and imaging modalities. At the time of obtain-
ing data set 4, the database search for fMRI experiments investigating the Stroop in-
terference task yielded experimental results reported in 16 peer-reviewed journal pub-
lications. The only criteria for the database search were the imaging modality fMRI
and that some form of Stroop interference was caused by the presented stimulus such
that an incongruent condition could be contrasted with a congruent or neutral one.
One publication was a meta-analysis and therefore excluded from further processing.
Within the remaining 15 studies, 17 data sets provided Talairach coordinates for the
contrasts 'incongruent' > 'congruent' and 'incongruent' > 'neutral'. These data, con-

taining 239 activation maxima in total, formed data set 4, which was used to test our meta-analysis technique presented in Section 5.4.

2.3.5   Data set 5: 728 Stroop activation coordinates

At the time of obtaining data set 5, the BrainMap database contained over 27,500 activation coordinates reported in 790 papers. Searching the database again for fMRI experiments investigating the Stroop interference task now resulted in 26 peer-reviewed journal publications. Within these studies, 728 Talairach coordinates for activation maxima were found. The majority of these coordinates (550 out of 728) represented the Stroop interference effect, i.e. significant activation found for the contrasts 'incongruent' > 'congruent', 'incongruent' > 'neutral', or 'incongruent' + 'congruent' > 'neutral'. As neutral condition, either the presentation of a neutral object (e.g. 'XXXX' instead of a color word) or a simple visual fixation were used. 55 coordinates were marked as deactivation in the database, i.e. they represented the contrast 'congruent' > 'incongruent'. The remaining coordinates were reported to represent other contrasts such as the contrast between different Stroop modalities or a conjunction of Stroop interference with comparable neuropsychological phenomena. 26 coordinates came from a meta-analysis on Stroop interference, 9 coordinates represented the interference effect in pathological gamblers.

As the focus of our work was on testing the newly developed meta-analysis tool described in Section 5.3 rather than on the investigation of the Stroop phenomenon as such, all 728 coordinates were subjected to the subsequent analysis without any further selection. This not only enabled us to test the method on a reasonably large data set, it also introduced some additional noise into the test data, required to assess the robustness of our approach.

Chapter 3

First-level analysis

3.1 The parcellation of functional cortical areas

3.1.1 Introduction

In this section we present the first of two exploratory methods for the analysis of single-subject fMRI data. The method is aimed at the detection of subregions in brain areas which in model-based analyses, as described in the previous chapter, appear as large homogeneously activated regions. These subregions are identified based on the similarity of their underlying fMRI signals. For this, the method facilitates two well-known mathematical concepts, *replicator dynamics* (Schuster and Sigmund, 1983) and *canonical correlation* (Hotelling, 1936). More specifically, it exploits the structure of a similarity matrix derived from the canonical correlation between fMRI time series by means of a replicator process.

Replicator dynamics originated from theoretical biology (Fisher, 1930; Schuster and Sigmund, 1983) but today has a wide range of applications, not least in the field of computer science. It is widely applied in game theory, optimization, hierarchical pattern matching and, in particular, for finding maximum cliques in graphs (Taylor and Jonker, 1978; Hofbauer and Sigmund, 1988a; Menon et al., 1995a; Pelillo et al., 1999; Pelillo, 2001; Bomze et al., 2000). In the context of functional imaging, it was first applied in the analysis of single-subject fMRI data by Lohmann and Bohn (2002). In Section 5.4 we introduce a further application facilitating replicator dynamics for the detection of functional networks from meta-analytic functional imaging data.

Prerequisite for the application of a replicator process is a non-negative, real-valued, symmetric similarity matrix, which in our case encodes the similarity between fMRI time series. Lohmann and Bohn (2002) previously suggested different suitable similarity measures for fMRI time series such as Mutual Information, Spearman's rank correlation, or the absolute value of Pearson's correlation coefficient. We propose to use canonical correlation as measure for the similarity of fMRI time series. The rationale behind this choice is as follows.

One critical question in the analysis of functional neuroimaging data is whether the obtained results are generalizable. That is, we wish to be able to compare results between different imaging sessions and subjects in order to arrive at conclusions that can be generalized to the population from which the subjects are drawn. This requires computational analysis tools that are robust against the spatial variability of the raw data which is unavoidably introduced by measurement artifacts and anatomical differences between subjects. Bivariate measures such as Pearson's correlation coefficient are strongly affected by such variability as they can only represent the relationship between two single voxels. A slight difference in the location of one voxel, for example the center of a cortical activation, can cause significantly different correlation coefficients. This will in turn affect any process that takes the similarity between voxels as input. Therefore, a similarity measure which compensates for the spatial variability between subjects is desirable. One such measure is the canonical correlation.

Canonical correlation is a multivariate extension to bivariate correlations, providing a measure of similarity between two sets of variables. It can thus be used to represent the similarity of fMRI time series between two groups of voxels. The variability analysis presented later in the chapter shows that using canonical correlation as input to the replicator process makes our method robust against the variability of fMRI data both across repeated measurements of single subjects as well as across different subjects.

In the following, we introduce replicator dynamics and canonical correlation as the two main building blocks of our analysis method. The former method is employed in the search for highly similar fMRI time series, the latter, providing a multivariate similarity measure for fMRI time series, defines the search space for the replicator process. We then present the application of our method in two experimental tasks and provide a detailed analysis of the within-subject and the between-subjects variability of the results.

3.1.2 Replicator dynamics

The aim of our method is to find groups of voxels whose fMRI time series represent the same brain functionality. Such groups of functionally coherent voxels can be detected by analyzing pairwise similarity measures between time series. Specifically, given a matrix $\mathbf{W} = (w_{ij})$ where w_{ij} represents a similarity measure between time series of voxel i and j, we wish to find a maximally coherent group, cluster, or network[1] of voxels. In this context a coherent network is defined as a network for which each member is closely connected with every other member.

A group of voxels exhibiting this coherence property can be found by determining the vector $\mathbf{x} = (x_1, x_2, \ldots, x_n)$ with $x_i \geq 0$ and $\sum x_i = 1$ that maximizes $\mathbf{x^T W x}$. Here, n is the number of voxels included in the analysis, and $x_i \in [0, 1]$ represents the degree of membership of voxel i in the maximally coherent group.

The maximization of $\mathbf{x^T W x}$ under the given constraints is know to be NP-hard, if $\mathbf{W}$ has positive eigenvalues (Gibbons et al., 1997; Pardalos and Vavasis, 1991). This is the case in our application. Note that the problem can also be interpreted as the search for the maximum weighted clique in an undirected graph which, again, is known to be a NP-hard problem (Bomze et al., 2000). Consequently, exact algorithms are guaranteed to return the global solution only in a time that increases exponentially with the number of entries in $\mathbf{W}$. They are therefore only applicable to very small problems. However, a local maximizer $\overline{\mathbf{x}}$ can be found using the following dynamical system:

$$\frac{d}{du}x_i(u) = x_i(u)[(\mathbf{W}\mathbf{x}(u))_i - \mathbf{x}(u)^{\mathbf{T}}\mathbf{W}\mathbf{x}(u)], \qquad i = 1, \ldots, n \qquad (3.1)$$

This equation is known as *replicator equation* from theoretical biology, where it is used to model the evolution of self-replicating interacting entities over time (Schuster and Sigmund, 1983; Hofbauer and Sigmund, 1988b). Replicators are entities in the evolutionary process. Each replicator i, $i = 1, \ldots, n$, which in our application are voxels in a region of interest, is associated with a *fitness*, encoded in the matrix $\mathbf{W}$, and with a *proportion* x_i indicating the percentage of the population that is identical

---

[1]In the following we will use the terms 'group', 'cluster' and 'network' interchangeably, accounting for the fact that our application of replicator dynamics addresses the coherence of voxels within the same cortical area, resulting in topologically connected groups or clusters of voxels. However, such groups are frequently termed 'network' in the replicator dynamics literature. The search for actual distributed networks of brain regions by means of replicator dynamics is pursued in Section 5.4.

to i. The term $\mathbf{x}(u)^{\mathbf{T}}\mathbf{W}\mathbf{x}(u)$ denotes the *mean fitness* of the population at a given time point u. The discrete-time version of the replicator equation is given by

$$x_i(u+1) = x_i(u)\frac{(\mathbf{W}\mathbf{x}(u))_i}{\mathbf{x}(u)^{\mathbf{T}}\mathbf{W}\mathbf{x}(u)} \;. \tag{3.2}$$

The dynamics of a replicator system are described by the fundamental theorem of natural selection (Fisher, 1930; Kimura, 1958):

Theorem:

Let $\mathbf{W}$ *be a non-negative, real-valued, symmetric matrix. Then the function* $\mathbf{x}(u)^{\mathbf{T}}\mathbf{W}\mathbf{x}(u)$ *is strictly increasing with increasing* u *along any non-stationary trajectory* $\mathbf{x}(u)$ *under both continuous-time (3.1) and discrete-time (3.2) replicator dynamics. Further, any such trajectory converges toward a stationary point* $\overline{\mathbf{x}}$*, and a vector* $\overline{\mathbf{x}} \in R^n$ *with* $\overline{x}_i \geq 0$*,* $i = 1,\dots,n$ *and* $\sum \overline{x}_i = 1$ *is asymptotically stable, if and only if it is a strict local maximizer of* $\mathbf{x}^{\mathbf{T}}\mathbf{W}\mathbf{x}$ *(Hofbauer and Sigmund, 1988b).*

This means that starting from an initial vector $\mathbf{x} = (x_1, x_2, \dots, x_n)$, we can find in an iterative process a vector $\overline{\mathbf{x}}$ that locally maximizes $\mathbf{x}^{\mathbf{T}}\mathbf{W}\mathbf{x}$. If no assumptions can be made about differences of the individual replicators, the vector $\mathbf{x} = (x_1, x_2, \dots, x_n)$ should be initialized with $x_i = 1/n$ for all $i = 1, \dots, n$ in order to avoid additional bias in the input data (Lohmann and Bohn, 2002).

During the iterative maximization process replicators with a particularly good fitness will be assigned a high proportion, while the proportion of replicators with poor fitness will decrease. Thus, when approaching stationarity a network of replicators with a particularly good fitness and consequently a high proportion in the population will emerge, given different initial fitness values of at least some replicators. This maximally coherent group of voxels is commonly referred to as *dominant network*. A detailed example for the process of finding the dominant network in a graph structure is given in Section 5.4.

Membership in the dominant network is a fuzzy concept, since the proportion of a replicator can take on any value between 0 and 1. However, we can define a membership function that classifies each replicator as belonging to this network or not based on its proportion after reaching stationarity. One obvious choice for this function is to declare a replicator to be a member of the dominant network, if its proportion increased during maximization of $\mathbf{x}^{\mathbf{T}}\mathbf{W}\mathbf{x}$ and thus, after reaching stationarity, exceeds

the average proportion of a replicator in the population. According to this membership function, replicator i is a member of the dominant network, if $x_i > 1/n$. Replicators within this dominant network can then be interpreted as the 'strongest' or 'most important' members of the system, according to the coherence measure chosen.

In order to find subsequent networks, which again show a coherent behavior among all voxels, the replicator process can be started again and re-applied to a fitness matrix from which the members of previous groups are deleted. Thus, it is possible to extract a series of coherent networks from the data. The degree of coherence among the voxels is thereby decreasing from application to application, i.e. the extracted networks are ranked according to their degree of coherence. If the true number of coherent networks in the data is not known in advance, a suitable stopping criteria for the repeated application of the replicator process needs to be defined. In our context, we wish to find groups of voxels whose coherence most likely reflects a common underlying functionality. Therefore, members of a coherent group should form a topologically connected cluster in the brain. Consequently, the application of the replicator process is repeated and groups of voxels are extracted, until they no longer form a topologically connected region. This seems intuitively plausible, since a common underlying functionality would not be expected for voxels which are not topologically connected, even if they show a stronger coherence than the remaining voxels in the investigated brain region.

The coherence within a cluster and the separability of the extracted groups of voxels can be visualized, for example, using multidimensional scaling (MDS) (Torgerson, 1952). MDS is a technique that transforms similarity values of high-dimensional data into Euclidean distance values in low-dimensional space as closely as possible. This allows for the visualization of high-dimensional data like fMRI time series in 2-dimensional space such that relative distances between data points represent the similarity between them. Details of the methodology and its application, in particular for clustering, can be found, e.g., in Davidson (1983) and Everitt et al. (2001). In the context of fMRI, MDS has been used, for example, by Friston et al. (1996); Welchew et al. (2002) and Fiebach et al. (2005).

In our application, data points in MDS are the time series of the voxels extracted as members of a coherent network by the replicator process. Therefore, MDS is applied to the similarity matrix $\mathbf{W}$ that serves as input to the replicator process. Labeling the voxels with their network membership then shows how well the different extracted networks are separable.

Given the constraints on the fitness matrix $\mathbf{W}$ described above, a number of different similarity measures for fMRI time series are conceivable. However, bivariate measures that only describe the similarity of pairs of voxels might be strongly affected by the spatial variability between subjects and completely ignore the spatial dependencies between neighboring voxels in functional images. We therefore suggest the use of canonical correlation as similarity metric for fMRI time series, a multivariate correlation that takes into account the spatial dependencies of neighboring voxels.

3.1.3 Canonical correlation

Canonical correlation analysis (CCA) was developed by Hotelling (1936) and has become a standard tool in multivariate data analysis. A thorough discussion of the method and its relation to other multivariate analysis techniques can be found, for example, in Anderson (2003) and Rencher (2002).

Consider two sets of variables $\mathbf{x} = (x_1, x_2, \ldots, x_p)$ and $\mathbf{y} = (y_1, y_2, \ldots, y_q)$. The covariance structure associated with $\mathbf{x}$ and $\mathbf{y}$ is

$$\mathbf{C} = \begin{bmatrix} \mathbf{C_{xx}} & \mathbf{C_{xy}} \\ \mathbf{C_{yx}} & \mathbf{C_{yy}} \end{bmatrix} , \tag{3.3}$$

where $\mathbf{C_{xx}}$ and $\mathbf{C_{yy}}$ are the $p \times p$ and $q \times q$ within-set covariance matrices of $\mathbf{x}$ and $\mathbf{y}$, respectively, and $\mathbf{C_{xy}} = \mathbf{C_{yx}^T}$ is the $p \times q$ between-sets covariance matrix. Let then $\rho_1^2, \rho_2^2, \ldots, \rho_s^2$ be the eigenvalues of

$$\mathbf{C_{xx}^{-1} C_{xy} C_{yy}^{-1} C_{yx}} \tag{3.4}$$

with $s = \min(p, q)$. The square roots of the eigenvalues $\rho_1, \rho_2, \ldots, \rho_s$ are called *canonical correlation coefficients*. Note that the same canonical correlation coefficients can be found as eigenvalues of

$$\mathbf{C_{yy}^{-1} C_{yx} C_{xx}^{-1} C_{xy}} , \tag{3.5}$$

however, the corresponding eigenvectors differ. It can be shown that the largest squared canonical correlation ρ_1^2 is the maximum squared correlation between the linear combinations

$$\begin{aligned} \mathbf{a^T x} &= a_1 x_1 + a_2 x_2 + \ldots + a_p x_p \tag{3.6} \\ \mathbf{b^T y} &= b_1 y_1 + b_2 y_2 + \ldots + b_q y_q , \tag{3.7} \end{aligned}$$

where $\mathbf{a}$ and $\mathbf{b}$ are the eigenvectors corresponding to ρ_1^2 obtained from Equations (3.4) and (3.5), respectively (Rencher, 2002).

In our context $\mathbf{x}$ and $\mathbf{y}$ represent time series of groups of voxels, i.e. $\mathbf{x}(t) = [x_1(t), x_2(t), \ldots, x_p(t)]$ and $\mathbf{y}(t) = [y_1(t), y_2(t), \ldots, y_q(t)]$ with $t = 1, \ldots, T$, where T is the number of measured time steps. For the application to our test data sets, presented later in this chapter, we used single voxels and their respective 6-neighborhoods in 3D space in order to form $\mathbf{x}$ and $\mathbf{y}$, i.e. $p = q = 7$. Thus, the largest canonical correlation coefficient ρ_1 provides a measure of how well the time series of a voxel and its 6 immediate neighbors correspond to the time series of another voxel and its 6-neighborhood.

It should be noted that canonical correlation coefficients are in general larger than (or at least equal to) the Pearson correlation coefficients for the same voxels. This can be explained by the fact that the canonical correlation coefficient between $\mathbf{x} = (x_1, x_2, \ldots, x_p)$ and $\mathbf{y} = (y_1, y_2, \ldots, y_q)$ represents the maximum correlation between linear functions of $\mathbf{x}$ and $\mathbf{y}$. Consequently, it exceeds the simple correlation between any x_i and y_j, $1 \leq i \leq p$ and $1 \leq j \leq q$ (Rencher, 2002). The absolute values of Pearson and canonical correlation coefficients are therefore not directly comparable. However, in our application this is not problematic, since the result of the replicator process is determined solely by the relative differences between canonical correlation coefficients of voxels.

The second, and for our application important difference between Pearson's and canonical correlation is illustrated in Figure 3.1 where both coefficients are calculated between a seed voxel and voxels in the same cortical area. As can be seen, moving away from the seed voxel, the gradient of Pearson's correlation coefficient is much steeper than the gradient of the canonical correlation coefficient. While the canonical correlation coefficient is still relatively high in two voxels' distance from the seed voxel, Pearson's correlation coefficient drops very quickly as we move away from the seed voxel. This can be explained by the fact that voxels which show a relatively weak similarity with the seed voxel in their fMRI time series, but are neighboring a voxel with a high similarity to the seed voxel, benefit from this proximity when canonical correlation is applied. To be more precise, assume that the time courses of the seed voxel and its immediate neighbors are represented by $\mathbf{x}$ and the time courses of another voxel of interest and its neighbors are stored in $\mathbf{y}$. If $\mathbf{y}$ contains a time course that is very similar to the time courses in $\mathbf{x}$, it will be assigned a high

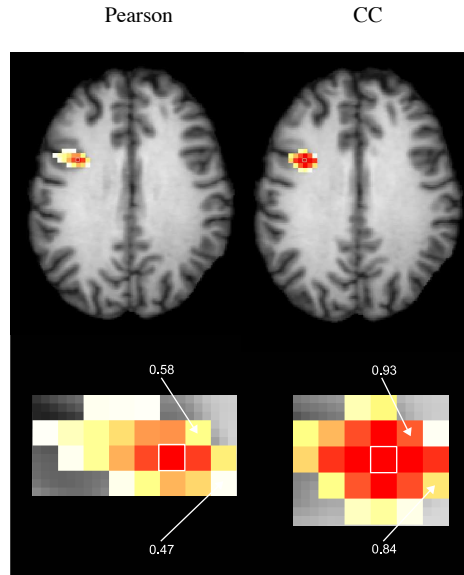*Figure 3.1: A comparison of the Pearson correlation coefficient (left) and the canonical cor-
relation (right) in a single subject performing the Stroop task (data set 1). The seed voxel is
marked with a white square. For the canonical correlation, sets were formed from a voxel and
its 6 immediate neighbors in 3D space. The voxels with the highest correlation to the seed
voxel in this slice are enlarged in the bottom row.*

coefficient in the linear combination in Equation (3.7), resulting in a high eigenvalue
corresponding to $\mathbf{b}$. This eigenvalue is then assigned as the canonical correlation
coefficient to the voxel of interest, even if its own time course is less similar to those
in $\mathbf{x}$. In other words, the similarity of a voxel's surrounding to the seed voxel is taken
into account when determining the canonical correlation coefficient. In that sense,
canonical correlation could be interpreted as a spatial filter on the correlation of time
series.

For n voxels, the largest canonical correlation coefficients for any pair of two vox-
els i and j with $1 \leq i \leq n$ and $1 \leq j \leq n$ and their respective neighborhoods can be
represented in an $n \times n$ similarity matrix. This matrix is symmetric, non-negative, and
real-valued and thus meets the criteria for an input to the replicator process described
above.

3.1.4   The algorithm

The various parts of our method for detecting groups of functionally coherent voxels from fMRI time series can be summarized in the following algorithm:

1. After defining a region of interest (ROI), calculate for each voxel in this region the canonical correlation with all other voxels in the region.

2. Store the largest canonical correlation coefficient between any pair of voxels in the similarity matrix $\mathbf{W}$.

3. Initialize the elements of the proportion vector $\mathbf{x} = (x_1, x_2, \ldots, x_n)$ with $1/n$, where n is the number of voxels in the ROI.

4. Apply the replicator process to the similarity matrix until convergence, thereby extracting the most coherent voxels.

5. Delete all members of the extracted group from the ROI and from $\mathbf{W}$.

6. Repeat the process from step 3, until the extracted group no longer forms a topologically connected cluster.

3.1.5   Relation to other methods

Adopting a broad view, replicator dynamics can be regarded as a form of clustering in the sense that similar items are grouped together and separated from less similar ones (Everitt et al., 2001). However, replicator dynamics differs from most clustering techniques in two aspects which makes it a particularly suitable tool for the functional parcellation of cortical areas based on fMRI signals.

Firstly, groups or clusters emerging from a replicator process exhibit different coherence properties than results from most other clustering techniques. While clustering typically identifies *star-shaped topologies*, where each cluster member is closely related to a single representative in the cluster center, replicator dynamics searches for *network topologies*, where each network member is closely related to *all other* network members. This complete pairwise similarity was suggested to be a plausible assumption on a functional cortical area, as such a region should only consist of voxels that *all* reflect the same underlying functionality (Lohmann and Bohn, 2002).

Secondly, in contrast to most other clustering techniques, replicator dynamics does not attempt to partition the entire feature space, i.e. to assign each voxel to one cluster. Rather, the replicator process *selects* from all input voxels groups of voxels exhibiting strong coherence properties among each other. This suits the task of detecting groups of strongly coherent voxels within a region of interest, as voxels not belonging to the 'core' of the region will not be artificially assigned to one of the clusters. Moreover, outliers and voxels showing no strong relation with other voxels are not forced to be a member of a cluster and thus do not affect the choice of cluster boundaries.

Replicator dynamics has close relations to other multivariate data analysis techniques. Like spectral clustering (Fiedler, 1973; Weiss, 1999; Ding et al., 2001), clustering by means of Hopfield networks (Hopfield, 1982; Jagota, 1995; Voultsidou et al., 2004, 2005), and cross-correlation clustering (Dodel et al., 2002) it explores the structural properties of some similarity matrix and can be reformulated as a graph partitioning problem for fully connected bidirectional (weighted) graphs. Cross-correlation clustering relies on graph-theoretic notions such as *cliques* and *connectivity components* in order to find subgraphs corresponding to clusters. In the Hopfield network approach, a thresholded similarity matrix provides the network weights. Clusters of correlated items are then identified as stationary states in the dynamics of the Hopfield network (Hopfield, 1982; Voultsidou et al., 2005). Spectral clustering explores the eigenstructure of the Laplacian $\mathbf{L}$, which is related to the similarity matrix $\mathbf{W}$ that serves as input to the replicator process. In the simplest form $\mathbf{L}$ is defined as $\mathbf{L} = \mathbf{D} - \mathbf{W}$, where $\mathbf{W}$ is some $n \times n$ similarity matrix as defined above and $\mathbf{D}$ denotes a diagonal matrix with entries $d_i = \sum_{j=1}^{n} w_{ij}$. The elements of the second eigenvector of $\mathbf{L}$ then indicate a partitioning of the input data into two disjunct groups. The inclusion of more eigenvectors provides a further sub-clustering.

Another related approach which, like replicator dynamics, operates directly on $\mathbf{W}$ rather than on the Laplacian, is principal component analysis (PCA) (Pearson, 1901). Given some data $\mathbf{z}$, the first principal component of $\mathbf{z}$ is the linear combination $y = \mathbf{x}^{\mathbf{T}}\mathbf{z}$ with maximum variance. This component is found by choosing the vector $\mathbf{x}$ which maximizes the objective function $\mathbf{x}^{\mathbf{T}}\mathbf{C_z}\mathbf{x}$ under the constraint $\sum x_i^2 = 1$. Here, $\mathbf{C_z}$ is the covariance matrix of the data, and it turns out that the first eigenvector of $\mathbf{C_z}$ is the solution for $\mathbf{x}$. In a similar manner, the replicator process finds a vector $\mathbf{x}$ that maximizes the objective function $\mathbf{x}^{\mathbf{T}}\mathbf{C}\mathbf{x}$. $\mathbf{C}$ is now the matrix of canonical correlation coefficients for our time series. The only difference in the two methods lies

in the different constraint $\sum x_i = 1$ with $x_i \geq 0$, $i = 1, \ldots, n$ on the solution of the replicator process. Note that the latter is *a priori* met in our application of the replicator process, since the matrix $\mathbf{C}$ contains only positive entries.
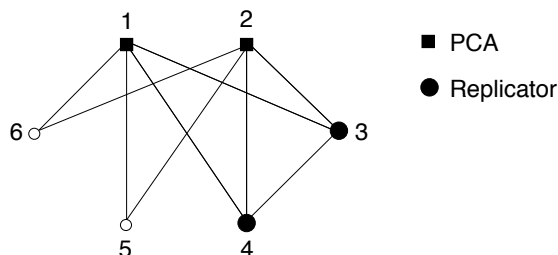

Figure 3.2: Example graph showing different solutions for the application of PCA and replicator dynamics to the problem of finding highly inter-connected subgraphs.

The difference between PCA and replicator dynamics might be negligible in some applications. However, the results of both methods can differ significantly as is shown in the following example. Consider the graph in Figure 3.2 where nodes represent voxels within some region of interest and arcs between nodes represent strong similarities between voxels. Our aim is to find groups of voxels, where each voxel shows a strong similarity with all other voxels in the group. Consequently, in the example graph, nodes should be chosen that have connections to other strongly connected nodes and are also all connected to each other. The replicator process provides the following result vector: $(0.001 \ 0.001 \ 0.499 \ 0.499 \ 0.000 \ 0.000)^{\mathbf{T}}$, where the value at position i indicates the importance of the ith node in the graph. Applying the membership function described above selects nodes 3 and 4 as solution of this problem, as the values for these two nodes increased above the initialization of 1/6 during the replicator process. In contrast, applying PCA results in the first eigenvector $(0.468 \ 0.468 \ 0.437 \ 0.437 \ 0.298 \ 0.298)^{\mathbf{T}}$. Here, nodes 1 and 2 have the highest values. While these two nodes are indeed connected to many other nodes in the graph, they do not have a direct connection and thus do not meet the requirements on our solution.

3.1.6   Application

Given a large region of voxels that are activated within the experimental contrast of interest, replicator dynamics may be able to further subdivide this region into voxels belonging to different coherent clusters, thus, indicating different functionalities of

subregions or a functional gradient. Neuroscience research recently suggested such a gradient for the lateral frontal cortex (LFC) in anterior-posterior direction, see e.g. Brass and von Cramon (2004) and Forstmann et al. (2005). More specifically, it should be possible to distinguish between an area lying in the vicinity of the inferior frontal sulcus and the precentral sulcus, the so-called inferior frontal junction (IFJ), and an area anteriorly located to the IFJ along the inferior frontal sulcus. Given this clear hypothesis, we used this cortical area as a test case for our algorithm. While its functional subdivision was impossible to detect based on the statistical significance of contrasts alone (Brass and von Cramon, 2004; Forstmann et al., 2005), we expected to find a clear parcellation when applying canonical correlation and replicator dynamics to this region.

We applied the method to data sets 1 and 2 (cf. Sections 2.3.1 and 2.3.2 and Figure 2.6), both of which contained extended activation in the LFC. A region of interest in the LFC was determined for each experiment separately using the group contrast images for the incongruent vs. neutral contrast in the Stroop paradigm (data set 1) and the contrast between the two different cue types in the task-switching paradigm (data set 2). These were the primary contrasts of interest for the investigated paradigms. In previous studies, the left IFJ was localized at Talairach coordinates x between -47 and -35, y between -4 and 10, and z between 27 and 40 (Brass et al., 2005; Derrfuss et al., 2004, 2005; Forstmann et al., 2005; Neumann et al., 2005). Therefore, ROIs were determined from the peak activation coordinate within these boundaries and all topologically connected voxels showing significant activation ($z > 3.09$, corresponding to $p < 0.001$). The resulting ROIs consisted of 100 and 210 voxels for the Stroop and the task-switching paradigm, respectively. For three subjects performing the task-switching paradigm, no functional data could be obtained for some voxels included in the mask volume. These three subjects were excluded from the subsequent analysis. Note again that the ROIs consisted of voxels that *all* showed significant activation in the contrasts relevant for the experimental tasks. The ROIs could not be further subdivided based on experimental manipulations.

Analysis Results

The Parcellation of cortical regions
The replicator process was applied repeatedly to the canonical correlation matrix of both data sets in order to find separable groups of coherent voxels. For each subject and
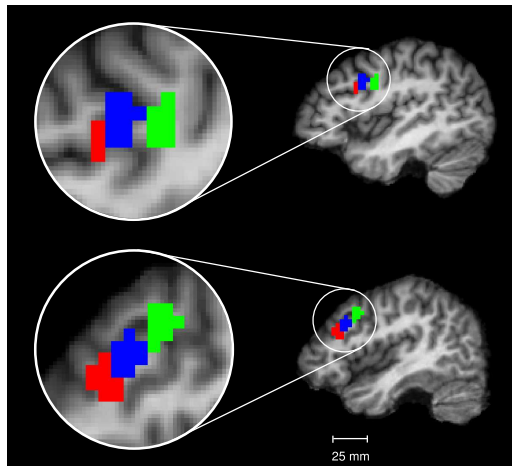
Figure 3.3: Example results from applying canonical correlation and replicator dynamics to the Stroop data (top) and the task switching data (bottom). Clusters 1 (blue), 2 (green), and 3 (red) are mapped onto anatomical slices ($x = -42$) of the individual subjects.

measurement, the application was stopped after the voxels in the extracted groups did not form a single topologically connected cluster anymore. For the Stroop paradigm, this procedure resulted in three consecutive clusters for all nine imaging sessions of each of the four subjects. Subsequent groups consisted of very few voxels distributed over the entire mask volume. The extracted voxels covered on average 75 % of the mask volume. A randomly chosen result is presented for a single imaging session of one subject in the top row of Figure 3.3.

The same procedure was applied to the data obtained for the task switching paradigm. Again, the application of the replicator process resulted in three topologically connected clusters of voxels for each subject. Subsequently extracted voxels did not form topologically connected regions. The three extracted networks covered on average 40 % of the mask volume. Note however that the mask volume was about twice as large as for the Stroop paradigm due to the higher anatomical variability between the 11 analyzed subjects. A randomly chosen result from a single subject can be seen in the bottom row of Figure 3.3.

The separation between groups was then visualized using MDS. MDS maps for the examples in Figure 3.3 are shown in Figure 3.4 (a). In both data sets the first and third group can clearly be separated from the second. Group 3 shows a relatively high

similarity to group 1, but is still clearly separable. This similarity corresponds to the physical location of the groups, with the third group neighboring the first one in anterior direction. Note that the apparent subdivision of group 2 in the task switching data is a result of the MDS projection together with group 1 and 3 rather than a representation of an inherent clustering. This can be seen in Figure 3.4 (b) where group 2 is mapped independently of the other two groups. This mapping shows an almost homogeneous cluster.

As already stated, canonical correlation can be viewed as a spatial filter on the correlation between time series. Specifically, one could argue that it favors the coherence between spatially close voxels, as they share some voxels in their respective neighborhoods. In order to verify that the parcellation of the ROIs is not exclusively driven by this effect, but reflects true similarity of the time series associated with voxels of the same coherent group, we applied the algorithm to ROIs containing spatially randomized voxels. Spatially permuted ROIs were constructed by 1,000 swaps of randomly selected pairs of voxels. This simple procedure ensures that the permuted ROIs are of the same size and physical location and contain exactly the same input time courses as the original data, only the positioning of the voxels within the ROI has changed. For comparability with the original data sets, a spatial Gaussian filter with 4.24 mm FWHM was applied after permutation.

Results of three permutation tests for the two subjects presented in Figure 3.3 are shown in Figure 3.5. For comparison, three groups of coherent voxels were extracted. If the parcellation obtained from the original data were exclusively driven by the use of canonical correlation, one would expect the application to spatially permuted ROIs to result in patterns very similar to the original data. In particular, extracted groups should have the same size, and the same number of extracted coherent groups should form topologically connected clusters. Visual inspection of the results in Figure 3.5 is sufficient to see that this is not the case. Positioning and size of the extracted groups varied between different permutation tests, and voxels within the same group do not always form topologically connected clusters. We would therefore argue that the results obtained from the original data truly reflect the intrinsic structure of the investigated regions.

(a)
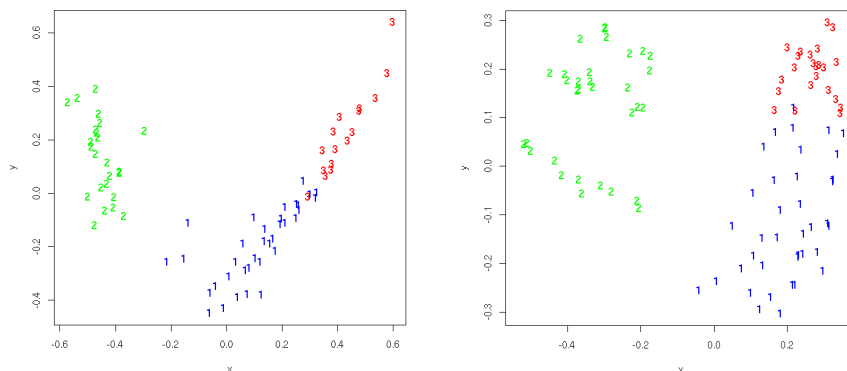
(b)
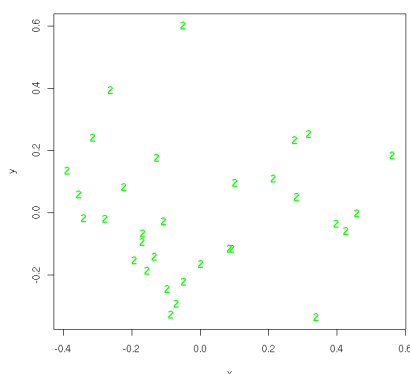
Figure 3.4: Multidimensional scaling maps for the coherent groups presented in Figure 3.3. (a) Stroop data (left), task switching data (right). Note that relative distances as determined by MDS are represented by the x- and y-axes. However, absolute distance values are arbitrary and can not be interpreted. (b) Individual MDS map of group 2 for the task switching data. The mapping shows that the apparent subdivision of group 2 in the task switching data (top right) is the result of simultaneously mapping all three groups rather than the reflection of an inherent sub-clustering.
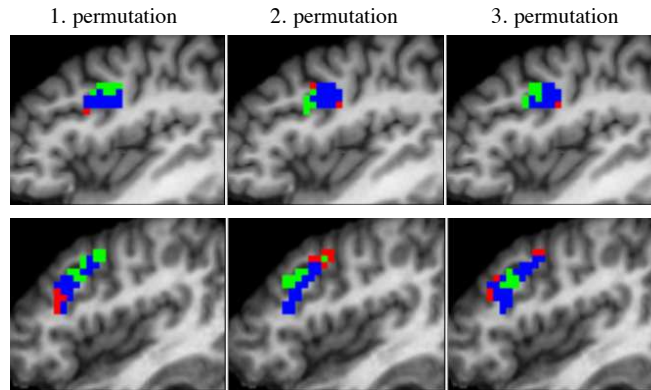
Figure 3.5: Three example results from applying the replicator process to the spatially randomized ROI of the Stroop data (top) and the task switching data (bottom). The same subjects and anatomical slices as in Figure 3.3 are shown. The extracted groups vary between different randomizations and some groups no longer form topologically connected clusters.

Analysis of the within-subject variability

The robustness of our algorithm against the within-subject variability of the BOLD signal was investigated for repeated measurements of single subjects contained in data set 1. For the first three subjects, the groups of coherent voxels were extracted in very similar anatomical locations and in the same order from all nine experimental sessions. This is exemplified for the first subject in Figure 3.6 (a) where results from all nine imaging sessions are shown.

From two sessions of subject 3, group 2 and group 3 were extracted in reversed order, i.e. the most anterior group was found second, the most posterior one last. For subject 4, the first and second group were comparable across all nine sessions. However, in three sessions the third group was not located anterior but posterior to the first, neighboring the second group in dorsal direction. Two randomly chosen example results for subjects 2 to 4 are presented in Figure 3.6 (b).

The results can be quantified by calculating the overlap of voxels in the extracted groups in all or the majority of imaging sessions. Table 3.1 shows for all subjects the average number of voxels forming a coherent group, the number of voxels that were members of the same group in all nine experimental sessions, and the number of voxels that were members of the same group in at least two thirds of the sessions.

(a)

(b)
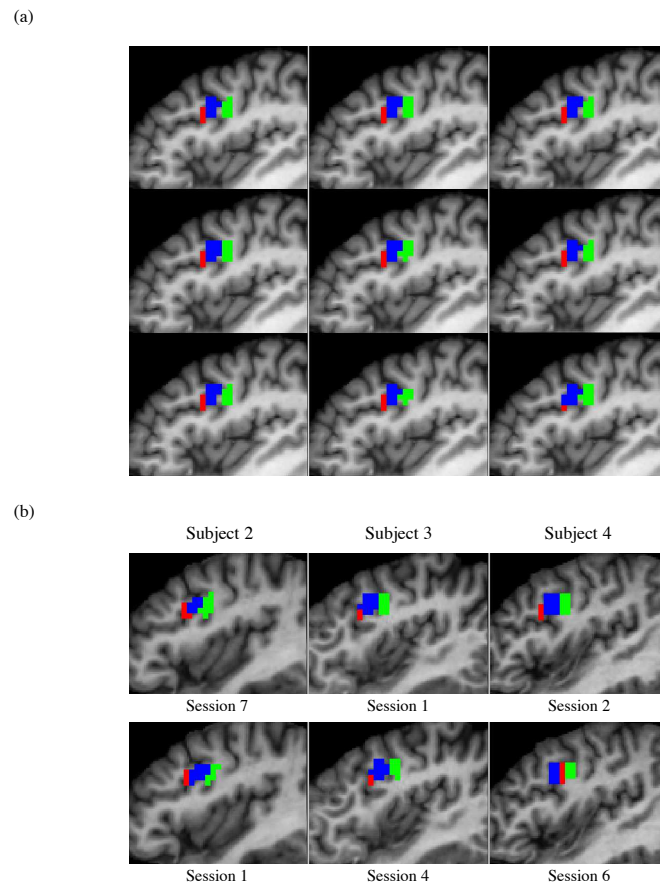
Figure 3.6: (a) Within-subject variability of the extracted groups from subject 1 performing the Stroop task. A sagittal view (x = −42) of the three coherent groups as extracted in all nine experimental sessions is shown. (b) Example results (x = −42) for the remaining three subjects contained in data set 1. Out of the nine repeated measurements, two measurements were chosen at random for each subject.

Subject	Average no. of voxels			Overlap (9 sessions)			Overlap ($\geq$ 6 sessions)		
	Net 1	Net 2	Net 3	Net 1	Net 2	Net 3	Net 1	Net 2	Net 3
1	29.3	28.0	18.8	25	18	14	27	26	19
2	31.7	27.7	18.6	17	14	13	29	26	18
3	32.4	20.7	22.0	14	12	14	31	20	22
4	30.3	23.8	17.1	11	7	—	24	19	10

Table 3.1:   The table shows for all four subjects the average number of voxels forming a coherent network, the number of voxels that were members of the same network in all nine experimental sessions, and the number of voxels that were members of the same network in at least two thirds of the experimental sessions.

For the first three subjects, at least half of the extracted voxels were members of the same group in all experimental sessions, in some cases even more than two thirds. For subject 4, at least some voxels were found consistent members of groups 1 and 2 in all sessions. The size of the overlap in two thirds of the sessions almost matched the average size of the extracted groups for all subjects.

For comparison, our analysis was repeated using the absolute value of Pearson's correlation coefficient first without spatial smoothing, then with spatial filters of different sizes. Gaussian filters of sizes between 4.24 mm and 8.48 mm FWHM were used. In Table 3.2, results are exemplified for subject 1. Similar results were obtained for the other three subjects. The analysis revealed again three separable groups of coherent voxels, showing that the strong coherence within the detected groups is not caused by spatial filtering alone. However, the size and location of the obtained groups were less stable across repeated measurements.

Analysis of the between-subject variability

We assessed the robustness of our method against the between-subject variability of the BOLD signal in a comparable manner, now using the multi-subject data provided by data set 2. The location of the three coherent groups was again very consistent across subjects. The order in which the groups were extracted was more variable than in the within-subject analysis. Most consistently, for 8 subjects the most posterior group was extracted second, in the other three cases it was extracted last. For 7 subjects the first coherent group was located between group 2 and 3 in anterior-posterior direction, in

Filter	Average no. of voxels			Overlap (9 sessions)			Overlap ($\geq$ 6 sessions)		
(FWHM)	Net 1	Net 2	Net 3	Net 1	Net 2	Net 3	Net 1	Net 2	Net 3
none	15.3	13.1	11.8	–	2	2	4	17	18
4.24 mm	15.3	13.1	11.7	–	2	2	4	17	16
7.06 mm	17.7	15.2	13.0	–	5	2	10	20	21
5.65 mm	16.7	14.1	11.9	–	8	3	11	22	18
8.48 mm	19.3	16.4	13.8	3	4	4	17	19	27

Table 3.2: The table shows for the first subject and different filter sizes the average number of voxels forming a coherent network, the number of voxels that were members of the same network in all nine experimental sessions, and the number of voxels that were members of the same network in at least two thirds of the experimental sessions. Coherent networks were obtained using the Pearson correlation coefficient instead of the canonical correlation coefficient and Gaussian spatial filtering.

the remaining 4 cases the most anterior group was extracted first followed by the most posterior one.

Table 3.3 presents the average number of voxels in the coherent groups as well as the number of overlapping voxels for 6 and more subjects. The maximum number of subjects for which voxels were found as members of the same group are 8, 9, and 11 for the three groups, respectively. Moreover, all three coherent groups contained a reasonable number of voxels that were group members in at least 6, i.e. more than half of the subjects.

Cluster	Average no. of voxels	Overlap ($\geq$ 6 subjects)					
		6	7	8	9	10	11
anterior	26.8	18	12	7	–	–	–
'middle'	29.0	20	14	9	3	–	–
posterior	26.9	26	15	13	6	4	2

Table 3.3: The table shows the average number of voxels in the coherent networks and the number of overlapping voxels for 6 and more subjects.

The overlap of groups across subjects is visualized in Figure 3.7. The top row shows group borders when all voxels were included that were group members in at

least 6 subjects. The bottom row shows the three groups separately in sagittal slices that contained the group centers, i.e. the voxels with the maximum overlap of subjects. The number of subjects where a voxel was a group member is color-coded from white (1) to red (11). As can be seen, the most consistent voxels, i.e. voxels within the borders in the top row and shown in red in the bottom row, form group centers that are clearly separated.
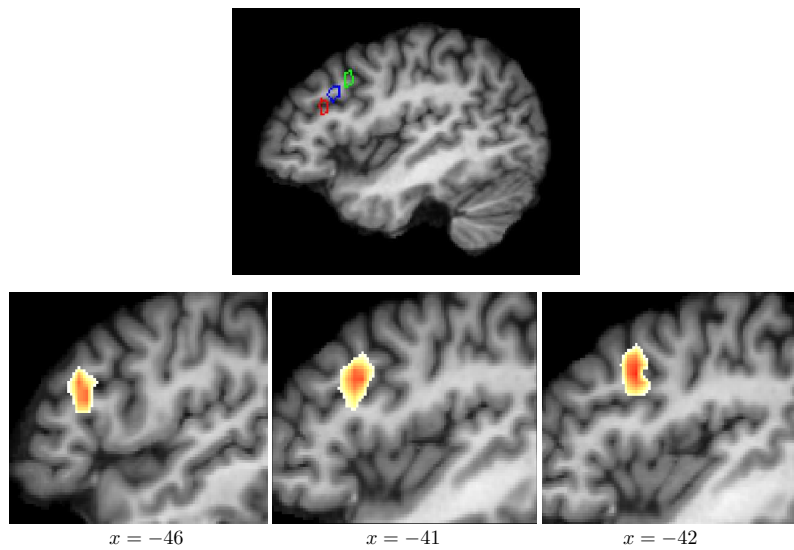

| $x = -46$ | $x = -41$ | $x = -42$ |

Figure 3.7: Between-subject variability of the extracted groups from all subjects performing the task-switching experiment (data set 2). Top: Sagittal view ($x = -42$) of the coherent groups including voxels that were group member in at least 6 subjects. Bottom: Group centers shown individually in anterior-posterior direction from left to right. The number of subjects for which a voxel was member of the coherent group is color-coded from white (1) to red (11).

3.1.7 Discussion

The method presented in this section was designed for the analysis of fMRI data on the first level, i.e. for investigating fMRI time series from individual subjects. However, an important aspect in functional imaging in general is the reproducibility of results and their comparability across multiple imaging sessions and subjects. Correlation-based analysis methods typically employ bivariate measures which capture the similarity between pairs of voxels. Using such bivariate measures as input to a replicator process

has been shown to yield good results in single subject analyses (Lohmann and Bohn, 2002). However, they are strongly affected by the spatial variability of the input data when results from repeated measurements or different subjects are compared. As our analysis of the within-subject variability shows (Table 3.2), hardly any overlap of clusters can be observed for multiple imaging sessions when bivariate correlation is used, even after applying large spatial filters to the input data. Such overlap would have been expected, however, given that the data were obtained from the same subject.

Our algorithm using canonical correlation as similarity measure between fMRI time series much alleviates this problem. It compensates for the variability of the input data, resulting in a largely increased overlap of clusters both between repeated measurements and across subjects. Canonical correlation accounts for the fact that fMRI time series always exist in a spatial context, and it captures the additional information provided by the relationship between neighboring voxels. It should be noted, however, that using canonical correlation might be disadvantageous in analyses where the precise anatomical localization of correlated cortical areas is critical. As canonical correlation operates like a spatial filter on the input data, boundaries between areas of correlated and uncorrelated voxels can be blurred, and areas of highly correlated voxels are likely to appear larger than in analyses based on bivariate correlation.

It is important to note that the replicator process is a selection process, i.e. it always extracts a network of voxels with maximal coherence from the input data, whether or not this reflects the natural structuring inherent in the data. The degree of coherence in the network, i.e. the similarity between the network members, and the dissimilarity to the remaining voxels cannot be deduced directly from the result. This can be problematic, if there exists no inherent structure in the input data at all. The extracted groups would then be meaningless and the underlying cortical area should really be treated as a homogeneous region. This problem is not exclusive to replicator dynamics. Most clustering techniques also result in some artificial structuring, if no natural structure is inherent in the data, in particular, if the number of clusters has to be specified in advance. However, a *post hoc* assessment of the compactness and separability of the obtained clusters or the use of re-randomization techniques can help to distinguish between an inherent and an artificial clustering of the data. We have used spatial randomization of the time courses in the mask volume in order to verify that the results of the replicator process reflect some inherent rather than artificial structure of the data. When three coherent groups of voxels were extracted from spatially random-

ized time courses, these groups differed in size from the original results and no longer formed topologically connected clusters. This indicates that the inherent structure of the data was destroyed by the randomization process and, most importantly, the obtained groups are not simply a result of the spatial filtering, the particular similarity measure, or the parcellation procedure used.

An important parameter in our method is the number of different networks or groups of voxels that should be extracted from the data, i.e. the number of subsequent applications of the replicator process once a network has been detected. Again, finding the optimal number of groups in some data is not specific to the repeated application of replicator dynamics, but is a common problem for most clustering techniques. Many clustering algorithms have to be provided with the expected number of clusters or, in case of hierarchical clustering, need some stopping criteria. Unfortunately, a universal solution to this problem does not exist, as the appropriate number of clusters not only depends on the structure of the input data but also on the objective of the clustering process (Halkidi et al., 2001). In our particular approach we propose to repeatedly apply the replicator process until the extracted network does not form a topologically connected cluster of voxels anymore. This is in line with the assumption that voxels which are not topologically connected would not be expected to reflect a common functionality.

3.2   Analysis of BOLD dynamics

3.2.1   Introduction

In this section we present a second techniques for the analysis of single-subject fMRI data. It is aimed at investigating temporal aspects of the BOLD signal. When examining fMRI time courses, a signal increase is often observed about 2 seconds after stimulus onset (DeYoe et al., 1994; Buckner et al., 1996; Buckner, 1998; Menon and Kim, 1999). The BOLD signal reaches its maximum approximately 5 to 8 seconds after stimulus onset and remains increased beyond the duration of the stimulus (Blamire et al., 1992; DeYoe et al., 1994; Menon et al., 1995b). The physiological and physical underpinnings of this particular shape of the BOLD response were outlined in Section 2.1 with the resulting schematic BOLD time course depicted in Figure 2.2. However, a number of studies have shown that this general shape can vary considerably when different brain regions and subjects are compared. Although it is reasonable to assume that the variability of the observed signal reflects, at least to some degree, the variability of the underlying neuronal activity, there might also be other causes. The BOLD response is, for example, sensitive to vessel diameter, whereby longer delays are found for larger vessels (Lee et al., 1995). The temporal behavior of the BOLD signal will thus be partly influenced by differences in the underlying vasculature. Moreover, differences in scanning hardware and procedures and experimental designs as well as analysis tools and data processing strategies will most likely affect any comparative study of brain activity. Knowing the exact amount of these influences on the measured fMRI signal is thus essential for the correct interpretation and statistical analysis of the obtained measurements.

In the following we present a method for the estimation of characteristic points such as time-to-peak from preprocessed fMRI time series and investigate the consistency of the estimated time points. The source of observable variation in the dynamics of the BOLD signal is again manifold, reaching from differences between repeated sessions to variations between brain regions and subjects. In a number of previous studies, the temporal behavior of the hemodynamic response has been found stable for repeated trials of a single session (Kim et al., 1997; Aguirre et al., 1998; Miezin et al., 2000). This suggests that averaging time courses over trials on a voxel-by-voxel basis preserves the temporal properties of the individual trials. However, differences on the order of a few seconds have been observed for estimates of time-to-onset, i.e.

the time point of the initial signal increase after stimulation, and time-to-peak when comparing trial-averaged time courses between subjects (Kim et al., 1997; Schacter et al., 1997; Buckner et al., 1998; Aguirre et al., 1998; Miezin et al., 2000). Variations of similar proportions have also been found in the timing and shape of the hemodynamic response across different cortical regions of individual subjects. Delays on the order of seconds and prolonged activation were, for example, observed for anterior prefrontal regions relative to visual areas (Schacter et al., 1997; Buckner et al., 1998). Miezin et al. (2000) reported considerable variation between motor and visual cortex of individual subjects, suggesting that the regional variation in the BOLD signal might even be substantially greater than any global factors influencing response properties across subjects. Delays in peak times between different cortical regions, some as long as a few seconds, were further observed by Thierry et al. (1999) and Kruggel and von Cramon (1999b).

So far, little work has been done addressing the within-subject variability of the BOLD response across a number of different experimental sessions, especially if they are several days or even weeks apart. Research into the between-session variability of the BOLD signal was largely directed towards the analysis of volume and overlap of activated voxels as well as the magnitude of their activation (Noll et al., 1997; Rombouts et al., 1998; McGonigle et al., 2000; Waldvogel et al., 2000; Maitra et al., 2002). One of the few reports addressing temporal aspects come from Aguirre et al. (1998) who found significant variability in the shape of the hemodynamic response in three out of four subjects who performed a simple visually induced motor task during five sessions taking place several days apart.

The availability of multiple sessions from single subjects in data set 1 enabled us to apply our method in the analysis of both the within-subject between-session and the between-subject variability. Given the stable temporal behavior of the BOLD response across trials in a single session, we expected the within-subject variability between scans to be considerably smaller than the between-subject variability, provided the experimental conditions do not substantially differ from one session to the next.

Finally, we tested the influence of the particular analysis technique used to obtain temporal parameters of the fMRI signal. Two general approaches can be found in the literature. Specific time points in the BOLD signal, typically time-to-onset and time-to-peak, can be derived directly from preprocessed data. Alternatively, parameters of functions assembling the assumed shape of the hemodynamic response and fitted to

the acquired data can serve as estimates of such time lags. Examples of this methodology, including linear and non-linear regressions of Gamma and Gaussian functions and linear combinations thereof, can be found in Henson et al. (2002), Liao et al. (2002), Miezin et al. (2000), Kruggel and von Cramon (1999a), and Cohen (1997). These two approaches are subject to different amounts of numerical instability and interpolation from the measured signal. In order to investigate such influences we implemented both approaches and compared the most commonly used parameter time-to-peak, once obtained directly from the fMRI signal and once estimated by fitting a Gamma function to trial-averaged time courses.

3.2.2   Temporal parameters of the BOLD response

As suggested by the works of Kim et al. (1997); Aguirre et al. (1998) and Miezin et al. (2000), averaging time courses over trials on a voxel-by-voxel basis preserves the temporal properties of the individual trials. Our method for determining characteristic time points of the BOLD signal therefore takes such trial-averaged time courses as input. Four characteristic points along trial-averaged time courses can be identified as follows. The minimum and the maximum signal amplitudes and their respective time lags with reference to the start of the stimulation, time-to-onset t_{min} and time-to-peak t_{max}, can be found by a simple search along the time course. For event-related designs, i.e. very short stimulus presentation, this search can be restricted to a time range of 0 to 5 seconds and 3 to 8 seconds, respectively, with $t_{min} < t_{max}$. These time ranges correspond to observations of the usual time ranges of time-to-onset and time-to-peak (Blamire et al., 1992; DeYoe et al., 1994; Menon et al., 1995b; Buckner et al., 1996). In addition, the first and second derivatives of the time courses at each time step can be calculated using Taylor polynomial approximations

$$\frac{df}{dt} = \frac{f(t+h) - f(t-h)}{2h} \tag{3.8}$$

$$\frac{d^2f}{dt^2} = \frac{f(t+h) - 2f(t) + f(t-h)}{h^2} \tag{3.9}$$

with $h = 1$. Time lags of the discrete approximations to the minimum and maximum of the second derivative are then determined by

$$t_{steep} : \quad \arg\max_t \quad \frac{d^2f}{dt^2} \tag{3.10}$$

$$t_{flat} : \quad \arg\min_t \quad \frac{d^2f}{dt^2} \tag{3.11}$$

for all $t \in [0 \ldots n]$, whereby

$$\frac{df}{dt}(t_{steep}) > 0, \quad \frac{df}{dt}(t_{flat}) > 0 \quad \text{and} \quad t_{steep} < t_{flat} \,. \tag{3.12}$$

The time t_{steep} marks the point along the time course where the BOLD response starts rising steeply from the base line after the presentation of a stimulus. This point can be interpreted as onset of the response function. The time t_{flat} marks the point along a time course where the function flattens out again before reaching the maximum of the activation. This point should be in close proximity to time-to-peak, but unlike time-to-peak should not be prone to estimation errors caused by prolonged activations. Such prolonged activations result in plateaus in the estimated BOLD response which make the exact identification of time-to-peak difficult. However, the length of activation should not affect t_{flat}, as this always marks the beginning of a period of increased activation, be it a single well-defined peak or a plateau of several seconds.

A typical trial-averaged time course with the four points of interest marked is shown in Figure 3.8. Note that reliably estimating t_{steep} and t_{flat} might require some temporal smoothing of the time course in order to reduce the number of local extrema of the first and second derivatives. In our application to data set 1, a temporal Gaussian filter with $\sigma = 1$ was used.

Our second method of analysis was chosen based on the observation that specific parameters or time points of the hemodynamic response can be derived directly from parameters of a model function fitted to the acquired data. As described in Section 2.2.4, it is believed that the course of the BOLD signal can be reasonably well approximated by a Gamma function (Friston et al., 1994; Boynton et al., 1996; Lange and Zeger, 1997). However, in our analysis fitting proved unsatisfactory when using the usually applied three-parameter Gamma function (Cohen, 1997; Miezin et al., 2000). In data set 1, no fit could be obtained within 10,000 iterations of the fitting algorithm for about 10 % of all activated voxels. Changes in parameter initialization had only marginal impact on these results. Moreover, visual inspection of the fitted functions revealed that, although in many cases the obtained fit resembled the shape of the underlying time course, the amplitude in particular was often not very well approximated. We therefore propose to use the four-parameter function

$$f(x) = (\frac{x}{a})^b \, e^{\frac{-(x-c)}{d}} \,, \tag{3.13}$$

leaving all four parameters subject to optimization. Note that the additional parameter allows for a more flexible modulation of the amplitude of the function.
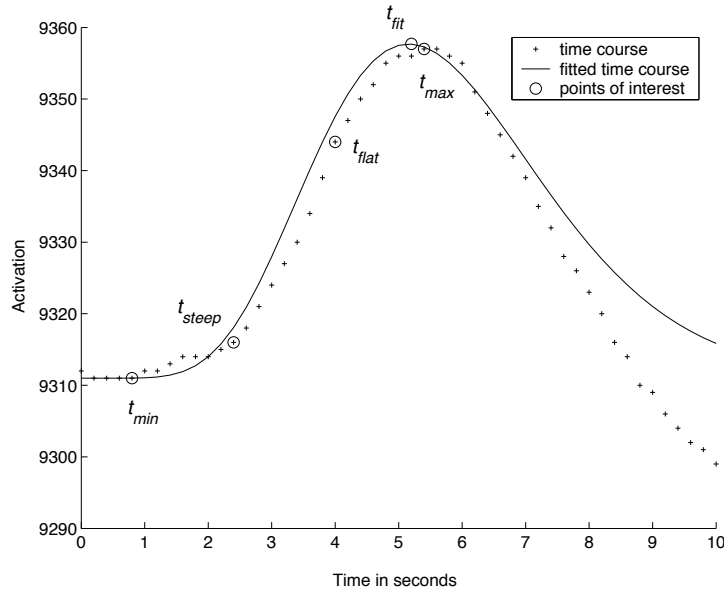
Figure 3.8: A typical trial-averaged time course of an activated voxel (data set 1, subject 1, session 2, neutral condition). The original time course and the fitted Gamma function are shown. The five parameter estimates describing distinct points between stimulus onset and maximum activation in the time course are marked.

In order to fit the trial-averaged time courses to this model function and to optimize model parameters we suggest the use of the Levenberg-Marqhardt algorithm, a standard routine for non-linear least-squares minimization (Seber and Wild, 1989). The standard deviations obtained for all data points when calculating the trial-averaged time courses can be used as additional input to the fitting procedure. This way, the adaptable parameters are forced to fit more reliably measured points better than highly unstable ones. In order to ensure the best fit of the model function in the area of our points of interest, the time range for fitting should again be restricted. In our test data, fitting was performed in the time range between the minimum of the average time course and 1 second after the maximum activation. From (3.13) the product of the parameters b and d can be taken directly as an estimate for the time-to-peak of the fitted data (Cohen, 1997; Glover, 1999; Liao et al., 2002). This point, subsequently called t_{fit} (see Figure 3.8), can be used for a direct comparison with time-to-peak obtained from the trial-averaged time course directly.

3.2.3   Application

The presented methods for obtaining characteristic time points of the BOLD response were tested using again data set 1. Data acquisition, preprocessing and the detection of significantly activated voxels are described in detail in Section 2.3.1.

Trial-averaged time courses were obtained on a voxel-by-voxel basis for each session, subject, and condition at a sampling rate of $200\,$ms which is twice the rate of image acquisition. Data for time points falling between two observed points were linearly interpolated from the weighted activation of their neighbors.

For each subject and condition, the values for the five time points of interest were averaged across the 9 sessions on a voxel-by-voxel basis. For representation, mean $\hat{t}$ and standard deviation σ can be color-coded and overlayed with intersection flat maps (Lohmann et al., 2002). These maps describe a projection of cortical regions onto a 2D plane which minimizes geometrical distortion along the lateral left-right direction and allows for a convenient inspection of wide parts of both hemispheres.

A typical distribution of the mean time lags and respective standard deviations is exemplified in Figure 3.9 for time-to-onset obtained for the incongruent condition in subject 4. Only voxels exceeding a z-value of 3.09 in the corresponding SPM$\{z\}$ are shown. Representing the results in an intersection flat map provides an overview of the time lags over the whole brain, also showing the variability of the estimates between different cortical regions. Mean time lags $\hat{t}_{steep}$ in this example range from 0.20 to 3.30 seconds with σ_{steep} between 0.14 and 2.33 seconds. Similar time differences in mean values between voxels were observed for the other subjects and conditions. Roughly speaking, $\hat{t}_{min}$ and $\hat{t}_{steep}$ were found between 0 and 4 seconds and between 0.1 and 4.4 seconds, respectively. Values for $\hat{t}_{flat}$ and $\hat{t}_{max}$ lay between 3 and 7.6 seconds and between 3.3 and 7.9 seconds, respectively. The values of the obtained standard deviations were usually in a range of 0.1 to 2 seconds, in most cases below 1 second, but in single cases as high as 4 seconds. A systematic difference in the estimates between the examined conditions could not be observed.

Note that the individual values $t_{min}, t_{steep}, t_{flat}, t_{max}$ obtained before averaging across sessions were distributed over the entire time range permitted, i.e. between 0 and 5 seconds for t_{min} and t_{steep} and between 3 and 8 seconds for t_{flat} and t_{max}, respectively. The values of t_{fit} were not restricted and even exceeded these boundaries. These widespread values can be explained by the fact that some voxels included in the SPM$\{z\}$ with a relatively small value were not activated in all individual sessions.
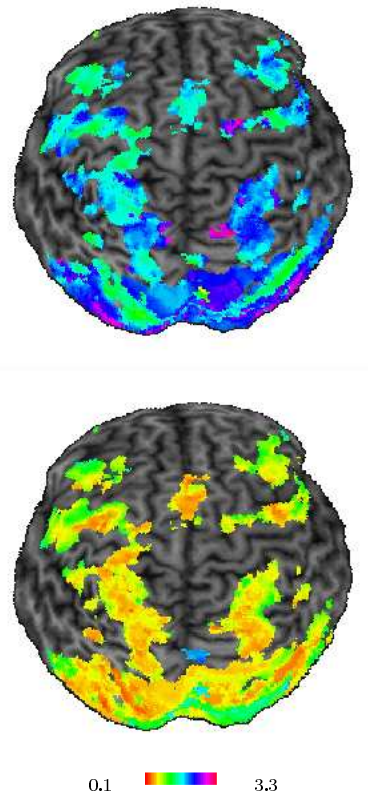
Figure 3.9: *Intersection flat maps showing means $\hat{t}_{steep}$ (top) and standard deviations σ_{steep} (bottom) in seconds for the incongruent condition in subject 4 after averaging across all sessions. Only voxels exceeding a z-value of 3.09 in the corresponding SPM$\{z\}$ are shown. Mean values vary in a range of 0.2 to 3.3 seconds in different cortical regions. Standard deviations obtained across all sessions range from 0.1 to 2.33 seconds, but exceed 1 second in very few cases only.*

For non-activated voxels, however, the shape of the time course might not be well approximated by the Gamma function and the points of interest can be placed outside the assumed range. This in turn accounts for the relatively high standard deviations observed in some cases when averaging across sessions.

The large variations observed for at least some voxels lead us to believe that the within-subject variability of the BOLD response is not significantly smaller than the between-subject variability observed in the literature. However, a closer inspection of the obtained estimates reveals that some cortical areas containing voxels with consistently small variation can be identified in all four subjects. This is exemplified in Figure 3.10 for the neutral condition of subject 1. Areas containing voxels with standard deviation $\sigma < 0.5$ are shown for all estimates. This particular threshold was chosen to be notably smaller than the variance of the estimates across subjects, which is typically as large as a few seconds.

With the exception of $\hat{t}_{steep}$, all estimates produced a number of cortical areas with such small temporal variation in the BOLD response. Although not completely identical, these areas largely overlap with highly activated areas. They include the presupplementary motor area, the left inferior frontal sulcus, the left intraparietal sulcus, and the left inferior temporal gyrus, which have also produced distinguished activations in previous Stroop studies (Zysset et al., 2001). As can be seen, however, different points of interest produced different such areas, indicating that the variances in the BOLD response do not only depend on the cortical region but might also vary along the time course of the signal.

One of our research questions pertains to the reliability of the five parameter estimates. We thus performed a comparison of the standard deviations of all five estimates for all subjects. Specifically, the most stable estimate was determined by comparing the standard deviations after averaging across all sessions of a subject on a voxel-by-voxel basis. The results are presented in Table 3.4. For each subject and condition the number of activated voxels was counted and the relative number of voxels (in %) where an estimated parameter produced the smallest standard deviation was determined. For example, 11,689 voxels were activated for the incongruent condition in subject 1. For 32.05 % of these voxels, σ_{fit} was smaller than the standard deviations of the other four estimates, which makes $\hat{t}_{fit}$ the least variable point along the trial-averaged time courses of this subject. σ_{max} was the smallest of all five estimates for 23.37 % of all voxels, followed by σ_{min} with 22.59 % of all voxels and so on. As can be seen in
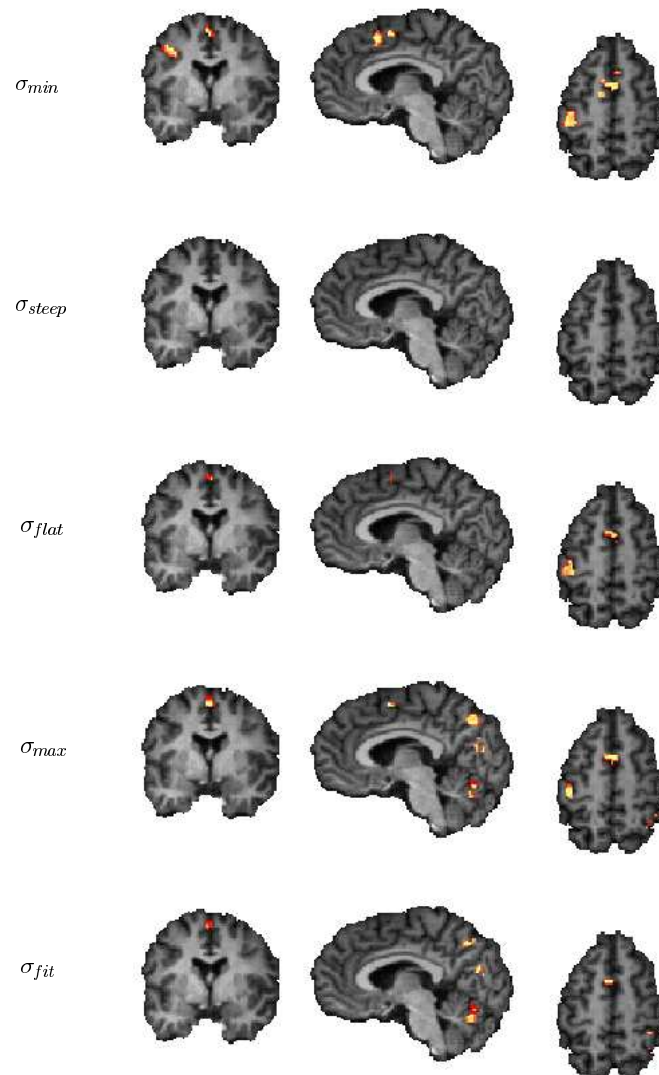
Figure 3.10: Cortical regions with standard deviations $\sigma < 0.5$ found for the five estimates in the neutral condition in subject 1. Regions with small variance largely overlapped with highly activated areas but differed between the five points along the BOLD response. In this example σ_{steep} produced no such region.

Subj.	Cond.	Voxels	σ_{min}	σ_{steep}	σ_{flat}	σ_{max}	σ_{fit}
1	incon	11,689	22.59 %	6.60 %	15.39 %	23.37 %	32.05 %
	neutral	6,769	21.23 %	2.98 %	16.65 %	29.71 %	29.43 %
2	incon	12,658	14.12 %	4.09 %	20.87 %	22.67 %	38.25 %
	neutral	10,341	11.81 %	2.82 %	32.88 %	20.10 %	32.38 %
3	incon	15,104	10.74 %	2.93 %	10.18 %	20.74 %	55.41 %
	neutral	14,081	10.18 %	2.25 %	10.87 %	26.59 %	50.10 %
4	incon	8,284	28.15 %	7.50 %	13.55 %	18.60 %	32.20 %
	neutral	6,982	25.78 %	6.86 %	18.17 %	22.16 %	27.03 %

Table 3.4: A Comparison of the standard deviations of all five estimates for the incongruent and neutral condition in all subjects. Each row shows the number of activated voxels and the relative number of voxels where each of the five estimates showed the smallest standard deviation (summing up to 100 %). In most cases, time-to-peak estimated from the fitted Gamma function shows the highest number, i.e., is the most stable among the five estimates. For the two exceptions, the neutral condition in subject 1 and subject 2, it is the second most stable estimate with nearly the same share of voxels as the first.

Table 3.4, with two exceptions, σ_{fit} was most often found the smallest of the five estimated standard deviations usually with about 30 %, in subject 3 with even over 50 % of all activated voxels. This means that $\hat{t}_{fit}$, i.e. time-to-peak estimated from the fitted Gamma function, was the most stable point along the average time courses. Among the four estimates obtained directly from the time courses, there also seems a tendency for time-to-peak to be the least variable point. This agrees with the observations reported by Miezin et al. (2000) who found time-to-peak the most stable point along trial-averaged time courses.

For the comparison of the two analysis methods, obtaining estimates directly from the averaged time courses and by means of fitting the data to a Gamma function, visual inspection of intersection flat maps for σ_{max} and σ_{fit} proves sufficient. This is exemplified in Figure 3.11 for the incongruent condition in subject 4. As becomes immediately obvious, standard deviations of both estimates are of the same magnitude. Similar results were obtained for all other subjects and conditions. We would thus argue that neither of the two approaches is generally to be preferred over the other.
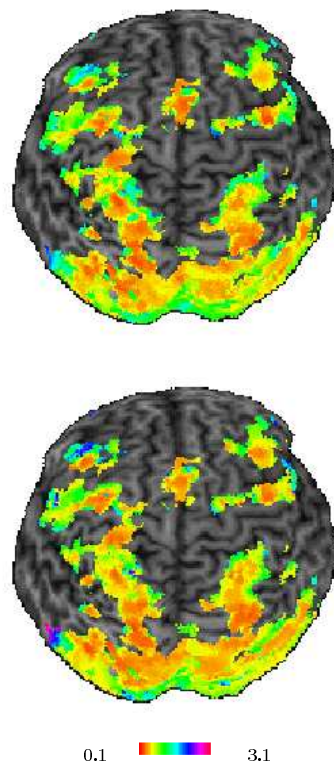
Figure 3.11: Intersection flat maps showing the two estimates for the variability of time-to-peak, σ_{max} and σ_{fit} (in seconds), for the incongruent condition in subject 4. Visual inspection already reveals that the two estimates obtained by two different methods provide nearly identical values.
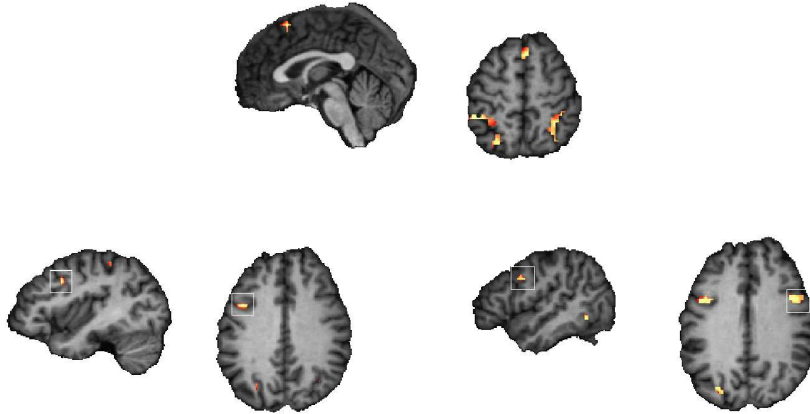
Figure 3.12:   Five cortical regions were found activated in all individual sessions for the incongruent condition in all four subjects. The regions are shown here for subject 4 and include presupplemental motor area, left and right intraparietal sulcus (top row), and left and right inferior precentral sulcus (bottom row left and right, respectively).

As shown above, using statistical parametric maps to mask activated voxels for further analysis is problematic, because they might include voxels that do not show activations in all individual sessions. We would expect considerably less variation in the parameters determined by our method when only voxels are considered which are activated in all sessions. We thus restricted another step of our test to cortical regions only containing such voxels. We chose the incongruent condition for this analysis which produced the strongest activations in all four subjects. Binary maps were obtained for all subjects marking voxels that exceeded the threshold $z = 2.33$ (corresponding to $p = 0.01$) in all contrast maps of the incongruent condition. Five cortical areas where found to contain such voxels in all four subjects. These areas are shown in Figure 3.12 and include the presupplemental motor area (preSMA), the left and right intraparietal sulcus (L SIP and R SIP), and the left and right inferior precentral sulcus (L IPCS and R IPCS). Regions of interest were formed within these areas, including the local maximum of activation together with all voxels within its 26-adjacency, i.e., voxels whose Euclidean distance from the maximum of activation did not exceed $\sqrt{3} \times$ *voxel size*. Voxels within this neighborhood that were not activated in all individual sessions were excluded from the ROI. For every ROI, average time courses, the estimates of the points of interest along the time courses, and their respective means and standard deviations across the nine sessions were obtained.

Subject	Estimate	preSMA	L IPCS	R IPCS	L SIP	R SIP
	σ_{min}	0.33	0.36	0.59	0.54	0.52
	σ_{steep}	0.69	0.74	0.55	0.97	1.15
1	σ_{flat}	0.57	0.70	0.63	0.30	0.65
	σ_{max}	0.52	0.60	0.55	0.33	0.36
	σ_{fit}	0.40	0.56	0.74	0.30	0.34
	σ_{min}	0.45	0.34	0.61	0.28	0.54
	σ_{steep}	0.87	0.72	0.99	0.68	1.17
2	σ_{flat}	0.27	0.14	0.28	0.59	0.37
	σ_{max}	0.57	0.40	0.33	0.61	0.48
	σ_{fit}	0.47	0.49	0.26	0.41	0.41
	σ_{min}	0.35	0.69	0.55	0.62	0.55
	σ_{steep}	0.79	0.94	1.07	0.44	0.84
3	σ_{flat}	0.35	0.48	0.50	0.68	0.26
	σ_{max}	0.24	0.26	0.18	0.37	0.15
	σ_{fit}	0.20	0.38	0.25	0.19	0.19
	σ_{min}	0.18	0.22	0.39	0.26	0.37
	σ_{steep}	0.21	0.48	0.83	0.55	0.65
4	σ_{flat}	0.19	0.24	0.31	0.13	0.26
	σ_{max}	0.19	0.28	0.27	0.22	0.25
	σ_{fit}	0.18	0.33	0.33	0.27	0.23

Table 3.5: Standard deviations (in seconds) of the five estimates in five comparable cortical regions for all subjects. Standard deviations for most estimates are considerably below 1 second, often found between 0.1 and 0.5 seconds. Only time-to-onset (σ_{steep}) shows higher variation for all four subjects.

The standard deviations for all estimates in the five ROIs of each subject are listed in Table 3.5. As expected, for these consistently activated voxels the standard deviations of the time lags are considerably below 1 second, often even below 0.5 seconds. An exception is σ_{steep}, the standard deviation of time-to-onset. However, time-to-onset was already found by far the most stable point along the BOLD response, shown in Table 3.4. Overall, the small standard deviations show that estimates of parameters describing the temporal behavior of the BOLD response can be meaningfully averaged across a number of sessions of the same subject as long as this averaging is performed only on voxels activated in all individual sessions.

Note that the standard deviations of the obtained time points were at least of the same magnitude when compared between subjects. However, the mean values of the estimates, shown in Table 3.6, varied considerably across cortical regions and across subjects. This is true even for the most stable estimate $\hat{t}_{fit}$ as can be seen in Figure 3.13. Subject 2 and subject 4 show a faster BOLD response in all cortical regions than subject 1 and subject 3. Moreover, mean time-to-peak varies considerably between cortical regions in all four subjects and, most notably, the order of activation in the five cortical regions differs immensely between subjects. This supports earlier observations by Miezin et al. (2000) who found that absolute estimates of time-to-peak and time-to-onset in the hemodynamic response have only a rough relation to the likely ordering of neuronal activity in different cortical regions. Despite the relatively low variance of these estimates within the same subject, their exact interpretation thus still remains an open question.

3.2.4   Discussion

The methods presented in this section facilitate the estimation of characteristic time points along the BOLD signal from preprocessed trial-averaged fMRI time series. The obtained estimates appear stable even across independent imaging sessions of the same subject, if they are derived from voxels that are significantly activated in every individual session. Among different points along a trial-averaged time course, time-to-peak appears to be the most stable one. Moreover, the two approaches to estimating parameters of the BOLD response, deriving them directly from trial-averaged time courses or from functions fitted to the acquired data, yielded nearly identical results when applied to our test data.

Subject	Estimate	preSMA	L IPCS	R IPCS	L SIP	R SIP
	$\hat{t}_{min}$	0.33	0.33	0.40	0.62	0.73
	$\hat{t}_{steep}$	1.16	1.42	1.29	1.24	1.78
1	$\hat{t}_{flat}$	4.20	4.33	4.62	4.38	5.04
	$\hat{t}_{max}$	4.71	4.91	4.87	4.89	5.47
	$\hat{t}_{fit}$	4.62	4.95	4.90	4.83	5.34
	$\hat{t}_{min}$	0.49	0.16	0.31	0.11	0.27
	$\hat{t}_{steep}$	0.96	0.40	0.82	0.29	1.38
2	$\hat{t}_{flat}$	4.16	4.07	4.13	4.00	4.04
	$\hat{t}_{max}$	4.53	4.42	4.69	4.51	4.27
	$\hat{t}_{fit}$	4.48	4.38	4.55	4.45	4.21
	$\hat{t}_{min}$	0.73	0.73	0.33	0.56	0.40
	$\hat{t}_{steep}$	1.22	1.29	1.07	2.16	1.20
3	$\hat{t}_{flat}$	4.40	4.44	4.31	4.36	4.18
	$\hat{t}_{max}$	4.73	5.04	4.96	5.20	4.73
	$\hat{t}_{fit}$	4.75	5.02	4.88	5.25	4.79
	$\hat{t}_{min}$	0.40	0.24	0.24	0.18	0.31
	$\hat{t}_{steep}$	0.89	0.76	0.80	0.96	0.67
4	$\hat{t}_{flat}$	4.00	4.20	4.07	4.00	4.07
	$\hat{t}_{max}$	4.27	4.38	4.31	4.29	4.40
	$\hat{t}_{fit}$	4.11	4.26	4.08	4.32	4.41

Table 3.6: Mean values (in seconds) of the five estimates in five comparable cortical regions for all subjects. Values vary considerably across cortical regions and subjects. A comparison of the mean values of the most stable estimate $\hat{t}_{fit}$ can also be seen in Figure 3.13.
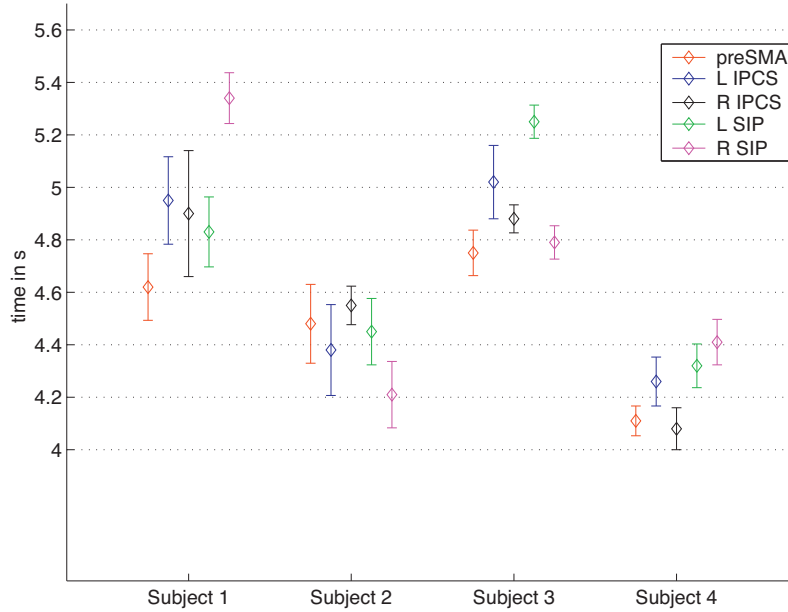
Figure 3.13: Mean values and standard errors of the time lags for the center of activation in the ROIs for all subjects. The most stable estimate $\hat{t}_{fit}$ obtained for the incongruent condition is shown. Even for this estimate, mean values vary significantly between cortical regions in single subjects and even more so between different subjects. Also note a different temporal order of the five cortical regions in all subjects.

We would argue that the relatively high variances we initially observed for some voxels in our test data set was largely caused by the fact that these voxels were not consistently activated in all sessions. As can be seen in Figure 3.11 for σ_{max} and σ_{fit}, such voxels were usually located close to the edge of clusters of activated voxels. Time courses of such voxels often did not clearly possess the parameters we wished to estimate, however. Frequently, we found a number of local minima before a larger signal increase, and the signal did not rise steadily, causing a large number of local extrema in the second derivatives. In some cases we observed two or even three peaks of activation, and some time courses did not show any increase of activation in the assumed time range at all. This of course makes providing meaningful estimates for parameters like time-to-peak and time-to-onset difficult. Moreover, the Gamma function fits only poorly to such data. Thus, given the relatively small number of sessions investigated, estimates from a voxel with no activation in only one session can already increase the

variance of the averaged data considerably. The temporal analysis of the BOLD response, though it can theoretically be applied to the whole brain, therefore only yields interpretable results for voxels activated in all investigated sessions.

It has been argued that model-based analysis methods for functional MR images are hampered by the fact that they make specific assumptions or require *a priori* knowledge about the shape of the time courses to be investigated (Duann et al., 2002; McKeown et al., 1998b). Building upon such model-based approaches in turn restricts any further analysis of temporal aspects of the BOLD response to voxels whose time course correlates well with a predicted response function. Recently, Duann et al. (2002) demonstrated that when abandoning any *a priori* assumptions about the shape of the hemodynamic response and adopting data-driven analysis approaches like independent component analysis (ICA), one finds marked variations of the derived components not only between subjects but also across stimulus types, sessions, and within sessions across trials, suggesting even higher variability of the hemodynamic response than observed with model-based analysis methods. However, components derived by PCA or ICA usually lack a clear physical interpretation. This makes a direct comparison of their variability to the variability of parameter estimates obtained with our methods impossible. It should also be noted that while abandoning *a priori* assumptions about the shape of the hemodynamic response, data-driven analysis methods are still based on assumptions which are not guaranteed to be met by fMRI data sets (Stone et al., 2002). PCA and ICA, for example, assume orthogonality and spatial independence of the derived components, respectively. Results from these methods thus depend crucially on the validity of such assumptions, just as the success of a model-based analysis depends on the correctness of the assumed hemodynamic response model. We still regard model-based approaches as appropriate tool for the analysis of BOLD response dynamics. This view is further supported by the point raised above, namely that only restricting the analysis to activated voxels, i.e., to voxels with time courses that roughly follow the assumed shape of a model function of the hemodynamic response, provides interpretable and comparable estimates of parameters describing the temporal behavior of the BOLD response.

The actual method of finding characteristic parameters of the BOLD response can also be viewed in the light of model-based versus data-driven methods. Such parameters or time points can be derived directly from preprocessed and trial-averaged data as done with our first approach. Alternatively, parameters of functions fitted to the

acquired data can serve as estimates for parameters of the BOLD response. This was demonstrated, for example, by Liao et al. (2002), Henson et al. (2002), Kruggel and von Cramon (1999a), Miezin et al. (2000), and in our second analysis approach. For our test data, both methods yielded comparable results. It is worth pointing out, though, that both methods have their advantages and disadvantages. The former approach is difficult to implement, if the data differs widely from the assumed shape and time range of the hemodynamic response. If these underlying assumptions are met, however, the derived parameters should be very accurate and close to reality, as the method operates directly on the acquired data, with few approximations and interpolations from the measured signal. In contrast, the method of fitting a function to the acquired data provides estimates of temporal properties of the BOLD response depending on parameters which are in turn estimated, and might thus be prone to high estimation errors. This was explicitly pointed out by Henson et al. (2002) who described latency differences across trial types by the ratio of two parameters of a fitted Gamma function and its derivative. On the other hand, fitting a function to the data has the advantage that obtained parameters of the model functions can directly and quickly serve as descriptions of temporal aspects of the underlying data, good fitting procedures provided. This is particularly true for fitting non-linear functions like the Gamma function to the data, where some of the obtained parameters have a clear physical interpretation such as time-to-peak or amplitude of the activation. This immediate interpretation of the fitted parameters is more difficult when fits of linear combinations of functions are employed.

Somewhat surprisingly, we observed a higher variability for the estimated time-to-onset than for all other points along the trial-averaged time courses. Moreover, the variability of $\hat{t}_{flat}$ was not, as initially expected, generally smaller than that of time-to-peak. These results might be caused by the fact that it is relatively difficult to exactly determine these two points from real fMRI signals. The estimation method can become inaccurate when applied to real data, as the first and second derivatives of the estimated hemodynamic responses still contain a large number of local extrema. The high variability of the two points might thus not so much reflect their instability in the measured signal but the still insufficient means used for their exact determination. Methods for a more accurate identification of the two points from the obtained measurements will have to be subject of further research.

Chapter 4

Second-level analysis

4.1 Introduction

In Chapter 2 we introduced the general linear model (GLM) as the most widely used mathematical model in the analysis of functional MRI data. After applying this model to the observed data, the detection of activated cortical regions is most commonly realized by means of statistical null hypothesis significance tests (NHST) based on classical[1] t- and F-statistics (Worsley and Friston, 1995; Ardekani and Kanno, 1998). These tests are typically performed on different levels of analysis. On the first level, values indicating the significance of an effect are obtained for individual subjects. On the second level of analysis, statistical random-effects analyses facilitate multi-subject or multi-session comparisons.

In recent years Bayesian techniques have been introduced to the field of functional MRI, providing a powerful alternative to linear modeling and NHST (Frank et al., 1998; Kershaw et al., 1999; Genovese, 2000; Højen-Sørensen et al., 2000; Gössl et al., 2001a,b; Friston et al., 2002a,b; Marrelec et al., 2003). These methods are aimed at a complete Bayesian analysis of the functional data on all levels. They facilitate Bayesian models for the estimation of the hemodynamic response and apply Bayesian inference for the detection of functional activation in both single subjects and groups of subjects.

---

[1]Following other authors such as Friston et al. (2002b) we will use the terms 'classical' or 'conventional' to refer to analysis methods based on a Fisher statistic and a frequentist interpretation of probability, although Bayesian methods were developed much earlier in the history of science.

In this chapter, we introduce an efficient new method that applies Bayesian techniques to the second level of analysis only. The method does not require a computationally expensive fully Bayesian approach on the first level. Rather, modeling on the level of single subjects is based on classical least-squares estimates of parameters for the GLM as described in Chapter 2. These parameter estimates are then viewed within a Bayesian framework as evidence for the presence or absence of some effect of interest in a group of subjects on the second level of analysis.

Traditionally, the analysis of functional MRI data has been viewed *either* within the frequentist *or* within the Bayesian framework. Our new approach draws on methodologies from each framework in different parts of the analysis. It is important to be clear that using Bayesian techniques on the second level of analysis does not presuppose Bayesian modeling and parameter estimation on the level of single subjects. Neither entails the classical treatment of the GLM frequentist methodologies on the level of inference. In our proposed method we view modeling and parameter estimation for single subjects within the classical, and statistical inference for groups of subjects within the Bayesian framework. This way we combine the relative simplicity of the GLM on the first level with the power and flexibility of Bayesian inference on the second level of analysis. Most notably, after parameter estimation of the GLM for single subjects, computation times for the Bayesian second-level analysis are under 10 seconds for a typical-sized group of subjects on a standard UNIX workstation.

Bayesian approaches are attractive alternatives to classical analysis methods, because they overcome a number of severe shortcoming of NHST. The p-values resulting from NHST describe the estimated probability of obtaining the observed data provided the null hypothesis of zero activation for an effect of interest is true. Consequently, sufficiently small p-values are used to reject the null hypothesis of zero activation. It is in the nature of the test that given a large enough sample size, p will always be small enough to reject the null hypothesis. On the other hand, the alternative to the null hypothesis can never be rejected. In other words, although we are testing against the hypothesis of zero activation, the test does not allow us to infer that no activation has occurred. The somewhat counterintuitive way of reading the results of a NHST has more than once led to false interpretations of the observed data (Krueger, 2001; Gigerenzer, 1993; Oakes, 1986). In contrast, Bayesian inference provides a means of directly assessing the probability for an effect of interest to take on a certain range of

values. For example, it allows us to infer the probability that a contrast between two experimental conditions is larger than zero.

Equally problematic for the application of NHST is the need for adjusting p-values according to the search volume of a test statistic in order to account for multiple comparisons. In the context of fMRI data analysis, multiple comparisons arise when repeatedly applying a t-test or F-test to assess the significance of activation in different voxels. The threshold indicating significance increases with the number of examined voxels (Friston et al., 2002b). The required adjustment implicates that inferences about some part of the brain depend on whether or not other parts have been inspected, which is not very plausible. The need to address this multiple comparison problem does not arise from the application of Bayesian inferences to individual voxels. The probability of activation in one brain area is independent of the inspection of other regions.

In the following we will summarize the main principles of Bayesian inference that are essential for the understanding of our analysis method. Excellent and more comprehensive introductions to Bayesian data analysis and Bayesian inference in statistical analyses are provided, for example, by Gelman et al. (2000) and Box and Tiao (1992). We will then demonstrate how these principles can be applied to parameters of the GLM as the results of first-level fMRI data analyses. The advantages of our method will be demonstrated and discussed using test data set 3 introduced in Section 2.3.3.

4.2 Bayesian second-level analysis

4.2.1 Bayes' theorem

Bayesian inference rests upon the posterior probability distribution of model parameters given some observed data. For a model parameter θ with the probability distribution $p(\theta)$ and the observed data y, the posterior probability distribution of the parameter given the data $p(\theta|y)$ can be calculated according to Bayes' theorem:

$$p(\theta|y) = \frac{p(y|\theta)\,p(\theta)}{p(y)}\,. \tag{4.1}$$

$p(\theta)$ is called the prior probability distribution or simply *prior* of the parameter θ, expressing our initial knowledge or belief about the value of the parameter. $p(\theta|y)$ is called the posterior probability distribution or *posterior* of the parameter, expressing our belief about the parameter in the light of evidence from the data y. Since the data

y are known and the true parameter θ is unknown, it is more convenient to express the conditional probability distribution $p(y|\theta)$ as the *likelihood function* of θ for given data y, which is written as $l(\theta|y)$ (Box and Tiao, 1992). Moreover, $p(y)$ is a function of the known data only and is constant with respect to the parameter θ. Equation (4.1) can thus be reformulated as

$$p(\theta|y) \propto l(\theta|y)\, p(\theta)\,. \tag{4.2}$$

In words, the posterior probability distribution for a parameter θ given the data y is proportional to the product of the distribution of θ prior to data observation and the likelihood of the parameter given the data. Bayes' theorem thus provides a mathematical means of combining previous knowledge with new evidence. This becomes particularly apparent when it is applied repeatedly. Assume some parameter θ and some initial data y_1. According to Bayes' theorem, the posterior of θ can be expressed as

$$p(\theta|y_1) \propto l(\theta|y_1)\, p(\theta)\,. \tag{4.3}$$

For a second observation y_2 with a distribution independent of y_1 we can state

$$
\begin{aligned}
p(\theta|y_2, y_1) &\propto l(\theta|y_2)\, l(\theta|y_1)\, p(\theta)\\
&\propto l(\theta|y_2)\, p(\theta|y_1)\,.
\end{aligned}
\tag{4.4}
$$

The posterior calculated in Equation (4.3) plays the role of the prior in Equation (4.4). The new posterior can in turn serve as prior in a subsequent step, as new data comes in. For this iterative process the posterior probability distribution can be easily calculated, if both prior and likelihood are normally distributed. Let the parameter θ be distributed as

$$p(\theta) = \frac{1}{\sqrt{2\pi}\,\sigma_0} \exp\left[-\frac{1}{2}\left(\frac{\theta - \theta_0}{\sigma_0}\right)^2\right]\,, \qquad -\infty < \theta < \infty\,, \tag{4.5}$$

and the likelihood function of the parameter be

$$l(\theta|y) \propto \exp\left[-\frac{1}{2}\left(\frac{\theta - y}{\sigma_1}\right)^2\right] \tag{4.6}$$

for an observation y, then the posterior distribution of θ is

$$p(\theta|y) = \frac{(\sigma_0^{-2} + \sigma_1^{-2})^{1/2}}{\sqrt{2\pi}} \exp\left[-\frac{1}{2}(\sigma_0^{-2} + \sigma_1^{-2})(\theta - \overline{\theta})^2\right]\,, \qquad -\infty < \theta < \infty\,, \tag{4.7}$$

which is the Normal distribution $N[\overline{\theta}, \overline{\sigma}^2]$ where

$$\overline{\theta} \;=\; \frac{1}{\sigma_0^{-2} + \sigma_1^{-2}}\, (\sigma_0^{-2}\theta_0 + \sigma_1^{-2}y) \tag{4.8}$$

$$\overline{\sigma}^2 \;=\; (\sigma_0^{-2} + \sigma_1^{-2})^{-1}. \tag{4.9}$$

The full proof is given in Box and Tiao (1992). See also Lee (1997) and Gelman et al. (2000) for discussion. Note that the resulting posterior mean can be interpreted as weighted average of the prior mean and the observed data, with weights proportional to the inverse variance.

4.2.2 Bayes' theorem and the general linear model

For k subjects, given the normal sampling distribution of the parameter estimates $\hat{\boldsymbol{\beta}}_i$, $i = 0, 1, \ldots, k$, in the GLM (cf. Section 2.2.2), we can use the formalism described above to infer about the mean of a contrast in a group of subjects. Parameter estimates obtained for single subjects can be combined in an iterative process outlined by Equations (4.3) to (4.9), given the same underlying model specification for all subjects. After establishing some prior which represents our initial belief about the mean of the contrast in the group of subjects, we can view the parameter estimates obtained from individual subjects as 'data' or evidence modifying this belief. As prior we use the probability distribution of the contrast estimated for a randomly chosen subject. In other words, we assume that the contrast of a randomly chosen subject is a good representative of the mean contrast in the entire group. This initial prior is then combined with the estimates for the contrast of a second subject, according to Bayes' theorem. Thus, we are updating our belief which is based on the observation of the first subject by evidence from the second subject.

More specifically, given the contrast of interest estimated for two subjects, $\mathbf{c}\hat{\boldsymbol{\beta}}_1$ and $\mathbf{c}\hat{\boldsymbol{\beta}}_2$, with respective variances $\sigma_1 = \mathrm{Var}(\mathbf{c}\hat{\boldsymbol{\beta}}_1)$ and $\sigma_2 = \mathrm{Var}(\mathbf{c}\hat{\boldsymbol{\beta}}_2)$, the posterior probability distribution of the combined contrast is the Normal distribution $N[\overline{\mathbf{c}\hat{\boldsymbol{\beta}}},\, \overline{\sigma}^2]$ where

$$\overline{\mathbf{c}\hat{\boldsymbol{\beta}}} \;=\; \frac{1}{\sigma_1^{-2} + \sigma_2^{-2}}\, (\sigma_1^{-2}\mathbf{c}\hat{\boldsymbol{\beta}}_1 + \sigma_2^{-2}\mathbf{c}\hat{\boldsymbol{\beta}}_2) \tag{4.10}$$

$$\overline{\sigma}^2 \;=\; (\sigma_1^{-2} + \sigma_2^{-2})^{-1}. \tag{4.11}$$

This probability distribution reflecting evidence from two subjects then serves as prior for the subsequent step of the iteration, where it is combined with the contrast estimated

for the next subject. The result of the iterative process is the posterior probability distribution of the weighted mean effect for the whole group of subjects.

If the number of subjects is known in advance, the iterative process can be replaced by a single step. For normally distributed estimated contrasts of interest $\mathbf{c}\hat{\boldsymbol{\beta}}_i$ with respective variances $\sigma_i = \mathrm{Var}(\mathbf{c}\hat{\boldsymbol{\beta}}_i)$, $i = 0, 1, \ldots, k$, obtained from k single subjects, the posterior of the combined contrast is the Normal distribution $N[\overline{\mathbf{c}\hat{\boldsymbol{\beta}}}, \overline{\sigma}^2]$ where

$$\overline{\mathbf{c}\hat{\boldsymbol{\beta}}} \;\; = \;\; \frac{\sum_i \sigma_i^{-2} \mathbf{c}\hat{\boldsymbol{\beta}}_i}{\sum_i \sigma_i^{-2}} \tag{4.12}$$

$$\overline{\sigma}^2 \;\; = \;\; \frac{1}{\sum_i \sigma_i^{-2}} \,. \tag{4.13}$$

These equations provide a very clear interpretation of the result. The mean of the posterior is the sum of the means of the individual parameter estimates weighted by their respective inverse variance. The resulting variance represents the pooled within-subject variance and is a measure of our certainty about the population mean. It further becomes obvious that the choice of the subject for the initial prior and the order of the remaining subjects does not influence the result of the iterative calculation of the posterior.

It is an important point to note from Equations (4.10) and (4.11) that the influence of each individual subject on the posterior for the entire group is determined by the estimated variance of the contrast specific to this subject. More precisely, the smaller the within-subject variance, the larger the influence of this subject on the posterior. This is illustrated in Figure 4.1, where the same prior $p(\beta) \sim N(2,1)$ is combined with parameter estimates[2] β_1 and β_2 from two subjects, whereby $l(\beta|\beta_1) \sim N(8,0.5)$ and $l(\beta|\beta_2) \sim N(8,1.5)$ for subject 1 and subject 2, respectively. The relatively small variance of β_1 causes the mean of the posterior to move towards the mean of β_1 and also results in a considerable decrease of variance from the prior to the posterior of β. The resulting posterior is $p(\beta|\beta_1) \sim N(6,0.33)$, calculated using Equations (4.8) and (4.9). In comparison, the posterior for the second subject is $p(\beta|\beta_2) \sim N(4.4,0.6)$. The larger variance of β_2 causes the mean of the posterior to move less towards the mean of β_2 and the posterior still shows a relatively high variance. This result is intuitively plausible, since the estimated variance of the parameters represents the stability of the corresponding contrast in the obtained measurement and the goodness of fit of

---

[2]For the sake of simplicity we will in the following refer to both a single parameter and a contrast of parameters as β, i.e. $\beta = \mathbf{c}\hat{\boldsymbol{\beta}}$.
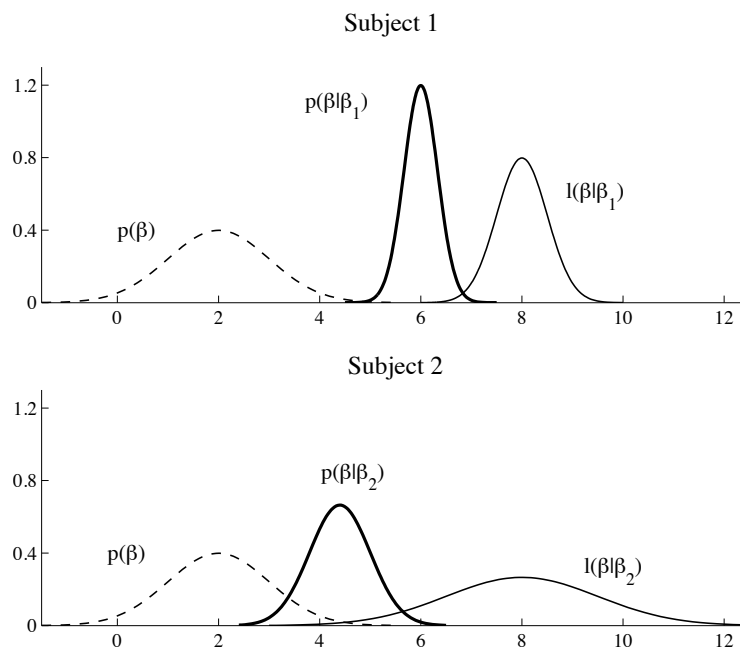
Figure 4.1: The likelihood functions $l(\beta|\beta_1)$ and $l(\beta|\beta_2)$ representing data from two subjects are combined with the same prior $p(\beta)$. The observed data have the same mean but different variances. Data with small variance (subject 1) have a larger influence on the prior than data with larger variance (subject 2). Consequently, the posterior $p(\beta|\beta_1)$ is shifted more towards the mean of the observed data than $p(\beta|\beta_2)$, and the variance of $p(\beta|\beta_1)$ is smaller than that of $p(\beta|\beta_2)$.

the model function to the data. The smaller this variance, i.e. the higher the stability of the measurements, the higher should our certainty be about the observed effect and, consequently, the more should the observed data influence or correct our belief about the true value of the effect of interest.

4.2.3   Bayesian inference

In the most simple case of a second-level analysis of fMRI data we wish to make inferences about the presence or absence of an effect of interest in a group of subjects. Such effect is usually the activation corresponding to an experimental condition or a contrast, i.e. the difference in activation between two conditions. With a NHST the null hypothesis of zero activation or zero contrast is assessed based on estimated contrasts

for individual subjects. The p-values resulting from the test describe the estimated probability of obtaining these individual contrasts provided the null hypothesis of zero activation is true. If this probability is small enough, we reject the null hypothesis. The test does not tell us, however, how certain we can be that the effect is present in the group, if the null hypothesis was rejected. This question can be directly addressed using Bayesian inference.

Bayesian inference is based on posterior probability distributions. From the posterior mean and variance of an effect of interest we can calculate the probability for the effect to take on a range of values. This provides us with a measure of certainty about the effect in the group. For any probability density function $f(\theta)$ the probability that the random variate θ takes on a value in the interval $[a, b]$ is

$$P\left(a \leq \theta \leq b\right) = \int_a^b f(\theta)\, d\theta\,. \tag{4.14}$$

For example, given estimated contrasts $\hat{\beta}^1, \hat{\beta}^2, \ldots, \hat{\beta}^n$ of n individual subjects, the probability of a positive contrast β for the entire group is

$$P\left(\beta > 0\right) = \int_0^\infty p\left(\beta \mid \hat{\beta}^1 \hat{\beta}^2 \ldots \hat{\beta}^n\right) d\beta\,. \tag{4.15}$$

Note that while the mean β of the posterior tells us something about the size of the effect of interest, the posterior probability enables us to mathematically express the strength of evidence for the effect (Genovese, 2000). The ability to calculate this probability thus facilitates statements such as 'Given the evidence of the observed data, we believe that in the investigated region the contrast is positive with 95 % probability.' Such statements very directly address the question of localizing regions with stimulus-related activation in the brain, and they are a much more intuitively plausible interpretation of the observed data than the rejection of a null hypothesis.

By means of Bayesian inference we can also tackle more complex research questions that are hard or impossible to formulate in terms of traditional hypothesis testing. This has been demonstrated, for example, for the comparison of activation amplitudes in different voxels (Frank et al., 1998) or the assessment of monotonicity of experimental conditions (Genovese, 2000). In the latter case, using NHST in order to assess the monotonicity of, say, four parameter estimates $\beta_1 \leq \beta_2 \leq \beta_3 \leq \beta_4$ for four different experimental conditions, requires the repeated application of single tests for the hypotheses $\beta_2 - \beta_1 \geq 0$ and $\beta_3 - \beta_2 \geq 0$ and $\beta_4 - \beta_3 \geq 0$. The resulting p-values have in turn to be corrected for multiple comparisons. In contrast, within a Bayesian

framework the posterior probability distributions for the parameters β_1, β_2, β_3, and β_4 allow the direct computation of $P\left(\beta_1 \leq \beta_2 \leq \beta_3 \leq \beta_4 \mid y\right)$.

When examining the results of fMRI experiments, we are often interested in comparisons of different groups of subjects, for example left-handed and right-handed participants of a study or subjects of different age or sex. The posterior distributions resulting from our analysis can directly be used to infer about differences between the means of two groups of subjects. Given the two posterior probability distributions $p\left(\beta_1 \mid \hat{\beta}^1 \ldots \hat{\beta}^k\right)$ and $p\left(\beta_2 \mid \hat{\beta}^{k+1} \ldots \hat{\beta}^n\right)$ for a contrast of interest in two groups of k and $n-k$ subjects, respectively, and assuming independence of the two groups, the joint distribution of the contrast is

$$p\left(\beta_1, \beta_2 \mid \hat{\beta}\right) = p\left(\beta_1 \mid \hat{\beta}^1 \ldots \hat{\beta}^k\right)\ p\left(\beta_2 \mid \hat{\beta}^{k+1} \ldots \hat{\beta}^n\right) \tag{4.16}$$

with $\hat{\beta} = \{\hat{\beta}^1 \ldots \hat{\beta}^k, \hat{\beta}^{k+1} \ldots \hat{\beta}^n\}$. The posterior of the difference in means $d = \beta_2 - \beta_1$ is then the correlation of the two independent distributions (Frank et al., 1998). For normally distributed $p\left(\beta_1\right) \sim N(\beta_1, \sigma_1^2)$ and $p\left(\beta_2\right) \sim N(\beta_2, \sigma_2^2)$ this correlation has the analytical form

$$p\left(d \mid \hat{\beta}\right) = \frac{1}{\sqrt{2\pi}\,\sigma} \exp\left[-\frac{1}{2}\left(\frac{d - \mu}{\sigma}\right)^2\right], \tag{4.17}$$

with $\mu = \beta_2 - \beta_1$ and $\sigma = (\sigma_1^2 + \sigma_2^2)^{1/2}$ (Box and Tiao, 1992; Frank et al., 1998). The probability for a difference in means between groups can then be calculated by integrating the posterior probability distribution in the respective interval. Other tests for differences in means such as comparisons of different contrasts in a single subject or within one group of subjects can be calculated in the same manner.

Posterior probability distributions can be summarized and visualized in various ways. Maps of the posterior mean of an effect provide estimates for the effect size in every voxel. Posterior probability maps reflect the probability with which we can expect the effect to be found in a population. As one is usually interested in both, how large is the effect of interest and how likely is it to occur, we will for our test data set present maps of the posterior means together with the corresponding posterior probability maps.

4.3   Application

The method introduced in the previous section was applied to data set 3 described in Section 2.3.3. The data set contains results from an experiment addressing differences between the left and right occipital cortex in response to spatial cueing (Pollmann and Morillo, 2003). Our Bayesian second-level analysis was implemented in C and computations were performed with an AMD Athlon(TM) XP1800+ processor and 768 MB working memory. Computation time for the complete second-level analysis was under 10 seconds for each contrast including the extraction of individual contrast means and variances from the results of the first-level analysis.

For the most prominent experimental contrast, valid-left against valid-right trials, the pattern of activation obtained from the classical analysis was replicated with our Bayesian approach. The most significant activations were found in the left and right lingual gyrus, in the lateral occipital gyri of both hemispheres, and in the junction of the right intraparietal sulcus and the transverse occipital sulcus (IPS/TOS). For the centers of activation in these regions, coordinates in the Talairach space, z-values obtained from the classical analysis, and means and standard deviations of the posteriors from the Bayesian analysis are listed in Table 4.1.

The regions are visualized in Figure 4.2. The top row shows the posterior mean values of the contrast obtained from the Bayesian second-level analysis (left) and the posterior probability maps for $P(\mathbf{c}\boldsymbol{\beta} > 0)$ (middle) and $P(\mathbf{c}\boldsymbol{\beta} < 0)$ (right). A threshold of 99.9 % was applied to the probability maps for visualization. The bottom row shows the corresponding SPM$\{z\}$ from the classical analysis, as described in Section 2.2, thresholded at $z = 3.09$.

As can be seen, the posterior probability maps correspond well with the SPM$\{z\}$. $P(\mathbf{c}\boldsymbol{\beta} > 0)$ and $P(\mathbf{c}\boldsymbol{\beta} < 0)$ exceed 99.9 % in regions with significant positive and negative z-values, respectively. However, whereas the SPM$\{z\}$ suggests five centers of activation within the significantly activated areas, the posterior probabilities are more homogeneously distributed over the significantly activated areas. In other words, the probability for activation is very high both in the centers of activation detected with the classical method and in their surrounding voxels. This seems intuitively plausible, since we would expect that the probability of activation in voxels close to an activation focus is still very high, even if the strength of the activation is smaller than in the center.

Differences between the five centers of activation which are not obvious from the SPM$\{z\}$ can be detected from the posterior means. The posterior means in the lingual

	ROI	Location	z (NHST)	Mean	SD
1	Lingual Gyrus	R (13 −83 −4)	4.36	0.23	0.02
2	Lingual Gyrus	L (−14 80 −1)	−4.79	−0.25	0.02
3	Lateral Occipital Gyrus	R (31 −70 2)	3.83	0.10	0.01
4	Lateral Occipital Gyrus	L (−38 −70 2)	−4.51	−0.09	0.01
5	IPS/TOS	R (28 −71 23)	4.13	0.12	0.01

Table 4.1: Centers of the most significant activations for the contrast between left-valid and right-valid trials. z-values resulting from a classical analysis and means and standard deviations (SD) from the second-level Bayesian analysis are shown. While z-values are comparable for all regions, differences between the lingual gyri and the remaining areas can be found for the posterior means and standard deviations.
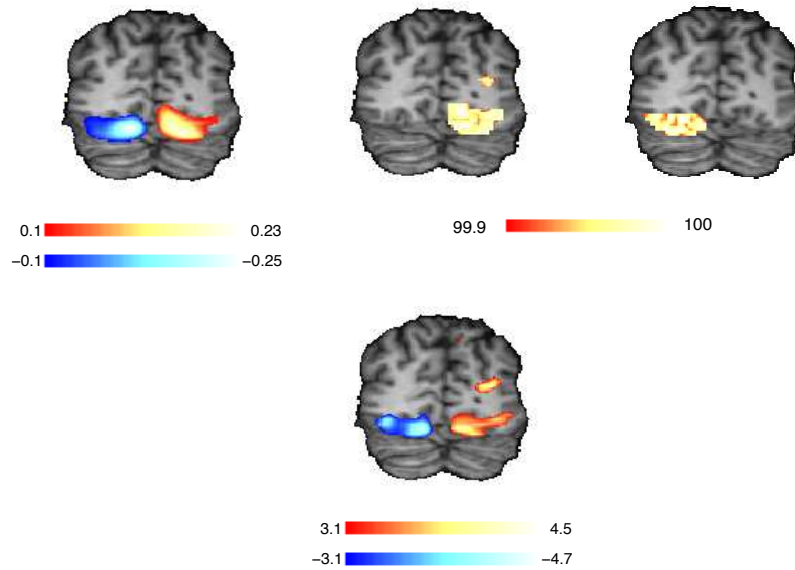
Figure 4.2: Posterior means (top left) and posterior probability maps (in %) for $P(\mathbf{c}\boldsymbol{\beta} > 0)$ (top middle) and $P(\mathbf{c}\boldsymbol{\beta} < 0)$ (top right) for the contrast between valid-left and valid-right trials. The corresponding SPM{z} is shown in the bottom row. Note that negative z-values and posterior means (blue) indicate activations for the inverse contrast, i.e. valid-right against valid-left trials.

gyri are considerably higher than in the other regions (see also Table 4.1) suggesting that the contrast between valid-left and the valid-right trials is much stronger there than in the remaining activated areas. Note that the value of the posterior mean in our model is proportional to the difference in the maximum signal amplitude for valid-left and valid-right trials.

A second possible contrast in the experimental data arises from the presentation of small or big cues (valid-small against valid-big trials). Previous behavioral experiments suggested much smaller effects than for the main contrast between valid-left and valid-right trials. However, our Bayesian analysis of the fMRI measurements revealed a number of cortical regions with posterior probabilities that were comparable to those estimated for the main contrast. Three of these regions are visualized in Figure 4.3. Here, the Bayesian analysis yielded posterior probabilities of 98 % or higher, whereas the values in the classical SPM$\{z\}$ did not exceed the threshold of $z = 3.09$.
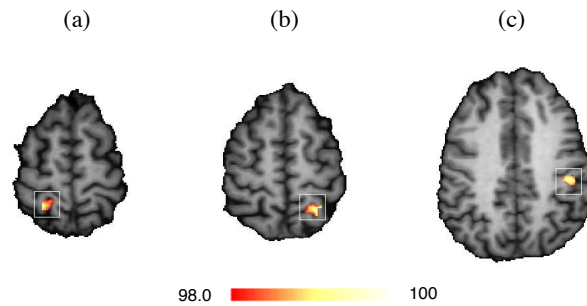
Figure 4.3: *Posterior probability maps (in %) for* $P(\mathbf{c}\boldsymbol{\beta} > 0)$ *for the contrast between valid-small and valid-big trials. The contrasts for all individual subjects in the three marked regions are further analyzed in Figure 4.4 (a-c).*

One reason for the different results obtained with both methods becomes visible when analyzing the contrast found for the individual subjects. Figure 4.4 (a-c) shows the contrast for all individual subjects for the centers of activation in the three cortical areas marked in Figure 4.3 (a-c), respectively. For comparison, Figure 4.4 (d) shows data for the center of activation in the inferior temporal sulcus obtained for the previous contrast. Here, the classical method yielded a z-value of 4.49 which is usually regarded as significant activation.
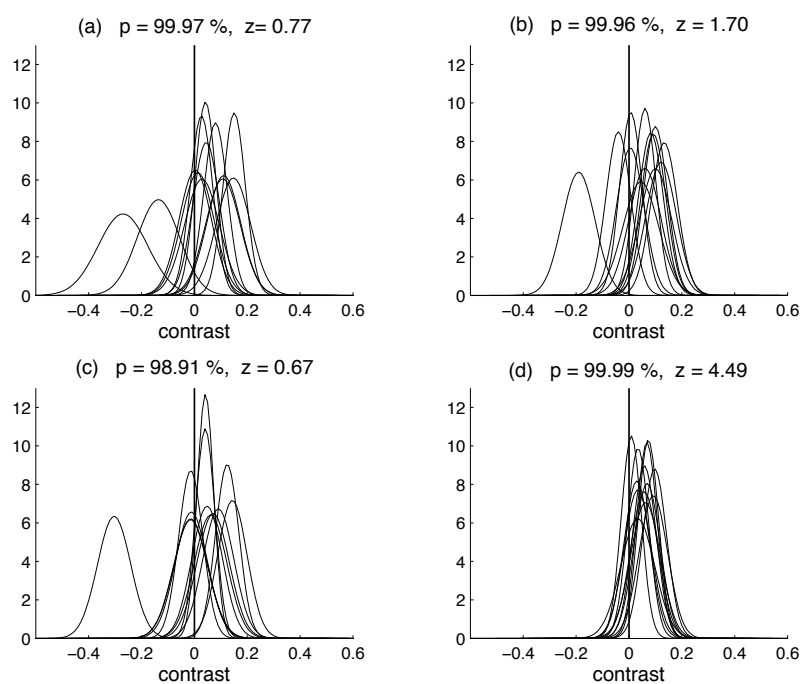
Figure 4.4: The sampling distribution of the contrast between valid-small and valid-big trials estimated for all individual subjects in three voxels (a-c). These three voxels have the highest posterior probability in the cortical areas marked in Figure 4.3 (a-c), respectively. These areas show no significant activation in the classical method. For comparison, (d) shows the contrast between valid-left and valid-right trials for all individual subjects for the center of activation in the inferior temporal sulcus. This region was also found significantly activated using the classical method. Posterior means and standard deviations were similar for all four voxels: (a) 0.052, 0.015, (b) 0.048, 0.014, (c) 0.031, 0.014, and (d) 0.054, 0.013, respectively.

For all four voxels in Figure 4.4 the estimated contrast is larger than zero for the majority of subjects. However, two subjects in voxel (a) and one subject in the voxels (b) and (c) differ from the general pattern with large negative contrasts. Such differences in the estimated contrasts can be caused, for example, by large anatomical variations between subjects which cannot fully be accounted for by the preprocessing procedures. Outliers can also result from differences in the temporal behavior of the BOLD response of individual subjects, as seen in Section 3.2. Relatively large temporal offsets between the observed data and the model function for the hemodynamic response can lead to poor fitting and estimation of the model parameters, in particular if the same model function is applied to all subjects which is a prerequisite for both the classical random-effects analysis and our Bayesian method.

In our test data, the estimated variances of the outliers are larger than the variances of the estimates for most other subjects. As we have already seen from Equations (4.10) and (4.11) and from Figure 4.1, with our approach the influence of each individual subject on the posterior for the entire group is determined by the estimated variance of the contrast specific to this subject. Given the relatively large variances of their estimates, the influence of the outliers' contrasts is not large enough to move the posterior mean close to zero. Consequently, the posterior probability $P\left(\mathbf{c}\boldsymbol{\beta} > 0\right)$ is still very high, whereas the significance values from the classical method do not exceed the required threshold due to the high between-subject variance caused by the outliers. For comparison, the classical second-level analysis was repeated for the three regions (a), (b), and (c), now omitting the two outliers for (a) and the single one for (b) and (c). The z-values for the three voxels increased to 3.14, 3.19, and 2.61, respectively. Although these values are still comparatively small, the large increase relative to the initial analysis shows clearly that the outliers had a huge impact on the results of the classical analysis.

4.4   Discussion

In this chapter, we have introduced a Bayesian method for the second-level analysis of fMRI data which is based on modeling the obtained measurements for single subjects by means of the GLM. The method is easy to implement and computationally inexpensive. The required computation time is on the order of seconds for a complete second-level analysis following a relatively simple classical modeling on the first level. This

is in stark contrast to alternative approaches using non-linear or hierarchical Bayesian modeling. Complex Bayesian models can be computationally expensive, and computation times on the order of hours or even days for single subjects have been reported (Genovese, 2000). Despite its mathematical simplicity, our method overcomes some of the drawbacks of NHST such as the need to address the problem of multiple comparisons. It provides estimates for both the size of an effect of interest and the probability of the effect to occur in the population. The results are easy to interpret and intuitively more plausible than results of classical NHST. Like other Bayesian approaches, our method permits complex inferences which are hard to derive from NHST.

Note that our Bayesian second-level analysis could also be combined with a different first-level analysis and is not restricted to the use of the GLM on the first level. One could for example conceive of a nonlinear model on the first level resulting in parameters with a clear physical interpretation such as a direct estimate of the amplitude or time delay of the observed signals. The only prerequisite for the straightforward application of Bayes' theorem presented here is that the effects of interest are described for single subjects as normally distributed variables.

NHST is based on a frequentist interpretation of probability. The probability of an event is defined as its relative frequency and is therefore, when viewed over a large number of trials, a constant. Consequently, a hypothesis about the event can only be true or false, and the observed data helps us to decide between these two possibilities. They do not allow us, however, to assess the probability of our hypothesis to be correct, although p-values resulting from NHST are often wrongly interpreted this way (Krueger, 2001; Gigerenzer, 1993; Oakes, 1986). Bayesian inference on the other hand provides us with exactly this information. Here probability is viewed as an individual's belief about an event which is modified by the observed data. Our initial belief (or hypothesis) about the event, represented as prior probability distribution, is modified by the observed data whereby we become more certain about the true nature of the event the more data we encounter.

Critics of Bayesian techniques often stress the subjectivity inherent in the methods. Clearly, the posterior probability distribution of an event crucially depends on the chosen prior, i.e. on the experimenter's belief about the event in question. However, we would argue in line with Lange (1997) and others that this subjective element should be regarded as a virtue rather than a disadvantage. It provides us with a method to incorporate knowledge and experiences from previous studies or subjects into our

model and combine them with newly acquired data. Making use of this virtue, we take the probability distribution estimated for one subject as our initial belief about the true distribution of the parameter in the entire group, i.e. as the first prior in the iterative application of Bayes' theorem. This way only actually observed data enter into the calculation of the posterior for the group of subjects.

It is also important to note that some degree of subjectivity enters into non-Bayesian models, too, when choosing the experimental design, formulating hypotheses, or selecting model parameters (Petersson et al., 1999; Krueger, 2001). As Gössl et al. (2001a) point out, model specifications such as the choice of basis functions in the GLM provide even harder constraints on the solutions than Bayesian priors. While the former specify a subspace in which the solution must lie, the latter only impose soft constraints on the solutions which can be violated, if a sufficient amount of acquired data provides appropriate evidence.

In the first-level analyses presented in Chapter 3 we observed that a considerable degree of variability in the fMRI signal can be found both in individual scans of the same subject and in different subjects of a group. Different classical models allow the within-subject and the between-subject variance of the observed signals to enter the second-level analysis to varying degrees. As outlined in Section 2.2.2, classical fixed-effect models do not take into account the between-subject variability of the responses. Rather, they are based on the assumption that all subjects respond with the same variance and thus utilize the within-subject variance as the only variance component. In other words, the acquired data are treated as if coming from a single subject. While the resulting large number of degrees of freedom facilitate a high sensitivity of the method, a classical fixed-effect model can produce significant effects by virtue of a single subject.

Mixed- or random-effect models are designed to take into account both the within-subject and between-subject variability of the responses. Subjects are viewed as randomly sampled from a population and the effects estimated for each subject are treated as random variables. The resulting variance of the estimated response across subjects contains both within- and between-subject variance components in a proportion determined by the ratio of scans per subject to the number of subjects (Friston et al., 1999). However, the analysis of random-effect models is often difficult (Searle et al., 1992) and the usually small number of subjects in fMRI experiments results in a low power of such analyses. As already noted in Section 2.2.2, a relatively simple two-stage model

implementing a random-effect analysis can be applied under the conditions that the model is balanced, i.e. the same experimental design is used for all subjects, and that the model is separable by subject, i.e. the parameter estimates for each subject are independent (Holmes and Friston, 1998). This implementation builds upon the idea of a simple summary statistics (Frison and Pocock, 1992). It should be noted, however, that the model rests on the simplifying assumption that the within-subject residual variance is constant for all subjects (McGonigle et al., 2000). This means that while the within-subject variance is one component of the overall variance, differences in the within-subject variance between subjects are neglected in the analysis. As our test data has shown, these models are still relatively sensitive to outliers which cause a high between-subject variance.

In our Bayesian approach the variance of the resulting posterior is the pooled within-subject variance of all subjects (Worsley et al., 2002). The between-subject variance is expressed by the spread of the means of the estimated parameters for individual subjects. Note that while the actual location of the means clearly influences the posterior mean for the group, their variance does not necessarily do so. In this respect our method implements a fixed-effect analysis. However, unlike in classical fixed-effect analyses, measurements are not viewed as coming from a single subject. Modeling on the first level is performed independently for each individual subject and the Bayesian inference allows for different within-subject variances. Most importantly, the within-subject variance, i.e. the stability of the measurements obtained for the individual trials and the goodness of the model fit, determines the influence of a subject on the posterior for the group. The influence of a few outliers in a group is small as long as their within-subject variance is not considerably smaller than those of the remaining subjects. This is a large advantage over conventional methods where the influence of each individual subject is not weighted by its within-subject variance. Consequently, outliers can cause the between-subject variance to increase considerably independent of their within-subject variance, which in turn results in small t- and subsequent z-values. Such outliers cannot always and completely be avoided given the large anatomical and physiological variability in the population. Therefore, robustness against such outliers is a prerequisite of powerful analytical tools for the evaluation of fMRI data. Our Bayesian second-level analysis meets this prerequisite.

Finally, a comparison between our Bayesian approach and a classical analysis on the same test data set revealed that the latter might disregard a number of activations on

the basis of their relatively low significance. We agree with Friston et al. (2002b) who observed that "there is no magical increase in power afforded by a Bayesian approach." However, we would also argue that we must not ignore cortical areas for the neuro-psychological interpretation of experimental results, where the posterior probability of an effect is as high as 99 % or above, just because they have missed a more or less arbitrary threshold in the classical analysis. They at least demand a closer look, which our Bayesian second-level analysis method affords.

Chapter 5

Meta-analysis

5.1 Introduction

Functional neuroimaging experiments enable us to investigate the relationship between particular brain activations and cognitive tasks performed by individual test subjects as well as groups of subjects or patients. However, despite the rapidly growing number of imaging studies, the goal of obtaining a clear and comprehensive understanding of cognitive brain function is still far beyond reach, not least because functional imaging experiments convey a rather variable picture of the brain at work. Often, considerable differences can be observed in the results of even closely related imaging experiments, not least due to the high variability of the observed data that was discussed in the previous chapters. The analysis of the consistency and convergence of results across experiments is therefore a crucial prerequisite for correct generalizations about human brain functions. This calls for analysis techniques on a meta-level, i.e. methods that facilitate the post-hoc combination of results from independently performed studies.

Sophisticated spatial normalization and registration methods, a standardized coordinate system, and the increasing availability of experimental results in publicly accessible neuroimaging databases have opened up the exciting possibility to compare and integrate results from a large number of studies independently performed in laboratories around the globe. An excellent overview of some of the first meta analyses in the context of functional neuroimaging can be found in Fox et al. (1998). More recently performed meta-analyses were published, for example, by Indefrey and Levelt (2000); Phan et al. (2002); Ullsperger and von Cramon (2004); Ridderinkhof et al. (2004); Laird et al. (2005c); Derrfuss et al. (2005) and Smith et al. (2009).

87

Despite the increasing number and significance of meta-analyses in functional imaging, there only exist a few formal mathematical tools for the qualitative and quantitative evaluation of meta-analytic data. In the most simple case, a meta-analysis consists of plotting activation coordinates from a number of experiments into a single figure. Visual inspection is then used to detect anatomical regions that show significant activation in all or the majority of the included studies. Unfortunately, this methodology introduces a high degree of subjectivity into the interpretation of the findings.

In this chapter, we present two new methods for the *coordinate-based meta-analysis* of functional imaging data. In coordinate-based meta-analyses activation coordinates reported from individual imaging experiments are analyzed in search for brain regions that are likely to be most relevant for the investigated cognitive function. With our first method presented in Section 5.3 we propose to apply a combination of *activation likelihood estimation* (ALE) and *model-based clustering* to this problem. The former is a form of kernel density estimation which was recently adapted for the analysis of functional imaging data (Turkeltaub et al., 2002; Chein et al., 2002). The latter provides a general framework for finding groups in data by formulating a clustering problem in terms of the estimation of parameters in a finite mixture of probability distributions (Everitt et al., 2001; Fraley and Raftery, 2002).

With our second meta-analysis technique described in Section 5.4 we address an increasingly important aspect of functional imaging research, namely the search for *functional networks* of brain regions related to a specific cognitive task. That is, we are interested in finding brain areas that jointly show significant activation for a particular cognitive task and can therefore be thought of as belonging to the network of regions required for solving the task in question. Finding such networks of co-activated brain regions is by no means trivial. The number of co-occurrences of two or three activated regions in the investigated studies might still be detectable upon visual inspection, for example by color coding or labelling of activated regions. However, as the potential number of activated areas within the same functional network increases, the process of finding co-occurrences quickly becomes too complex to manage by visual inspection only.

In Section 3.1 we introduced replicator dynamics as a tool for finding groups or networks of highly correlated fMRI time series. In Section 5.4 we propose to use replicator dynamics in the search for cortical areas that jointly show significant activation across a number of studies. The method takes as input activated regions derived

from ALE or the combination of ALE and model-based clustering presented in Sections 5.2 and 5.3. Replicator dynamics is then applied to a so-called co-occurrence matrix, which stores the number of studies for which pairs of the obtained cortical regions have been found co-activated. The dominant network resulting from the replicator process should represent the group of regions which are likely to play a critical role in processing the investigated cognitive task.

In the following section we describe activation likelihood estimation as prerequisite for our two new meta-analysis techniques. In Section 5.3 we present model-based clustering and its application to data set 5 in a meta-analysis of 26 Stroop experiments (cf. Section 2.3.5). In Section 5.4 our network analysis technique based on replicator dynamics is introduced and exemplified using data set 4. This data set contains 15 fMRI experiments again addressing the Stroop interference task (cf. Section 2.3.4).

5.2 Activation likelihood estimation (ALE)

Activation likelihood estimation, developed concurrently but independently by Chein et al. (2002) and Turkeltaub et al. (2002), was among the first methods aimed at modeling activated cortical areas or so-called *activation foci* from meta-analytic imaging data. It was recently extended by Laird et al. (2005a) and Eickhoff et al. (2009) to account for multiple comparisons and random-effects analyses and to enable statistical comparisons between two or more meta-analyses. For the meta-analyses presented in this chapter, ALE was implemented in C as part of the software package LIPSIA (Lohmann et al., 2001).

The general approach of ALE is depicted in Figure 5.1. Activation maxima, reported as results of an imaging experiment, are modeled by 3-dimensional Gaussian probability distributions centered at their Talairach coordinates (Figure 5.1, left). Specifically, the probability that a given activation maximum lies within a particular voxel is

$$p = \frac{1}{(2\pi)^{3/2}\,\sigma^3}\,\exp\left[\frac{-d^2}{2\,\sigma^2}\right]\,, \tag{5.1}$$

where σ is the standard deviation of the distribution and d is the Euclidean distance of the voxel to the activation maximum. For each voxel, the union of these probabilities calculated for all activation maxima yields the activation likelihood estimate (ALE)[1]

---

[1]Note that both the method 'activation likelihood estimation' and the actual estimate are typically abbreviated as 'ALE'.
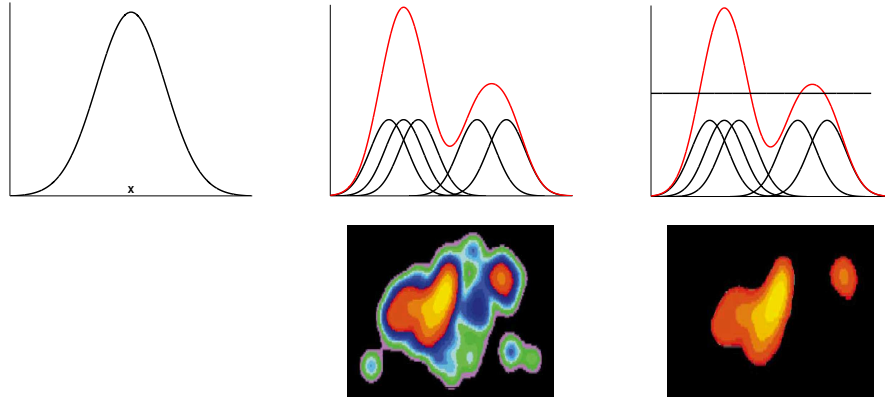
Figure 5.1: Three steps of the ALE procedure in 1D and 2D view (bottom row adapted from Chein et al. (2002)). Activation maxima are modeled by 3-dimensional Gaussian probability distributions (left). The union of the distributions yields ALE maps (middle). A permutation test is employed to determine an appropriate ALE threshold that marks significant activation across experiments (right).

(Figure 5.1, middle). That is, given two likelihood estimates p_a and p_b, representing the likelihood that activation coordinates a and b lie within a particular voxel x, respectively, the ALE representing the likelihood that at least one of the two coordinates lies within this voxel x can be calculated as

$$\mathrm{ALE}_x = p_a + p_b - p_a p_b = p_a + p_b(1 - p_a) \,.$$

In regions with a relatively high density of reported activation maxima, voxels are assigned a high ALE value in contrast to regions where only few and widely spaced activation maxima were reported.

From the resulting ALE maps, one can infer whether activation maxima reported from different experiments are likely to represent the same functional activation. A non-parametric permutation test is utilized to test against the null hypothesis that the activation maxima are spread uniformly throughout the brain (Turkeltaub et al., 2002). Given some desired level of significance α, ALE maps are thresholded at the $100(1-\alpha)$th percentile of the null distribution. Topologically connected voxels with significant ALE values are then considered distinct activated functional regions (Figure 5.1, right).

The extent and separability of the resulting regions critically depend on the choice of σ in Equation (5.1). This is depicted in Figure 5.2, where ALE was applied to data set 5, using both $\sigma = 4\,\text{mm}$ (top) and $\sigma = 5\,\text{mm}$ (bottom). The former resulted in 17 ALE regions, covering $31{,}402\,\text{mm}^3$ and 167 of the input coordinates. The latter yielded only 13 ALE regions, covering $54{,}810\,\text{mm}^3$ and 210 activation coordinates. Thus, decreasing σ leads to smaller regions of significant voxels and to an increase in the number of discrete above-threshold regions which, however, represent only few of the original activation maxima. Increasing σ has the opposite effect with larger regions representing more of the original data.
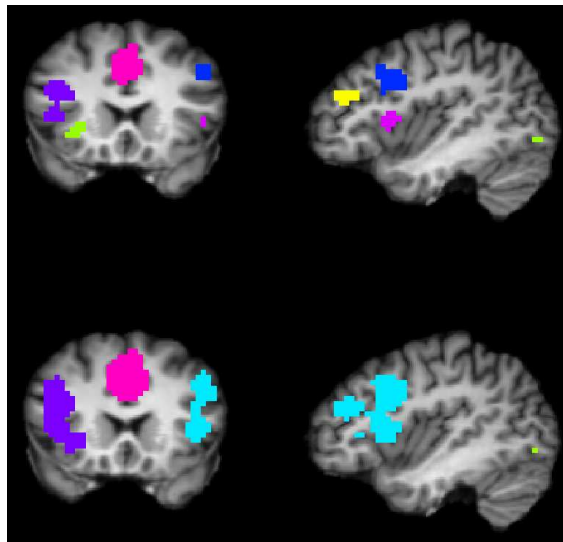
Figure 5.2: Example ALE regions obtained for $\sigma = 4\,\text{mm}$ (top) and $\sigma = 5\,\text{mm}$ (bottom) shown in a sagittal ($x = 40$) and coronal ($y = 17$) slice. Results differ widely, in particular in the right prefrontal cortex, where 3 distinct ALE regions obtained for $\sigma = 4\,\text{mm}$ are no longer separable for $\sigma = 5\,\text{mm}$.

Commonly, σ is chosen to correspond to the size of spatial filters typically applied to fMRI data. In most published ALE analyses (see Fox et al. (2005b) for some examples) σ varies between 4 and 6 mm, with the vast majority of studies using 4.24 mm (10 mm FWHM). However, this value is still chosen more or less arbitrarily.

The strong dependency of the results on the standard deviation of the Gaussian is a major drawback of ALE. In general, choosing the standard deviation too small results

in many small activation foci which cover only a small part of the original input data and do not carry significantly more information than provided by the individual activation maxima alone. In contrast, using a large standard deviation results in activation foci which represent more of the original activation maxima. However, the size of such foci can by far exceed the extent of corresponding activations typically found in single fMRI studies. Such ALE foci might thus comprise more than one functional unit. This can be observed, in particular, in studies with a very inhomogeneous distribution of activation coordinates. In this case a more adaptive or a hierarchical approach would be desirable.

With the method described in the following section, we propose to alleviate this problem by first applying ALE to the original data and then subjecting activation maxima lying within the resulting activation foci to further clustering. Using a large standard deviation of the Gaussian in the first step yields a new set of activation maxima from which coordinates with no other activation maxima in their vicinity are removed. The subsequent model-based clustering then explores the statistical distribution of the remaining coordinates.

5.3   Model-based clustering

5.3.1   The principle of model-based clustering

Model-based clustering in general assumes that the observed data are generated by a finite mixture of probability distributions. Each probability distribution corresponds to a cluster or component. In our application, coordinates surviving activation likelihood estimation are subjected to subsequent clustering based on such finite mixtures.

Our particular implementation closely follows the general model-based clustering approach proposed by Fraley and Raftery (2002). This approach considers mixtures of multivariate Gaussians. Maximum likelihood estimation of the mixture models is performed via the expectation-maximization (EM) algorithm (Hartley, 1958; Dempster et al., 1977), which determines the parameters of the mixture components as well as the posterior probability for a data point to belong to a specific component or cluster. Since a suitable initialization is critical in the successful application of EM, hierarchical agglomerative clustering is performed as an initializing step.

Varying the parameterization of the covariance matrix of a Gaussian mixture provides a set of models with different geometric characteristics, reaching from spherical clusters of equal shape and volume to ellipsoidal clusters with variable shape, volume, and orientation (Banfield and Raftery, 1993). We propose a set of 10 different parameterizations. The best parameterization of the model and the optimal number of clusters are determined using the Bayesian information criterion (BIC) (Schwarz, 1978).

In the following, the individual parts of the clustering procedure are described in detail. These parts were implemented for our application using the software package MCLUST (Fraley and Raftery, 1999, 2003).

5.3.2 Gaussian mixture models

For n independent multivariate observations $\mathbf{x} = (x_1, \ldots, x_n)$, the likelihood of a mixture model with M clusters can be written as

$$\mathcal{L}\left(\theta_1, \ldots, \theta_M, \mathbf{p} \,|\, \mathbf{x}\right) = \prod_{i=1}^{n} \sum_{k=1}^{M} p_k \, f_k(x_i| \,\theta_k)\,, \tag{5.2}$$

where f_k is the density of the cluster k with parameter vector θ_k, and $\mathbf{p} = (p_1, \ldots, p_M)$ is the vector of mixing proportions with $p_k \geq 0$ and $\sum_{k=1}^{M} p_k = 1$. Since any distribution can be effectively approximated by a mixture of Gaussians (Silverman, 1985; Scott, 1992), the probability density function is most commonly represented by

$$f_k(x_i|\mu_k, \mathbf{\Sigma_k}) \quad = \quad \frac{1}{\sqrt{(2\pi)^d|\mathbf{\Sigma_k}|}} \, exp\left(-\frac{1}{2}(x_i - \mu_k)^{\mathbf{T}}(\mathbf{\Sigma_k})^{-1}(x_i - \mu_k)\right) \tag{5.3}$$

for d-dimensional data with mean μ_k and covariance matrix $\mathbf{\Sigma_k}$. Geometrical features of the clusters can be varied by parameterization of the covariance matrices $\mathbf{\Sigma_k}$. Banfield and Raftery (1993) suggest various parameterizations through the eigenvalue decomposition

$$\mathbf{\Sigma_k} = \lambda_k \mathbf{D_k A_k D_k^T}\,. \tag{5.4}$$

$\mathbf{D_k}$ is the matrix of eigenvectors, $\mathbf{A_k}$ is a diagonal matrix with elements that are proportional to the eigenvalues of $\mathbf{\Sigma_k}$ such that $|\mathbf{A_k}| = \mathbf{1}$, and λ_k is a scalar. Treating $\mathbf{D_k}, \lambda_k$, and $\mathbf{A_k}$ as independent parameters and keeping them either constant or variable across clusters varies the shape, volume, and orientation of the clusters. In the simplest case $\mathbf{\Sigma_k} = \lambda\,\mathbf{I}$, all clusters are spherical and of equal size. The least constraint case given in Equation (5.4) accounts for ellipsoidal clusters of variable shape,

Parameterization	Clusters		
	Shape	Volume	Orientation
a) $\boldsymbol{\Sigma_k} = \lambda\mathbf{I}$	equal	equal	–
b) $\boldsymbol{\Sigma_k} = \lambda_k\mathbf{I}$	equal	variable	–
c) $\boldsymbol{\Sigma_k} = \lambda\mathbf{A}$	equal	equal	along coordinate axes
d) $\boldsymbol{\Sigma_k} = \lambda_k\mathbf{A}$	equal	variable	along coordinate axes
e) $\boldsymbol{\Sigma_k} = \lambda\mathbf{A_k}$	variable	equal	along coordinate axes
f) $\boldsymbol{\Sigma_k} = \lambda_k\mathbf{A_k}$	variable	variable	along coordinate axes
g) $\boldsymbol{\Sigma_k} = \lambda\mathbf{DAD^T}$	equal	equal	equal
h) $\boldsymbol{\Sigma_k} = \lambda\mathbf{D_k AD_k^T}$	equal	equal	variable
i) $\boldsymbol{\Sigma_k} = \lambda_k\mathbf{D_k AD_k^T}$	equal	variable	variable
k) $\boldsymbol{\Sigma_k} = \lambda_k\mathbf{D_k A_k D_k^T}$	variable	variable	variable

Table 5.1: Different parameterizations of the covariance matrices applied in model-based clustering. It can be seen that $\mathbf{A_k}$ accounts for the shape of a cluster, λ_k for its volume, and $\mathbf{D_k}$ for its orientation. Table adapted from Fraley and Raftery (1999, Table 1).

volume, and orientation. All parameterizations applied to our experimental data are presented in Table 5.1. The first two models have spherical, all other models have ellipsoidal clusters, whereby clusters in models with diagonal covariance matrices (c-f) are oriented along the coordinate axes. Models with identical matrix $\mathbf{A}$ for all clusters have equally shaped clusters, whereas models with identical λ for all clusters have clusters of the same volume.

5.3.3   Maximum likelihood estimation

Maximum likelihood estimation of a Gaussian mixture model as defined in Equations (5.2) and (5.3) can be performed via the widely used EM algorithm which provides a general approach to parameter estimation in incomplete data problems (Hartley, 1958; Dempster et al., 1977; Neal and Hinton, 1998). Given a likelihood function $\mathcal{L}(\theta|\mathbf{y}) = \prod_{i=1}^{n} f(y_i|\theta)$, for parameters θ and data $\mathbf{y} = (y_1, \ldots, y_n)$, we wish to find $\hat{\theta}$ such that

$$\hat{\theta} = \arg\max_{\theta} \mathcal{L}(\theta|\mathbf{y}).$$

In the presence of some hidden data $\mathbf{z}$ such that $\mathbf{y} = (\mathbf{x}, \mathbf{z})$ with $\mathbf{x}$ observed and $\mathbf{z}$ unobserved, we can equivalently maximize the so-called complete-data log likelihood and find $\hat{\theta}$ such that

$$\hat{\theta} = \arg\max_{\theta} \log \mathcal{L}\left(\theta, \mathbf{z} \,|\, \mathbf{x}\right).$$

Starting from an initial guess, the EM algorithm proceeds by alternately estimating the unobservable data $\mathbf{z}$ and the unknown parameters θ. Specifically, in the E-step, the algorithm calculates the expected value of the complete-data log likelihood with respect to $\mathbf{z}$ given $\mathbf{x}$ and the current estimate of θ. In the M-step, this expected value is maximized in terms of θ, keeping $\mathbf{z}$ fixed as computed in the previous E-step.

In our application, the complete data $\mathbf{y} = (y_1, \ldots, y_n)$, consists of $y_i = (x_i, z_i)$ where each x_i is a 3-dimensional vector containing coordinates of activation maxima in Talairach space and $z_i = (z_{i1}, \ldots, z_{iM})$ is the unknown membership of x_i in one of the M clusters, i.e.

$$z_{ik} = \begin{cases} 1 & x_i \text{ belongs to cluster } k \\ 0 & \text{otherwise}. \end{cases}$$

With the density of observation x_i given z_i as $\prod_{k=1}^{M} f_k(x_i | \mu_k, \mathbf{\Sigma_k})^{z_{ik}}$, the complete-data log likelihood in our problem can be formulated as

$$\ell\left(\mu_k, \mathbf{\Sigma_k}, p_k, z_{ik} \,|\, \mathbf{y}\right) = \sum_{i=1}^{n} \sum_{k=1}^{M} z_{ik} \log\left[p_k \, f_k(x_i | \mu_k, \mathbf{\Sigma_k})\right] \tag{5.5}$$

assuming that each z_i is independently and identically distributed according to a multinomial distribution of one draw from M categories with probabilities $p_1, \ldots, p_M$ (Fraley and Raftery, 1998).

Maximum likelihood estimation is performed by alternating between the calculation of z_{ik} given x_i, μ_k, and $\mathbf{\Sigma_k}$ (E-step) and maximizing Equation (5.5) with respect to μ_k, $\mathbf{\Sigma_k}$, and p_k with z_{ik} fixed (M-step):

E-step:

$$z_{ik} \leftarrow \frac{p_k \, f_k(x_i | \mu_k, \mathbf{\Sigma_k})}{\sum_{j=1}^{M} p_j \, f_j(x_i | \mu_j, \mathbf{\Sigma_j})}.$$

M-step:

$$p_k \leftarrow \frac{n_k}{n} \qquad \mu_k = \frac{1}{n_k} \sum_{i=1}^{n} z_{ik} \, x_i$$

with

$$n_k \leftarrow \sum_{i=1}^{n} z_{ik} \, .$$

The calculation of $\mathbf{\Sigma_k}$ in the M-step depends on the parameterization and differs for the investigated models. Let $\mathbf{W_k}$ be the within-cluster scattering matrix of cluster k

$$\mathbf{W_k} = \sum_{i=1}^{n} z_{ik}(x_i - \mu_k)(x_i - \mu_k)^{\mathbf{T}} \tag{5.6}$$

and

$$\mathbf{W} = \sum_{k=1}^{M} \mathbf{W_k} \, . \tag{5.7}$$

Then, the covariance matrices of the densities are calculated as follows (for details see Celeux and Govaert (1995)).

a) Model $\mathbf{\Sigma_k} = \lambda \mathbf{I}$:

$$\lambda = \frac{\text{trace}(\mathbf{W})}{d\,n} \, .$$

b) Model $\mathbf{\Sigma_k} = \lambda_k \mathbf{I}$:

$$\lambda_k = \frac{\text{trace}(\mathbf{W_k})}{d\,n_k} \, .$$

c) Model $\mathbf{\Sigma_k} = \lambda \mathbf{A}$:

$$\mathbf{A} = \frac{\text{diag}(\mathbf{W})}{|\text{diag}(\mathbf{W})|^{\frac{1}{d}}} \qquad \lambda = \frac{|\text{diag}(\mathbf{W})|^{\frac{1}{d}}}{n} \, .$$

d) Model $\mathbf{\Sigma_k} = \lambda_k \mathbf{A}$:

$$\mathbf{A} = \frac{\text{diag}(\sum_{k=1}^{M} \frac{\mathbf{W_k}}{\lambda_k})}{|\text{diag}(\sum_{k=1}^{M} \frac{\mathbf{W_k}}{\lambda_k})|^{\frac{1}{d}}} \qquad \lambda_k = \frac{\text{trace}(\mathbf{W_k}\mathbf{A}^{-1})}{dn_k} \, .$$

e) Model $\mathbf{\Sigma_k} = \lambda \mathbf{A_k}$:

$$\mathbf{A_k} = \frac{\text{diag}(\mathbf{W_k})}{|\text{diag}(\mathbf{W_k})|^{\frac{1}{d}}} \qquad \lambda = \frac{\sum_{k=1}^{M} |\text{diag}(\mathbf{W_k})|^{\frac{1}{d}}}{n} \, .$$

f) Model $\mathbf{\Sigma_k} = \lambda_k \mathbf{A_k}$:

$$\mathbf{A_k} = \frac{\text{diag}(\mathbf{W_k})}{|\text{diag}(\mathbf{W_k})|^{\frac{1}{d}}} \qquad \lambda_k = \frac{|\text{diag}(\mathbf{W_k})|^{\frac{1}{d}}}{n_k} \, .$$

g) Model $\boldsymbol{\Sigma_k} = \lambda \mathbf{DAD^T}$:

$$\boldsymbol{\Sigma_k} = \frac{\mathbf{W}}{n}\,.$$

h) Model $\boldsymbol{\Sigma_k} = \lambda \mathbf{D_k A D_k^T}$:

Given the eigenvalue decomposition $\mathbf{W_k} = \mathbf{L_k}\,\boldsymbol{\Omega_k}\,\mathbf{L_k^T}$ with eigenvalues in $\boldsymbol{\Omega_k}$ in decreasing order,

$$\mathbf{A} = \frac{\sum_{k=1}^{M}\boldsymbol{\Omega_k}}{|\sum_{k=1}^{M}\boldsymbol{\Omega_k}|^{\frac{1}{d}}} \qquad \lambda = \frac{|\sum_{k=1}^{M}\boldsymbol{\Omega_k}|^{\frac{1}{d}}}{n}\,.$$

i) Model $\boldsymbol{\Sigma_k} = \lambda_k \mathbf{D_k A D_k^T}$:

Given the eigenvalue decomposition $\mathbf{W_k} = \mathbf{L_k}\,\boldsymbol{\Omega_k}\,\mathbf{L_k^T}$ with eigenvalues in $\boldsymbol{\Omega_k}$ in decreasing order,

$$\mathbf{A} = \frac{\sum_{k=1}^{M}\frac{\boldsymbol{\Omega_k}}{\lambda_k}}{|\sum_{k=1}^{M}\frac{\boldsymbol{\Omega_k}}{\lambda_k}|^{\frac{1}{d}}} \qquad \lambda = \frac{\text{trace}(\mathbf{W_k}\,\mathbf{L_k}\,\mathbf{A^{-1}}\,\mathbf{D_k^T})}{d\,n_k}\,.$$

k) Model $\boldsymbol{\Sigma_k} = \lambda_k \mathbf{D_k A_k D_k^T}$:

$$\boldsymbol{\Sigma_k} = \frac{1}{n_k}\,\mathbf{W_k}\,.$$

Note that in models (d) and (i), estimation of the covariance matrix has to be performed iteratively.

The procedure of alternating between E- and M-step is terminated after the relative difference between successive values of $\ell\left(\mu_k, \boldsymbol{\Sigma_k}, p_k, z_{ik} | \mathbf{y}\right)$ falls below some threshold ϵ, which in our application was set to $\epsilon = 0.00001$. The value of z_{ik} at the maximum of Equation (5.5) is the estimated probability that x_i belongs to cluster k, and the maximum likelihood classification of x_i is the cluster k, with

$$k = \arg\max_k\ z_{ik} \qquad \text{for} \quad k = 1, \ldots, M\,.$$

5.3.4 Hierarchical agglomeration

Following the suggestion by Fraley and Raftery (1998), we employ model-based hierarchical agglomeration for initialization. This method tends to yield reasonable clusterings in the absence of any information about a possible clustering inherent in the data (Fraley and Raftery, 2002).

Hierarchical agglomeration techniques typically start with a pre-defined number of clusters and in each step merge the two closest clusters into a new cluster, thereby reducing the number of clusters by one. The implementation used here starts with n clusters, each containing a single observation x_i. Then, two clusters are chosen such that merging them increases the so-called *classification likelihood*,

$$\mathcal{C}\left(\theta_1, \ldots, \theta_M, \mathbf{c} \,|\, \mathbf{x}\right) = \prod_{i=1}^{n} f_{c_i}(x_i \,|\, \theta_{c_i}), \qquad (5.8)$$

with $f_k(x_i)$ given in Equation (5.3). The vector $\mathbf{c} = (c_1, \ldots, c_n)$ encodes the classification of the data, i.e. $c_i = k$ if x_i is classified as member of cluster k. For an unrestricted covariance matrix as defined in Equation (5.4), approximately maximizing the classification likelihood (5.8) amounts to minimizing

$$\sum_{k=1}^{M} n_k \, \log |\frac{\mathbf{W_k}}{n_k}|,$$

where n_k is the number of elements in cluster k, and $\mathbf{W_k}$ is the within-cluster scattering matrix of cluster k as defined in Equation (5.6). Computational issues on this clustering procedure are discussed in detail by Banfield and Raftery (1993) and Fraley (1998), in particular regarding the initial stages with a single data point in each cluster which leads to $|\mathbf{W}| = 0$.

From the values of $\mathbf{c}$ at the maximum of $\mathcal{C}$, initializations for the unknown membership values z_{ik} are derived, and first estimates for the parameters of the Gaussian components can be obtained from an M-step of the EM algorithm as described above.

5.3.5   Model selection via BIC

A problem of most clustering techniques is to determine the number of clusters inherent in the data. One common technique in model-based clustering is to apply several models with different pre-defined numbers of clusters and subsequently choose the best model according to some model selection criterion. For models with equal number of parameters, the simplest approach is to compare estimated residual variances. This is not applicable, however, when models with varying number of parameters are considered.

An advantage of using mixture models for clustering is that approximate Bayes factors can be used for model selection. Bayes factors were developed originally as

a Bayesian approach to hypothesis testing by Jeffreys (1935, 1961). In the context of model comparison, a Bayes factor describes the posterior odds for one model against another given equal prior probabilities. It is determined from the ratio of the integrated likelihoods of the models. In conjunction with EM for maximum likelihood estimation, the integrated likelihood of a model can be approximated under certain regularity conditions by the Bayesian information criterion (Schwarz, 1978), which is defined[2] as

$$\text{BIC} := 2\,\hat{\ell}\,(\theta|\,\mathbf{x}) - m\,\log(n), \tag{5.9}$$

where $\hat{\ell}$ is the maximized mixture log likelihood of the model, m is the number of independent parameters of the model, and n the number of data points. With this definition, a large BIC value provides strong evidence for a model and the associated number of clusters.

The relationship between Bayes factors and BIC, the regularity conditions, and the use of Bayes factors for model comparison are discussed in detail, e.g., by Kass and Raftery (1995). They also provide guidelines for the strength of evidence for or against some model: A difference of less than 2 between the BIC of two models corresponds to weak, a difference between 2 and 6 to positive, between 6 and 10 to strong, and a difference greater than 10 to very strong evidence for the model with the higher BIC value.

5.3.6   The algorithm

Taking together the individual parts described above, our algorithm for deriving activated functional regions from meta-analytic imaging data can be summarized as follows:

1. Given a list of coordinates encoding activation maxima in Talairach space from a number of individual studies, calculate ALEs for all voxels using a large standard deviation of the Gaussian. Determine those coordinates that fall within the regions above the ALE threshold.

2. Define a maximum number of clusters M. Perform hierarchical agglomeration for up to M clusters using the reduced coordinate list obtained in Step 1 as

---

[2]Note that in some publications, BIC is defined as $\text{BIC} := -2\,\hat{\ell}\,(\theta|\,\mathbf{x}) + m\,\log(n)$ with smaller BIC values indicating stronger evidence for a model.

input, thereby approximately maximizing the classification likelihood as defined in Equation (5.8).

3. For each parameterization and number of clusters of the model as defined in Equation (5.5) perform EM, using the classification obtained in Step 2 as initialization.

4. Calculate the BIC for each parameterization and number of clusters in the model according to Equation (5.9).

5. Choose the parameterization and number of clusters with a decisive maximum BIC value as solution according to the guidelines above.

5.3.7 Application

Our method was applied to data set 5 described in Section 2.3.5 which contained results from 26 fMRI experiments employing the Stroop paradigm (Stroop, 1935). This paradigm, already used in data set 1, was again chosen as a test case, because the interference effect has been investigated in a large number of functional imaging studies and is known to produce very reliable cortical activation. Activation is most commonly reported in the left inferior frontal region, the left inferior parietal region, and the left and right anterior cingulate (McKeown et al., 1998a; Banich et al., 2000; Liu et al., 2004). Our own network analysis presented in Section 5.4 revealed a frontal network including the presupplementary motor area (preSMA), the inferior frontal sulcus, the anterior cingulate cortex (ACC) of both hemispheres, and the inferior frontal junction area (IFJ). Other frequently reported areas include frontopolar cortex, occipital cortex, fusiform gyrus, and insula (Laird et al., 2005c; Zysset et al., 2001).

Despite the high agreement in the reported activated areas, the actual location of associated coordinates in Talairach space differs widely between studies. For example, the left IFJ was localized in previous studies at Talairach coordinates x between -47 and -35, y between -4 and 10, and z between 27 and 40 (Brass et al., 2005; Derrfuss et al., 2004, 2005; Neumann et al., 2005). Such high variability makes the classification of the data into distinct functional units difficult.

Plots of all coordinates in data set 5 projected onto a single axial, sagittal and coronal slice are shown in the top row of Figure 5.3. Coordinates reported from different studies are represented by different colors. As can be seen, activation maxima are dis-

tributed over large parts of the cortex, although some areas with a higher density of activation coordinates are already apparent, in particular in the left lateral prefrontal cortex and the medial frontal cortex. These can be seen more clearly in the example slices in the bottom row of Figure 5.3.

(a)

(b)

$$z = 37 \qquad x = 1 \qquad x = -40$$
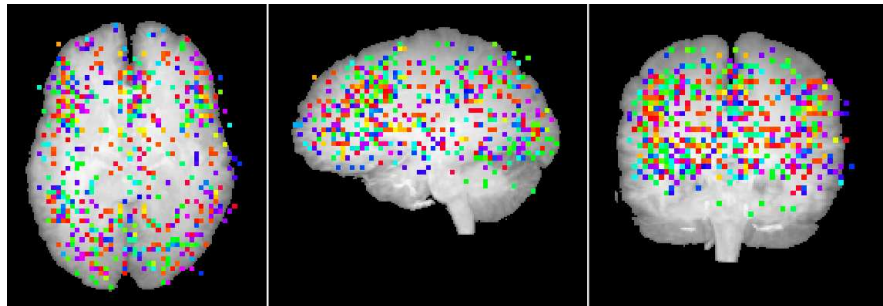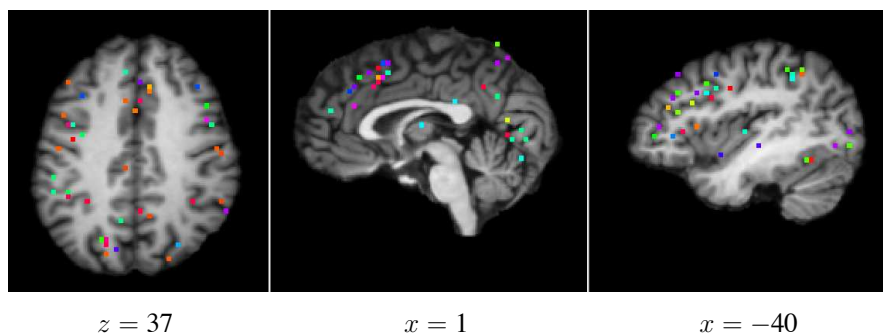
Figure 5.3: (a) 728 activation coordinates in data set 5 projected onto three orthogonal single coronal, sagittal, and axial slices. (b) Three example slices showing activation coordinates projected onto an individual brain. Slices were chosen to show cortical areas which are frequently reported as significantly activated in the Stroop task (ACC, IFJ, preSMA). Activation coordinates from the same study are plotted in the same color.

All activation coordinates were first subjected to an ALE analysis with standard deviation of $\sigma = 5$ mm, corresponding to 11.8 mm FWHM. The null distribution was derived from 1,000 iterations of randomly placing 728 activation coordinates over a mask brain volume defined by the minimum and maximum Talairach coordinates in the orig-

<div align="center">$z = 37$ $x = 1$ $x = -40$</div>
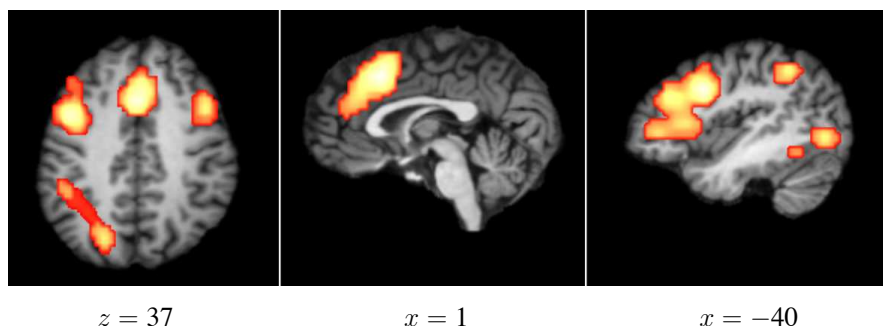
Figure 5.4: ALE map derived from the 728 activation coordinates. The map was thresholded at $\alpha = 0.01\%$ corresponding to an ALE value of 0.0156 (maximum ALE value $ALE_{max} = 0.049$). Axial and sagittal slices correspond to the example slices shown in Figure 5.3 (b).

inal data set. The brain mask spanned a volume of 61,408 voxels, each $3 \times 3 \times 3\,\text{mm}^3$ in size. As suggested by Turkeltaub et al. (2002), the resulting ALE map was thresholded at an α-level of $\alpha = 0.01\%$. This corresponded to an ALE threshold of 0.0156. Figure 5.4 shows sagittal and axial example slices of the ALE map containing only voxels above threshold.

The analysis yielded 13 regions of topologically connected voxels above threshold which covered a total volume of $54{,}810\,\text{mm}^3$ and contained 210 of the original activation maxima. Table 5.2 shows size, maximum ALE value, location of the center in Talairach space, and the number of original activation coordinates covered by the detected ALE regions.

Note that the four largest regions cover 89.65% ($49{,}140\,\text{mm}^3$) of the total ALE regions' volume. They contain 83.8% of all above-threshold coordinates. This can be explained by the very inhomogeneous distribution of the original input coordinates: More than 40% of the original activation maxima fell within regions spanned by the minimum and maximum Talairach coordinates of the four largest ALE regions. The remaining coordinates were distributed more evenly over other parts of the cortex.

Note further that some smaller regions surviving the ALE threshold contain only single activation maxima. This seems counterintuitive at first, as a single coordinate should not result in a relatively high ALE value. However, one could imagine, for example, a situation where three coordinates are arranged in a 'row', i.e. at three voxels in the same row of a slice with one voxel between them. The voxel in the middle will get a higher ALE value than the ones at both ends, as it has two other coordinates

	volume	max ALE			location			# of coords
1	19,494	0.05	L	(	−44	6	33)	66
2	13,716	0.05	R	(	1	18	39)	49
3	9,882	0.04	R	(	43	9	30)	36
4	6,048	0.03	L	(	−41	−51	45)	25
5	3,105	0.03	L	(	−38	−72	3)	16
6	1,134	0.02	L	(	−47	−54	−3)	7
7	297	0.02	R	(	49	−45	30)	3
8	324	0.02	L	(	−5	36	−3)	2
9	297	0.02	R	(	46	−51	−6)	1
10	189	0.02	R	(	10	−60	15)	2
11	162	0.02	R	(	7	−75	−9)	1
12	81	0.02	R	(	19	48	21)	1
13	81	0.02	R	(	37	−72	−3)	1

Table 5.2: ALE regions and number of activation coordinates falling within these regions as result of the ALE analysis of 728 activation maxima in data set 5. Regions are ordered by size.

in close proximity (only two voxels away) whereas the other two voxels have one coordinate in close distance and another one four voxels further away. Depending on the distribution of other coordinates, thresholding the ALE values could now shape the surviving ALE region such that only the coordinate in the middle will be inside the region, whereas the values at the other two voxels might just be too small to survive the thresholding. Thus, ALE regions containing only a single coordinate are caused by very small groups of activation maxima that are quite isolated from the remaining ones. The fact that some of our ALE regions contain only a single coordinate indicates that all remaining activation coordinates, not surviving the thresholding, are very isolated from each other. They can therefore be regarded as noise.

Despite the use of a very small α-level in ALE thresholding, some of the determined ALE foci clearly exceed the size of cortical activations typically found in these regions for the Stroop paradigm, see e.g. Zysset et al. (2001) for a comparison. Moreover, as seen in Figure 5.4, within such foci, in particular in the left prefrontal cortex, sub-maxima of ALE values are visible, indicating a possible sub-clustering of the rep-

resented activation coordinates. We therefore subjected all above-threshold activation coordinates to model-based clustering as described above.

Hierarchical agglomeration of the coordinates was first performed for up to 30 clusters. Using the results as initialization for the EM algorithm, models as defined in Equation (5.5) with the parameterizations introduced in Section 5.3 with up to 30 clusters were then applied to the data set, and BIC values were calculated for each number of clusters and parameterization.

The three models with $\lambda_k = \lambda$, i.e. models with clusters of equal volume, out-performed the remaining models which all allowed for clusters of variable volume. This seems counterintuitive at first, as a more variable model would be expected to fit the data better than a more restricted one. However, as described above, the BIC value penalizes model complexity, which is larger for models with variable clusters than for models with equal clusters. Thus, for our data, allowing the clusters' volume to vary did not increase the log likelihood of the models sufficiently in order to justify the increased number of model parameters. Note also that for very large cluster numbers, some more variable models failed to provide a clustering due to the singularity of the associated covariance matrices. This was not the case for models with fewer free parameters, however.

Figure 5.5 shows plots of the BIC values of the best three models for up to 30 clusters. BIC values of these models are very similar, in particular for models with more than 20 clusters. The right side shows an enlarged plot of the BIC values for models with 20 up to 25 clusters. All three models yielded the highest BIC value when applied with 24 clusters. The more complex models with ellipsoidal clusters slightly outperformed the spherical one, whereby the difference between a variable and a fixed orientation of the clusters was negligible.

Figure 5.6 shows the results of the model-based clustering exemplified for the two largest ALE regions which were situated in the left lateral prefrontal cortex (left LPFC) and the medial frontal cortex (MFC), respectively (cf. Table 5.2). The clustering of activation coordinates within the left LPFC is shown in five consecutive sagittal functional slices at Talairach coordinates between $x = -34$ and $x = -46$. The coordinates in this ALE region were subdivided into five groups in anterior-posterior and superior-inferior direction. In the most posterior and superior part of the region a further division in lateral-medial direction can be observed (shown in green and blue). Interestingly, cluster centers of the more anterior and inferior clusters corresponded
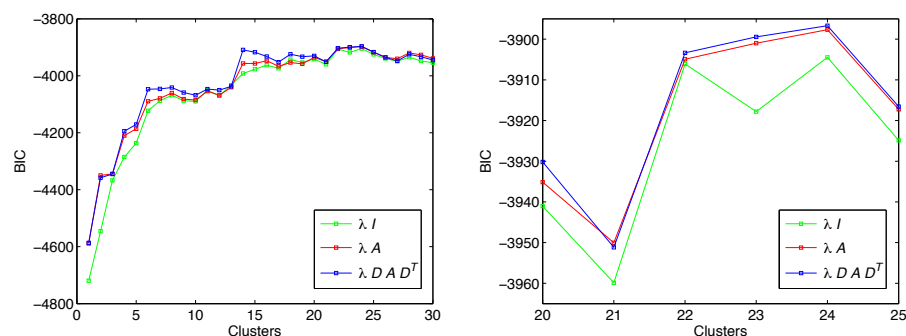
Figure 5.5: Plot of the BIC values of the best three models for up to 30 clusters (left) and enlarged plot of the BIC values for the best three models with cluster numbers between 20 and 25 (right).

closely to the sub-maxima in the ALE focus visible in Figure 5.4. However, the division of posterior and superior parts of the region into two clusters could not have been predicted from the ALE sub-maxima. The same holds for the clustering of coordinates in the MFC, where no sub-maxima could be observed in the ALE map. The clustering of coordinates in the MFC is shown in the right panel of Figure 5.6 in four consecutive sagittal slices. The best model provided four clusters, again dividing the region in anterior-posterior and superior-inferior direction. Thus, model-based clustering revealed some additional structure in the data that would have remained undetected when using ALE alone. To get some feeling for the actual shape of the clusters and their relative location, the extracted clusters are presented again in views from different angles in Figure 5.7.

The robustness of our method against noisy input data was finally tested in a post-hoc analysis including only the 550 activation coordinates that truly represented the Stroop interference effect, i.e. a contrast between the incongruent and the congruent or the neutral condition. The results did not significantly differ from the results of the original analysis. The noise in the original input data thus did not have a noteworthy impact on the results of the model-based clustering.
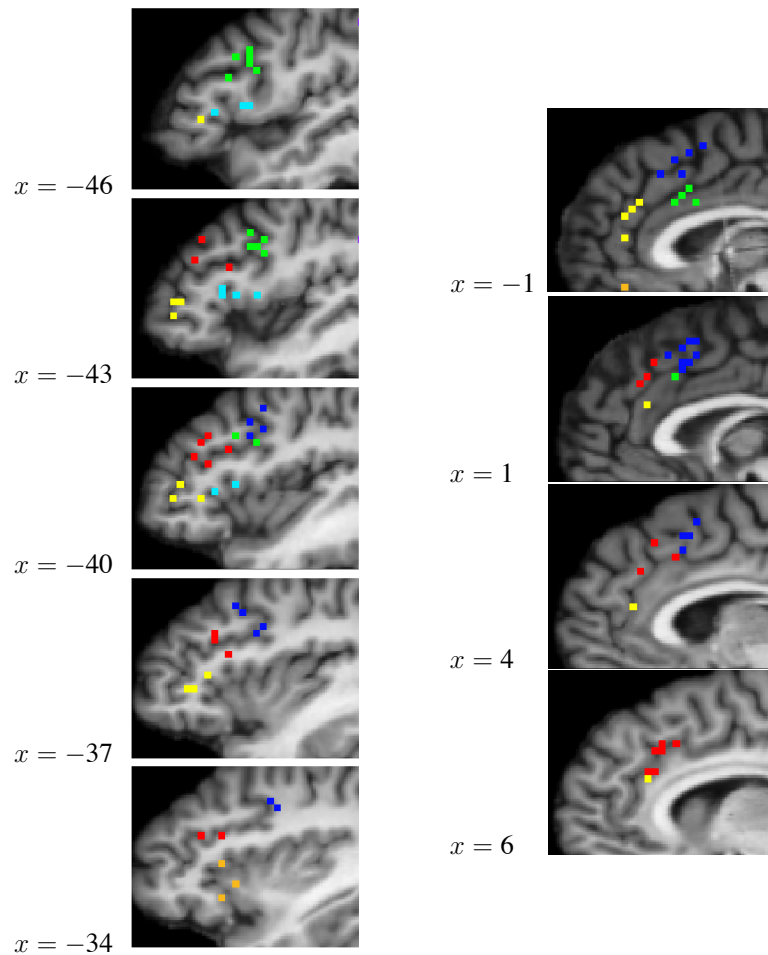
Figure 5.6: Left: Clustering results for the largest ALE region (left LPFC), shown in five consecutive sagittal slices. The clustering yielded five clusters (shown in green, light blue, yellow, red, and blue), dividing the region primarily in anterior-posterior and superior-inferior direction. The most posterior and superior part of the region was further divided in lateral-medial direction. An additional cluster centered around the left insula can be seen in orange at x = −34. These coordinates were not part of the largest ALE region. Right: Results for the second largest ALE region (MFC) shown in four consecutive sagittal slices. Clustering yielded four clusters (blue, green, yellow, red), again dividing the region in anterior-posterior and superior-inferior direction. Note that the single coordinate shown in orange at x = −1 was not part of the second largest ALE region.
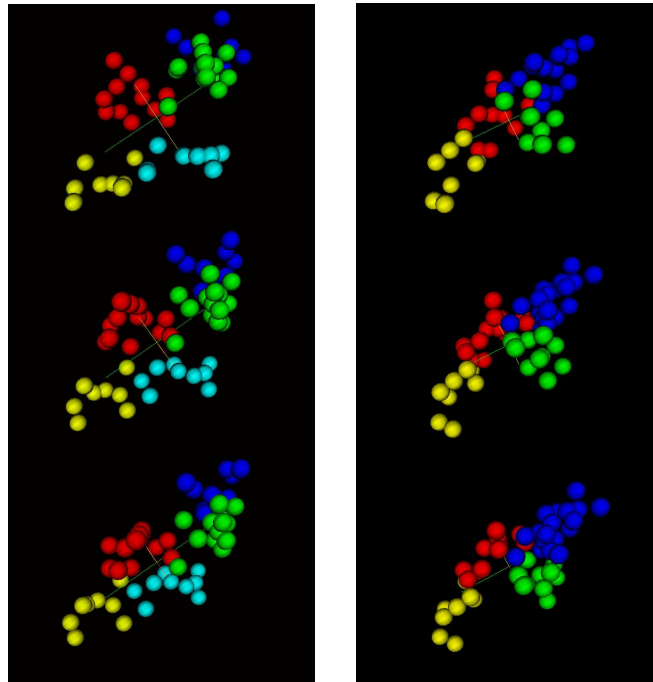
Figure 5.7: Clustering results, on the left for the largest ALE region (left LPFC) and on the right for the second largest ALE region (MFC). Clusters are shown in a sagittal view (top) corresponding to the view in Figure 5.6 and twice turned around the y-axis by a few degrees in order to visualize the shape and separation of the clusters. Colors correspond to the colors in Figure 5.6.

5.3.8 Discussion

Activation likelihood estimation facilitates the detection of cortical activation from activation maxima reported in independently performed functional imaging studies. The resulting areas reflect the distribution of activation maxima over the cortex. In particular, clusters of activation maxima in a region reflect the likely involvement of this region in processing a cognitive task, whereas isolated activation maxima are regarded as noise.

Our analysis shows that the extent of ALE regions can vary considerably due to the heterogeneous distribution of the input data across different parts of the cortex. As seen in Table 5.2 and Figure 5.4, the size of some regions obtained by ALE alone by far exceeded the extent of comparable activations reported in single fMRI experiments. For example, Zysset et al. (2001) reported for the Stroop paradigm two clearly separated activated regions with maxima in the posterior (Talairach coordinates: $-38, 5, 30$) and the anterior (Talairach coordinates: $-38, 35, 5$) inferior frontal sulcus. These coordinates were both located within the same ALE region in our analysis. This was caused by the high number of activation coordinates within this region together with their high spatial variability. Moreover, within the largest ALE focus located in the left LPFC, sub-maxima could be observed, indicating a possible sub-clustering of the region.

One simple way to separate several areas within such a large ALE region would be the choice of a higher ALE threshold. However, this is problematic if a whole-brain analysis is performed, since ALE values in other regions might be significantly lower despite a high concentration of activation coordinates. For example, in Figure 5.4 (b) a cluster of activation coordinates can clearly be seen in the anterior part of the left intraparietal sulcus. However, the resulting ALE focus representing no less than 25 activation coordinates has a maximum ALE value of only 0.027 in comparison to 0.05 in the left LPFC. Thus, by simply choosing a higher ALE threshold, some clusters of activation coordinates might remain undetected.

We tried to alleviate this problem by following a hierarchical approach. In a first step, ALE was used to identify regions with high concentration of activation coordinates. In a second step, large ALE regions were further investigated in search for a possible subdivision. Applying this two-step procedure to activation maxima from 26 Stroop experiments first resulted in relatively large ALE regions, in particular in the frontal lobe (cf. Fig 5.4). This is in line with earlier findings on frontal lobe activity, in particular in a meta-analysis by Duncan and Owen (2000) who reported cortical

regions of large extent to be recruited by a variety of cognitive tasks. However, in contrast to this study, our analysis pointed to a possible further sub-clustering of these areas. The two largest ALE regions found in the left lateral prefrontal cortex and the medial frontal wall were partitioned into five and four clusters, respectively. While our exploratory analysis technique does not have the power to associate specific cognitive functions to these clusters, this finding could serve as a hypothesis for a further functional specialization of these regions.

The main directions of the clustering were in parallel to the coordinate axes, primarily in anterior-posterior and superior-inferior direction. This corresponds well with recent results from single-subject and group analyses obtained from a variety of analysis techniques as well as from other meta-analyses, see e.g. the results of the first-level analysis presented in Section 3.1 as well as Forstmann et al. (2005); Koechlin et al. (2003); Müller et al. (2003) for LPFC, and Forstmann et al. (2005) and Amodio and Frith (2006) for MFC clustering.

It is important to be clear about the implicit assumptions made in the application of our analysis technique. Meta-analyses are aimed at consolidating results from several studies in order to find general mechanisms related to a particular task, class of paradigms, etc. Thus, if we want to generalize the findings of any meta-analysis, we must assume that the data extracted from the included studies are a representative sample of all the data collected for the investigated phenomenon. It should be noted that, while this is a rather strong assumption, it has to be met by any empirical analysis technique relying on sampled data. A second, closely related, assumption specific to clustering activation coordinates is that the inherent distribution of activation for the investigated phenomenon is completely represented by the investigated data.

In a meta-analysis, these assumptions are sometimes hard to meet due to the selective publication of activation coordinates from particular cortical regions, a problem often referred to as 'publication or literature bias'. In the majority of experimental studies, only a specific aspect of a paradigm or a particular cortical region is investigated and, consequently, some significantly activated regions found for a stimulus might be neglected in the publication of the results. This can result in over-emphasizing some regions while neglecting others, which in turn can lead to a non-representative distribution of the input data. A careful and informed selection of studies included in such an analysis and the inclusion of as much data as possible is thus indispensable.

For our example analysis we used a very large data set, in order to minimize the effects of the publication bias. However, our method also works for smaller analyses. For very small numbers of activation maxima, the maximum number of clusters might have to be reduced, in order to avoid singularity problems in the estimation of the covariance matrix. Moreover, for small or very homogeneously distributed data sets, the problem of very large ALE regions might not arise in the first place. In this case, the results of the model-based clustering should not differ significantly from the application of ALE alone.

The clustering technique presented here is purely data-driven. That is, the results are exclusively derived from the spatial distribution of the input data and restricted only by the constraints on the geometry of the mixture model clusters. Here, additional constraints such as anatomical or cytoarchitectonic boundaries between cortical regions are conceivable. How such constraints can be incorporated into the mathematical framework of mixture modelling is a question that should be addressed in future work.

As noted earlier, in ALE the extent and number of above-threshold clusters critically depend on the choice of a suitable standard deviation of the Gaussian. Nielsen and Hansen (2002) offer an interesting approach to this problem by optimizing the standard deviation of a Gaussian kernel when modelling the relation between anatomical labels and corresponding focus locations. Similar to ALE, activation maxima are modelled by 3-dimensional Gaussian probability distributions and the standard deviation is optimized by leave-one-out cross validation (Nielsen and Hansen, 2002). In our hierarchical approach, the choice of σ is less critical and the use of a large standard deviation is feasible, as ALE is used only as a preprocessing step for model-based clustering. We can thus make use of as much information present in the data as possible. Note that in our test data set the use of an even larger standard deviation did not have any effect on the choice of activation coordinates entering the second step of our analysis, although some ALE regions were merged and slightly extended. The results of the model-based clustering for a larger standard deviation would therefore be identical to the results presented here for $\sigma = 5\,\mathrm{mm}$.

A second parameter influencing the outcome of an ALE analysis, is the size of the mask volume used for deriving the null hypothesis. Clearly, the size of the volume has some influence on the ALE threshold corresponding to the desired α-level. Therefore, the mask volume chosen should match the volume spanned by the activation maxima

included in the analysis. In our example, the activation coordinates in data set 5 from the database were distributed over the entire brain volume, including subcortical regions and even some white matter. We therefore chose as a mask the entire volume of a brain, normalized to standard size. The distribution of the random activation foci was then restricted to the area spanned by the minimum and maximum Talairach coordinates of the 728 empirical maxima. Note, however, that the particular choice of the mask volume is less critical than might appear at first sight. This is due to the large ratio between the empirical maxima and the number of voxels in the mask (in our analysis 728 and 61 408 voxels, respectively). For example, reducing the mask volume by 1/2 in our example analysis would change the ALE threshold only from 0.0156 to 0.018. The resulting thresholded ALE map would still contain the vast majority of the activation maxima that exceed the threshold when the full mask volume is used. This shows that slight variations in the mask volume do not significantly change the outcome of the subsequent model-based clustering.

Note that in our example data, ALE values were not corrected for multiple comparison (Laird et al., 2005a). Rather, as suggested in the original work by Turkeltaub et al. (2002), values were thresholded at a very small α-level of 0.01 % ($p = 0.0001$) to protect from family-wise Type I errors. Correction was omitted for the sake of simplicity, keeping in mind that 1) in our approach ALE serves as a preprocessing step to model-based clustering and thus should not be performed too conservatively, and 2) the aim of model-based clustering is the sub-clustering of large ALE foci which would in any case survive the correction procedure. Moreover, Laird and colleagues, when introducing multiple comparison correction for ALE, compared it to uncorrected thresholding with small thresholds and observed: "It is clear that thresholding the ALE maps at $p < 0.0001$ (uncorrected) produced results that most closely matched the FDR-corrected results (Laird et al., 2005a, p.161)." This confirms our own empirical observation that correcting ALE values, though statistically sound, in practical terms often amounts to using a smaller threshold without correction, as was done in our example.

The second step of our analysis procedure pertains to fitting Gaussian mixtures to the activation coordinates that survived the ALE threshold in the first analysis step. Although Gaussians are the most commonly used components in mixture modelling, they have a well-known limitation: Gaussian mixture models have a relatively high sensitivity to outliers which can lead to an over-estimation of the number of clusters (Svensén and Bishop, 2004). However, we would argue that this is not a critical issue

in our particular application, since such outliers are removed by ALE before the actual clustering.

Like in many clustering problems, the true number of clusters for a given set of activation maxima is not known in advance. This can be problematic as most clustering techniques require the number of clusters to be pre-specified. In the model-based clustering approach suggested here, this problem is solved by fitting a set of models with different numbers of clusters to the data and applying a model selection criterion afterwards. An alternative to this so-called deterministic approach (Abd-Almageed et al., 2003) are stochastic methods which allow an automatic determination of the number of clusters in the process of modelling (Richardson and Green, 1997; Svensén and Bishop, 2004; Abd-Almageed et al., 2005, e.g.). However, as these techniques are based on Markov Chain Monte Carlo or variational methods, they tend to be computationally very expensive.

The use of BIC as model selection criterion allows us to select the best number of clusters and the model parameterization simultaneously. Like most model selection criteria, BIC follows the principle of Occam's razor and favors from two or more candidate models the model that fits the data sufficiently well in the least complex way. In our context, this idea can be expressed formally using the estimated log likelihood of the models and a fixed penalizing term encoding the number of parameters of each model. Here, alternative approaches such as the Akaike information criterion (AIC) (Akaike, 1973) or the deviance information criterion (DIC) (Spiegelhalter et al., 2002) are conceivable. AIC, for example, is strongly related to BIC as it only differs in the simpler penalty term $2\,m$ (cf. Equation 5.9). This means, however, that for large sample sizes, AIC tends to favor more complex models compared to BIC. Other conceivable strategies include model selection procedures based on data-driven rather than fixed penalty terms (Shen and Ye, 2002, e.g.), or stochastic methods which allow an automatic determination of the number of clusters in the process of modelling (Richardson and Green, 1997; Svensén and Bishop, 2004; Abd-Almageed et al., 2005, e.g.). The application of different model selection criteria and their influence on the result of the clustering is one possible direction of future research.

Finally, note the relationship of different parameterizations of the Gaussians to other clustering criteria. For example, for the spherical model $\Sigma_\mathbf{k} = \lambda\mathbf{I}$, maximizing the complete-data log likelihood in Equation (5.5) refers to minimizing the standard k-means clustering criterion $\text{trace}(\mathbf{W})$ where $\mathbf{W}$ is the within-cluster scatter ma-

trix as defined in (5.6) and (5.7). Maximizing the likelihood of the ellipsoidal model $\mathbf{\Sigma_k} = \lambda \mathbf{D} \mathbf{A} \mathbf{D^T}$ is related to the minimization of $\det(\mathbf{W})$. Thus, allowing the parameterization of the covariance matrices to vary, model-based clustering encompasses and generalizes a number of classical clustering procedures[3]. The general problems of choosing an appropriate clustering technique and the optimal number of clusters are then formulated as model selection problem (Fraley and Raftery, 2002).

5.4 Network analysis

5.4.1 Replicator dynamics and meta-analyses

In Section 3.1 we demonstrated how replicator dynamics can be applied in the analysis of single-subject fMRI data. Specifically, replicator dynamics was used for the parcellation of cortical areas into subregions. In this section, we propose to apply replicator dynamics in a different context. We demonstrate how replicator dynamics can be used in the search for functional networks from meta-analytic imaging data. As stated in the introduction of this chapter, we define a functional network in the context of meta-analyses as a network or group of jointly activated brain regions.

In contrast to single-subject analyses that are based on fMRI time series, our only source of information for finding such networks on the meta-level are coordinates of activation maxima from several comparable studies. We therefore make use of the methods introduced in Sections 5.2 and 5.3, ALE and model-based clustering, to first calculate activation foci or clusters of activation coordinates from the input data. Each ALE focus or cluster of coordinates can then be represented by a replicator. The fitness of each replicator is derived from the number of co-occurrences with other brain regions in all experiments included in the analysis. The rational behind this choice of fitness function is as follows.

We are interested in finding a functional network underlying a particular cognitive task, which will usually include a number of different brain regions. It is safe to assume that brain regions which play a critical role in solving this task will be observed jointly activated in the majority of experiments. Additional activation foci might appear in only some of the experiments, arising, for example, from specific

---

[3]For a more detailed discussion on relations between classical cluster criteria and constraints on the model covariance matrix see, e.g., Everitt et al. (2001); Celeux and Govaert (1995); Banfield and Raftery (1993).

characteristics in the experimental setup and the stimulus material, differences in scanning hardware or the use of different software packages for the statistical evaluation of the data. Consequently, in our analysis we wish to find those brain regions that are most often jointly activated and separate them from regions whose activation is the result of measurement noise and the particularities of an experimental setup. The fitness assigned to each brain region should therefore be determined by the number or relative frequency of co-occurrences with other areas. This information can be captured in a co-occurrence matrix, which by definition is a non-negative, real-valued, and symmetric matrix and thus meets the requirements for the application of replicator dynamics (cf. Section 3.1).

In this new meta-analysis context the proportion of a replicator evolving during the replicator process can be interpreted as the 'importance' of a brain region for performing the experimental task. As discussed in Section 3.1, in order to avoid any bias in the input data, the proportion vector $\mathbf{x}$ should be initialized as $\mathbf{x}(0) = (1/n \ 1/n \ldots 1/n)$ where n now denotes the number of activation foci included in the analysis.

What happens when a replicator process is applied to a number of activation foci? First, for each focus the number of joint activations with other foci is represented in a co-occurrence matrix $\mathbf{W}$. Then Equation (3.2) is iteratively applied, where the new proportion of a replicator is determined from its individual fitness, its proportion, and the mean fitness of the population. More precisely, if for a focus i the term $(\mathbf{W}\mathbf{x}(t))_i$ exceeds the mean fitness $\mathbf{x}(t)^{\mathbf{T}}\mathbf{W}\mathbf{x}(t)$, its proportion will increase, if it does not exceed the mean fitness, the proportion will decrease.

We can see that in the first step of the iteration the numerator in Equation (3.2) only depends on the matrix $\mathbf{W}$ and on the overall number of replicators, since the initial proportion x_i is the same for all activation foci:

$$(\mathbf{W}\mathbf{x}(0))_i = \sum_j \mathbf{W}_{ij} \, x_j(0) = \sum_j \mathbf{W}_{ij} \, 1/n = 1/n \sum_j \mathbf{W}_{ij} \qquad (5.10)$$

In other words, only the total number of joint activations with other foci determines the new proportion for a particular replicator. Consequently, two replicators i and j with the same number of co-occurrences with other activation foci will be assigned the same proportion in the first step of the iteration: $x_i(1) = x_j(1)$. For foci that are relatively often jointly activated with other activation foci, the ratio of individual and mean fitness is high. Such areas will be assigned a high new proportion in the population. In contrast, foci with only a few co-occurrences will be assigned a lower proportion.

In the following steps of the iteration, the term $(\mathbf{W}\mathbf{x}(t))_i$ no longer only depends on the total number of joint activations for a focus, it now also becomes important with which other foci it occurs. A focus i which co-occurs with other 'important' foci, i.e. with foci that were assigned a high proportion in the previous step, will increase its own proportion more than a focus j which is equally often jointly activated but with less 'important' other foci.

As the theorem of natural selection states, the mean fitness $\mathbf{x}(t)^{\mathbf{T}}\mathbf{W}\mathbf{x}(t)$ is maximized as the process goes on. Therefore, only the fitness and the proportion of the most important foci will be high enough to exceed the mean fitness in later stages of the replicator process and, as a result, a network of the most important activation foci will emerge. Again, the replicator process becomes stationary when $\mathbf{x}^{\mathbf{T}}\mathbf{W}\mathbf{x}$ reaches its maximum (cf. Section 3.1).

5.4.2 An example

The following example is designed to illustrate the behavior of a replicator process for a number of activation areas with different fitness values. While the example is not derived from real data, it is constructed to demonstrate some of the mechanisms involved in determining in an iterative process a dominant network of replicators with equal initial fitness. Results from applications to real meta-analytic data will be presented later in this chapter.

	1	2	3	4	5	6	7	8	9	10	11	12	13	14
A	x	x		x	x		x	x	x			x	x	x
B	x	x	x	x			x	x		x			x	
C		x			x	x		x				x		x
D		x		x	x					x	x			
E					x						x			
F										x				

Table 5.3: Activation for 6 areas (A to F) as they could have been observed in 14 experiments.

Suppose within 14 experiments investigating the same cognitive task we had observed the activation of 6 different brain regions as shown in Table 5.3. The co-occurrence of activated brain regions can be conveniently pictured in an undirected
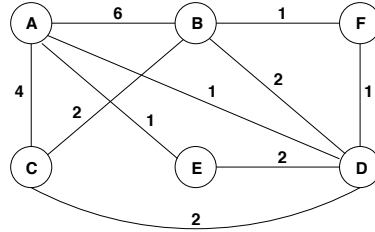
*Figure 5.8: Co-occurrences of activated brain regions from Table 5.3 represented as an undi-
rected graph. Nodes represent activation foci and weighted connections represent the number
of co-occurrences of two foci.*

graph where nodes represent activation foci and connections between nodes represent
joint activations. The weight of a connection is the number of joint activations found
in all experiments. The graph corresponding to Table 5.3 is depicted in Figure 5.8.

As graph and table show, brain regions A and B have the most co-occurrences with
other activation foci, 12 and 11, respectively. Focus C and focus D have the same total
number of co-occurrences, but co-occur partly with different other brain regions. E and
F appear only rarely. From the graph we can easily derive the following co-occurrence
matrix:

$$\mathbf{W} = \begin{bmatrix} 0 & 6 & 4 & 1 & 1 & 0 \\ 6 & 0 & 2 & 2 & 0 & 1 \\ 4 & 2 & 0 & 2 & 0 & 0 \\ 1 & 2 & 2 & 0 & 2 & 1 \\ 1 & 0 & 0 & 2 & 0 & 0 \\ 0 & 1 & 0 & 1 & 0 & 0 \end{bmatrix} \tag{5.11}$$

The order of rows and columns in the matrix $\mathbf{W}$ corresponds to the order of the foci
in Table 5.3. Since 6 cortical regions are included in the analysis, the proportion vector
$\mathbf{x}$ is initialized as

$$\mathbf{x}(0) = (\,1/6 \ \ 1/6 \ \ 1/6 \ \ 1/6 \ \ 1/6 \ \ 1/6\,)^{\mathbf{T}}. \tag{5.12}$$

After the first iteration through Equation 3.2 we get the new proportion vector

$$\mathbf{x}(1) = (\,0.27 \ \ 0.25 \ \ 0.18 \ \ 0.18 \ \ 0.07 \ \ 0.04\,)^{\mathbf{T}}. \tag{5.13}$$

As we would expect from the distribution of co-occurrences, regions A and B were
assigned the largest proportion. C and D with the same total number of co-occurrences
also have the same proportion after the first iteration. The proportion of E and F has
decreased from the initial value, since $(\mathbf{W}\mathbf{x}(0))_{\mathbf{E}} = 0.50$ and $(\mathbf{W}\mathbf{x}(0))_{\mathbf{F}} = 0.33$ do

not exceed the mean fitness $\mathbf{x}(0)^{\mathbf{T}}\mathbf{W}\mathbf{x}(0) = 1.22$. After applying the membership function, we see that the proportions of A, B, C, and D exceed the mean proportion 1/6. These foci are therefore member of the preliminary dominant network.

In the second iteration, the proportion vector changes to

$$\mathbf{x}(2) = (\,0.35\ \ 0.31\ \ 0.18\ \ 0.12\ \ 0.02\ \ 0.01\,)^{\mathbf{T}}. \tag{5.14}$$

The proportion of A and B has increased again, the proportion of E and F has further decreased. The proportion of C now differs from that of D. While the proportion of C stayed approximately the same in this iteration, the proportion of D has decreased. This is caused by the fact that C and D co-occur with different activation foci. C is most often jointly activated with regions A and B, which in turn have the two highest proportions among all activation foci after the previous iteration. In contrast, focus D occurs more often with E and F which have a relatively low proportion after the previous step. Consequently, $(\mathbf{W}\mathbf{x}(1))_{\mathbf{C}} = 1.95$ is higher than $(\mathbf{W}\mathbf{x}(1))_{\mathbf{D}} = 1.32$. In fact, $(\mathbf{W}\mathbf{x}(1))_{\mathbf{D}}$ no longer exceeds the new mean fitness $\mathbf{x}(1)^{\mathbf{T}}\mathbf{W}\mathbf{x}(1) = 1.93$. Thus, the proportion of D decreases, and after applying the membership function again, only foci A, B, and C remain in the dominant network.

The third and forth iteration result in

$$\mathbf{x}(3) \;=\; (\,0.39\ 0.35\ 0.17\ 0.07\ 0.00\ 0.00\,)^{\mathbf{T}} \tag{5.15}$$
$$\mathbf{x}(4) \;=\; (\,0.42\ 0.37\ 0.15\ 0.03\ 0.00\ 0.00\,)^{\mathbf{T}}. \tag{5.16}$$

The proportion of C decreases for the first time in the third iteration, since $(\mathbf{W}\mathbf{x}(2))_{\mathbf{C}} = 2.26$ is smaller than the new mean fitness $\mathbf{x}(2)^{\mathbf{T}}\mathbf{W}\mathbf{x}(2) = 2.41$. However, after the third iteration the proportion of C is still large enough to keep C as a member of the dominant network, whereas after the forth iteration it drops below the threshold. The membership of only A and B in the dominant network now remains unchanged for all following iterations. As result of the replicator process, regions A and B are determined as dominant network in this example.

5.4.3   From lists of coordinates to replicators

Prior to the application of the replicator equations, we need to determine the replicators, i.e. brain regions that serve as candidates for the dominant network, from the lists of activation maxima in the considered experiments. For this, we can resort to

the methods introduced in Sections 5.2 and 5.3. ALE can be applied to automatically transform activation coordinates into ALE foci. Thus, ALE can be viewed as a prepro-cessing step to our search for functional networks. In cases where ALE results in large activation foci that are unlikely to represent distinct functional regions, model-based clustering can be applied additionally. The number of co-occurrences of an activation focus surviving ALE or a cluster of coordinates found by model-based clustering is de-termined by simply counting, for all reported coordinates within a focus or cluster, the number of co-occurrences with coordinates laying within every other focus or cluster across all studies.

5.4.4 Application

Our network analysis technique was tested using data set 4 which is described in detail in Section 2.3.4. Plots of all coordinates within this data set projected onto a single axial, sagittal and coronal slice are shown in the top row of Figure 5.9. Coordinates reported from different studies are represented by different colors. As can be seen, activation maxima are distributed over large parts of the cortex, and visual inspection does not reveal any obvious clustering of activation foci.

Activation likelihood estimates were calculated for all voxels using 3D Gaussian distributions with variance $\sigma = 3$ mm. The subsequent search for topologically con-nected regions exceeding an ALE threshold of 0.05 yielded the 15 activation foci de-picted in the bottom row of Figure 5.9. The foci are again projected onto a single axial, sagittal and coronal slice. Subsequent model-based clustering did not reveal any sub-clustering of the ALE regions. The best model fit was obtained for a model with 15 clusters, where cluster membership corresponded to the membership of the activation maxima to the 15 ALE regions.

A co-occurrence matrix was calculated from all coordinates laying within these 15 brain regions and the replicator process was applied until the membership of activation foci in the dominant network remained stable for 20 iterations.

During the replicator process five regions were selected as members of the domi-nant network: the presupplementary motor area (preSMA), the inferior frontal sulcus (IFS) extending onto the middle frontal gyrus, the anterior cingulate cortex (ACC) of both hemispheres extending into BA 8, and the inferior frontal junction area (IFJ). These five activation foci are depicted in Figure 5.10. The fitness of the remaining 10 activation foci dropped to values close to zero. Volume, mean ALE, the number
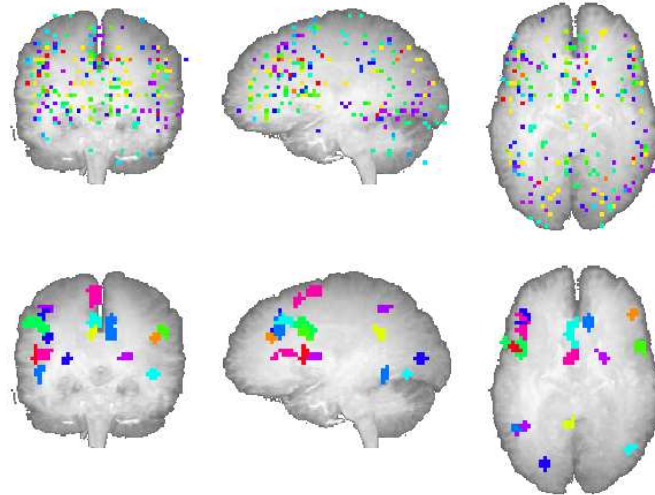
Figure 5.9: Top: 239 activation maxima obtained for fMRI experiments addressing the Stroop interference task (data set 4). Bottom: 15 activation foci derived from ALE of the above co-ordinates. Activation maxima and foci are shown in the glass brain view, i.e. projected onto three orthogonal single coronal, sagittal and axial slices.

of joint activations and the fitness of the five selected regions are listed in Table 5.4. The highest fitness was observed for the ACC of the right hemisphere which is the second largest activation focus and also shows the second largest number of joint activations. The largest number of joint activations was found for the IFS. While this is the smallest region in the dominant network, it was assigned the second highest fitness by the replicator process. These results confirm our earlier observation that the fitness of a replicator does not merely reflect the total number of joint activations with other regions. Neither does it only reflect its anatomical size. Rather, while the size and ALE of a focus are derived from the number and exact location of reported activation maxima within this brain region, the result of the replicator process is determined by the relationship between different activation foci.

5.4.5 Discussion

The application of replicator dynamics in meta-analytic studies addresses an important aspect of functional imaging research, the search for functional networks of brain regions underlying a specific cognitive task. It is based on the assumption that cor-

Region	Volume (mm^3)	mean ALE	# of joint activations	fitness
IFJ	3,012	0.09	38	0.17
IFS	1,912	0.08	46	0.27
preSMA	2,023	0.08	30	0.14
ACC left	2,134	0.07	34	0.12
ACC right	2,500	0.08	40	0.29

Table 5.4: The five members of the dominant network resulting from the replicator process. Volume, mean ALE, number of joint activations, and fitness of the five regions selected by the replicator process.

(a)

(b)   (c)
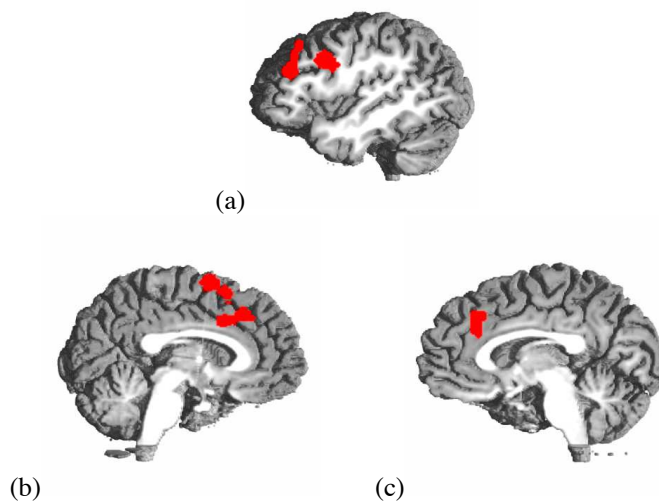
Figure 5.10: The five selected members of the dominant network. They include the IFJ, IFS extending into middle frontal gyrus, preSMA, and ACC of both hemispheres extending into BA 8. (a) and (b) show the left hemisphere at $x = -44$ and $x = -3$, respectively, (c) shows the right hemisphere at $x = 7$.

tical areas involved in solving the same cognitive task are likely to be found jointly activated in experiments addressing this task. Consequently, the method searches for regions that are jointly activated in the majority of experiments considered in the meta-analysis. It is important to be clear that this search goes beyond the usually performed simple counting of activation maxima in particular cortical regions. Counting the number of studies that activated an area can provide useful information about the role this area might play in performing a particular task. However, it leaves any information about the relationship between different areas unexplored. Our new method draws on both sources of information. The number of studies that activated a particular region and the distribution of activation over anatomical regions first control the choice of activation foci that enter the replicator process. The co-occurrences of brain regions then determine the members of the dominant network.

We chose a well-studied paradigm for the application of our methodology and, in order to include a reasonable amount of data, placed only minimal restrictions on the selected studies. While the results of the replicator process agree well with results reported in many other Stroop studies, we have to be aware of the fact that they might be biased at least to some degree. Some of the included studies, for example, were aimed at investigating the possible functionality of the anterior cingulate cortex. It is thus not surprising to find a high proportion of reported activations in this particular area, while other activation locations might have been omitted in the publication of these studies. This might also explain the absence of any parietal areas in the result of the replicator process. Some activation foci were derived from parietal activation maxima reported in the considered experiments (see Figure 5.9, bottom). However, their co-occurrences with other foci were outnumbered by those of frontal regions which were of particular interest in some of the experiments. The careful selection of studies included in the analysis is therefore, like in all meta-analytic approaches, the most important prerequisite for the successful application of the replicator process and the meaningful interpretation of its results.

One way of alleviating the problems introduced by a publication bias is to include large amounts of data in a meta-analysis. This in turn requires efficient analysis tools. The method presented here is very fast and easy to implement. After extracting activation maxima from the database, the results of our example analysis were calculated within a few seconds using MATLAB and an AMD Athlon(TM) XP1800 processor. Our method should therefore be easily applicable to large-scale studies.

With our example data set we have restricted the meta-analysis to different variants of the same experimental paradigm. One could also conceive of applying the method to data gathered for different paradigms which, for example, are expected to employ similar neuropsychological mechanisms. The application of the replicator process could then reveal a functional network of cortical areas that these paradigms have in common, and dissociate it from areas that are specific to the different paradigms investigated.

Chapter 6

Conclusion

With the development and use of structural and functional neuroimaging techniques we have greatly advanced our understanding of the human brain. Today, in vivo imaging techniques like MRI and fMRI are at the heart of modern cognitive and neurosciences. The ongoing development of neuroimaging techniques is accompanied by the continuous search for efficient and robust computational analysis methods for the acquired data.

It is widely agreed that the brain has a modular structure and, on a functional level, implements a variety of working principles such as functional segregation, functional integration, and temporal synchronization (Posner et al., 1988; Schwarz et al., 2008). The investigation of these principles calls for a multitude of techniques that can be applied on different analysis levels and that are able to extract and operate on different properties of the obtained signals. For example, many imaging experiments are conducted to investigate where in the brain specific cognitive processes are implemented. For this, analysis methods have to be employed that can detect and segregate individual brain regions where the observed signal closely correlates with certain aspects of the experimental design. Other experiments are aimed at studying the interplay between different brain regions. In this case, network analysis methods are of particular importance.

In developing techniques for the analysis of fMRI data, we face a number of serious challenges. Firstly, the functional MRI signal is only an indirect marker of the neuronal activity that we wish to investigate. Thus, complex neurovascular mechanisms have to be taken into account in the analysis of the fMRI signal (Logothetis, 2003). This can be achieved by the development of models that account for what is known about the

spatial and temporal properties and the precise shape of the BOLD response. Unfortunately, the exact link between neuronal activity and the BOLD signal is still not fully understood and current models often disregard significant parts of the fMRI signal (Fox and Raichle, 2007). Thus, in addition to model-based approaches, exploratory analysis techniques are needed that detect and elucidate the yet unexplained components of the fMRI signal.

Secondly, the fMRI signal suffers from a relatively poor signal-to-noise ratio. For most applications, this necessitates heavy preprocessing, such as temporal and spatial smoothing of the data (Tabelow et al., 2009). In addition, we have to resort to statistical approaches that facilitate data averaging over many experimental trials and across several subjects. This, however, decreases the local specificity of the analysis results, making the precise anatomical localization of the detected brain activation difficult (Logothetis, 2008).

Thirdly, fMRI signals are known to be highly variable across space and time. That is, the temporal dynamics and precise shape of the BOLD response can vary significantly across different subjects, across multiple measurements of the same subject, and across brain regions (Schacter et al., 1997; Buckner et al., 1998; Aguirre et al., 1998; Miezin et al., 2000; McGonigle et al., 2000). This calls for analysis techniques that are flexible and robust enough to produce generalizable results despite the unavoidable intra-individual, inter-individual, and inter-regional differences. In addition, the properties of the BOLD response and the relations between different brain regions are subject to developmental changes and disease (D'Esposito et al., 2003; Meunier et al., 2009). Analytical methods are thus needed that enable us to investigate these changes and to incorporate them into the data analysis of patients studies and studies including children or elderly participants.

In this work, we have presented a series of fMRI data analysis techniques that are aimed at investigating different aspects of the fMRI signal and related cognitive processing and, at the same time, alleviate some of the aforementioned problems. On the level of individual fMRI data sets we proposed a method for the parcellation of functional brain regions. This exploratory method complements model-based approaches where the functional segregation of brain regions is investigated solely on the grounds of experimental manipulations of the underlying cognitive processes. With our new method we can search for subclusters of voxels within seemingly homogeneously ac-

tivated functional regions, where a further subdivision by means of experimental manipulation alone failed.

It is important to note that the results obtained in our example analyses were highly reproducible both across different subjects and across multiple measurements of the same subject. This underlines the robustness of our method against the temporal and spatial variability of the BOLD signal and the anatomical variability between subjects (Neumann et al., 2006). This robustness is achieved by employing replicator dynamics as parcellation mechanism and canonical correlation as similarity measure for fMRI time series. In contrast to traditional clustering techniques, where every data point must be assigned to a cluster, replicator dynamics is a selection process, where clusters of highly similar items are formed, while outliers are disregarded. In addition, canonical correlation as a multivariate similarity measure takes into account the similarity of fMRI time series in neighboring voxels, this way extenuating the influence of the anatomical between-subject variability on the parcellation results. The resulting high reliability of the clustering could not be achieved by means of simpler univariate similarity measures and clustering methods. This demonstrates that the choice of appropriate techniques is of paramount importance for the analysis of functional MRI data, given the specific properties of fMRI time series and the complex relationships between them.

Our second method for the analysis of fMRI time series was developed for the examination of temporal properties of the BOLD signal. The results of this development are two-fold. Firstly, we could demonstrate that one can indeed reliably identify properties of averaged fMRI time series, such as time-to-peak, that make temporal aspects of the BOLD signal accessible and comparable. Secondly, our method confirmed temporal offsets between different brain regions in the range of several seconds and an equally large between-subject variability in the temporal dynamics of the BOLD signal. While the origin of this high variability is still not entirely clear, these results remind us to remain cautious in the interpretation of temporal aspects of the BOLD signal, for example in the analysis of brain networks or the investigation of the information flow between functional regions. Such analyses must not be based on temporal aspects of the BOLD signal alone.

Despite these limitations on temporal information from the BOLD signal, the application of our analysis method in several imaging studies provided useful insights, for example, into the neural correlates of decision making processes (Zysset et al.,

2006) and the BOLD variability observed in very long imaging sessions (Menz et al., 2006). Moreover, we could identify temporal differences in language processing between adults and children (Brauer et al., 2008) and age-related changes in the BOLD signal measured for the Stroop interference task (Zysset et al., 2007).

Given the high variation of the BOLD signal across space and time, both the within- and between-subject variability of fMRI time series need to be accounted for in the statistical analysis of imaging experiments, in particular when assessing data from multiple subjects. On this analysis level, statistical parametric mapping within the framework of frequentist statistics is most commonly employed. This approach, however, by and large neglects the within-subject variability of the individual model parameters obtained on the first level.

We have provided an alternative approach for the second-level analysis of fMRI data based on Bayesian statistics. In this approach, the influence of individual subject data on the group analysis is weighted by the subject's within-subject variability of the estimated model parameters on the first level. Consequently, outliers are less likely to strongly affect second-level statistical inferences (Neumann and Lohmann, 2003), and relatively small effects resulting from experimental manipulations can still be reliably detected. A number of studies could already draw on this advantage, for example, studies that addressed the processing of language, music, and sound (Ruschemeyer et al., 2005; Krumbholz et al., 2005; Tervaniemi et al., 2006; Tillmann et al., 2006), the neural correlates of intuition (Volz and von Cramon, 2006) and error processing (Fiehler et al., 2004). Most notably, with our Bayesian method we were able to infer subtle differences in learning from errors that related to small genetic variations between two groups of subjects (Klein et al., 2007b).

With the exponentially growing number of imaging studies each year, meta-analyses have recently received growing interest in the neuroscience community (Wager et al., 2007; Derrfuss and Mar, 2009; Costafreda, 2009). The first neuroimaging meta-analysis, published in 1991, merely consisted of plotting activation coordinates from three previous studies into a common coordinate system (Frith et al., 1991; Fox et al., 2005a). Since then, a number of automatic coordinate-based techniques have emerged that allow for objective and statistically formal, quantitative meta-analyses. Most notably, Activation Likelihood Estimation (Turkeltaub et al., 2002) and concurrently developed comparable methods (Chein et al., 2002; Wager et al., 2004) are frequently used to model functional brain regions based on activation maxima obtained in

independently performed imaging experiments. We previously applied ALE in meta-analyses of cognitive control (Derrfuss et al., 2005; Neumann et al., 2005) and error processing (Klein et al., 2007a), and extended its application to anatomical data in the investigation of fronto-temporal degeneration in patients with dementia (Schroeter et al., 2007, 2008, 2009) and neural correlates of obesity (Chaudhry et al., 2010).

With our work presented here, we also provided two new meta-analysis techniques, addressing the principles of functional segregation and functional integration, respectively. Firstly, we applied exploratory clustering based on Gaussian mixture models to activation coordinates, in order to determine functional regions that were frequently activated in the examined imaging experiments. This new approach should be viewed as an extension rather than a replacement of ALE, as it facilitates a further sub-clustering, where ALE yields results that are too large to be interpretable as homogeneous functional regions (Neumann et al., 2008). The aim of our method is thus comparable to the parcellation of functional brain regions that we proposed in the context of first-level analyses: Where model-based approaches and experimental manipulations fail, exploratory analysis techniques are applied to yield a functional segregation of brain regions on the grounds of statistical properties of the obtained BOLD signals or activation coordinates.

Our second meta-analysis technique was developed to aid the detection of functional brain networks. We again employed the concept of replicator dynamics, this time making use of a co-activation matrix rather than a similarity measure for individual time series. This way, we were able to capture commonalities in the activation of functional regions across several experiments and infer the most frequently co-activated functional regions for a specific cognitive paradigm. Using this method, we could, for example, confirm the major regions involved in a fronto-temporal cognitive control network (Neumann et al., 2005) and identify an extended language network facilitated in the comprehension of coherent or incoherent text (Ferstl et al., 2008).

In its current state, our technique yields undirected functional networks. That is, from a matrix recording pairwise co-activations of brain regions, we can infer the common activation of an entire network of brain regions across experiments. However, as it stands, the method does not provide us with any information about possible directed dependencies between these regions. One approach to the search for directed dependencies between functional brain regions can be found in the field of Bayesian networks. With Bayesian networks, we can represent statistical dependencies between

entities, in our context between functional brain regions. Most importantly, the structure of Bayesian networks can be learned from observational data (Heckerman et al., 1995; Pearl, 2000). In our current research we investigate how, on the meta-analysis level, structure learning of Bayesian networks can be applied to co-activation matrices recording the co-activation of functional regions across different imaging experiments (Neumann et al., 2010).

The investigation of directed or even causal dependencies between functional brain regions is not restricted to the meta-analysis level. In fact, the search for connectivities and dependencies between functional brain regions is among the most challenging questions for future work in the development of fMRI analysis methods. On the level of individual fMRI experiments, different approaches to connectivity analyses have already been proposed, including structural equation modelling (McIntosh and Gonzalez-Lima, 1994; Büchel and Friston, 1997) and dynamic causal modelling (Friston et al., 2003). However, as these methods are heavily hypothesis-driven and based on strong model assumptions about the BOLD signal, their feasibility for the investigation of directed functional networks is still subject to heated debates.

In addition to connectivity analyses, there are a number of important research questions for the future development of functional neuroimaging. First, how can we exploit information from several imaging modalities at the same time? Can we, for example, efficiently combine EEG measurements, given their high temporal resolution, and fMRI data, given their unsurpassed spatial resolution, in a multi-modal approach? While a simultaneous measurement of the two modalities is now technically possible, computationally, the conjoint data analysis is still a challenging task, as data from different imaging modalities vary widely in temporal and spatial scale, data representation and statistical properties (Halchenko et al., 2005).

Second, how can we integrate anatomical and functional brain data? For example, is it possible to consolidate directed statistical dependencies between brain regions obtained from fMRI measurements with anatomical connections derived from MRI data? And how can anatomical constraints, such as the spatial distribution of grey and white matter in the brain, be incorporated into the analysis of functional imaging data? Using anatomical information to guide fMRI data analysis is expected to greatly improve the anatomical specificity of detected brain activation (Kiebel et al., 2000), but this again calls for new analysis techniques that can take into account data from different imaging modalities and of highly varying mathematical properties.

A third line of future research is model-based fMRI. Model-based fMRI builds on individual differences in parameters that quantify specific cognitive processes. Specifically, parameters are estimated for individual subjects based on computational models of the subjects' behavior. These parameter estimates are then incorporated into the analysis of fMRI measurements in search for brain structures that are involved in parameter setting and processing. For example, in a recent study we investigated the certainty that participants developed about the correctness of a specific choice during a decision making experiment (Klein et al., 2007b). While this certainty was not behaviorally measurable on a trial-by-trial basis, it could be estimated from a neural network model mimicking the subject's behavior. Subsequently, the estimated parameter was used as additional regressor in the statistical analysis of the obtained fMRI data, and brain regions encoding and processing this parameter were identified.

In a similar vain, we were able to model decision making processes under time pressure (Forstmann et al., 2008) and decision processes incorporating prior knowledge (Forstmann et al., 2010) by means of a linear ballistic accumulator. Daw et al. (2005) employed two computational systems implementing tree-search and temporal difference learning, respectively, in order to model two competing cognitive control mechanisms. These few examples of model-based fMRI studies already show that the integration of computational models and fMRI data analysis can greatly enhance our understanding of cognitive processes and their implementation in the brain.

Finally, new fMRI data analysis techniques are currently emerging that explore the use of multi-dimensional pattern analysis tools like artificial neural networks and support vector machines for the classification and prediction of distributed cortical activation patterns. The earliest application of this approach already showed that activation patterns encoding objects of different categories can be clearly distinguished even by means of a very simple pattern classifier (Haxby et al., 2001). Given the inherently distributed nature of cortical representations and processes, the use of sophisticated multi-dimensional pattern analysis methods, as developed in computer science and machine learning, could bring us a large step forward in the understanding of the brain's computations.

While this list of future research questions is by no means comprehensive, it provides some examples that illustrate the importance of continuous methodological development in the field of functional neuroimaging. Without new computational meth-

ods for the analysis of functional imaging data, we will not be able to analyze and understand the human brain in its entire complexity.

Unquestionably, understanding the brain requires a truly interdisciplinary research effort. Computer science and mathematics, among other disciplines, play a key role in this ambitious enterprise.

Bibliography

Abd-Almageed, W., El-Osery, A., and Smith, C. E. (2003). Non-parametric expectation maximization: A learning automata approach. In *Proceedings of the IEEE International Conference on Systems, Man and Cybernetics, Washington, DC*.

Abd-Almageed, W., El-Osery, A., and Smith, C. E. (2005). Estimating time-varying densities using a stochastic learning automaton. *Soft Computing*. Issue: Online First, 17. November 2005.

Aguirre, G. K., Zarahn, E., and D'Esposito, M. (1998). The variability of human BOLD hemodynamic responses. *NeuroImage*, 8(4):360–369.

Akaike, H. (1973). Information theory and an extension of the maximum likelihood principle. In *Proceeding of the Second International Symposium on Information Theory, Budapest*, pages 267–28.

Amodio, D. A. and Frith, C. D. (2006). Meeting of minds: the medial frontal cortex and social cognition. *Nature Reviews Neuroscience*, 7:268–277.

Anderson, T. W. (2003). *An Introduction to Multivariate Statistical Analysis*. John Wiley & Sons, New York, 3rd edition.

Ardekani, B. A. and Kanno, I. (1998). Statistical methods for detecting activated regions in functional MRI of the brain. *Magnetic Resonance Imaging*, 16(10):1217–1225.

Banfield, J. and Raftery, A. (1993). Model-Based gaussian and Non-Gaussian Clustering. *Biometrics*, 49:803–821.

Banich, M. T., Milham, M. P., Atchley, R. A., Cohen, N. J., Webb, A., Wszalek, T., Kramer, A. F., Liang, Z.-P., Wright, A., Shenker, J., Magin, R., Barad, V., Gullett, D., Shah, C., and Brown, C. (2000). fMRI studies of stroop tasks reveal unique roles of anterior and posterior brain systems in attentional selection. *Journal of Cognitive Neuroscience*, 12(6):988–1000.

Blamire, A. M., Ogawa, S., Ugurbil, K., Rothman, D., McCarthy, G., Ellermann, J. M., Hyder, F., Rattner, Z., and Shulman, R. G. (1992). Dynamic mapping of the human

visual cortex by high-speed magnetic resonance imaging. *Proc Natl Acad Sci USA*, 89:11069–11073.

Bomze, I. M., Pelillo, M., and Stix, V. (2000). Approximating the maximum weight clique using replicator dynamics. *IEEE Transactions on Neural Networks*, 11(6):1228–1241.

Box, G. E. P. and Tiao, G. C. (1992). *Bayesian Inference in Statistical Analysis*. John Wiley & Sons, New York.

Boynton, G. M., Engel, S. A., Glover, G. H., and Heeger, D. J. (1996). Linear systems analysis of fMRI in human V1. *Journal of Neuroscience*, 16:4207–4221.

Brass, M., Derrfuss, J., Forstmann, B., and von Cramon, D. Y. (2005). The role of the inferior frontal junction area in cognitive control. *Trends in Cognitive Sciences*, 9(7):314–316.

Brass, M. and von Cramon, D. Y. (2004). Decomposing components of task preparation with functional magnetic resonance imaging. *Journal of Cognitive Neuroscience*, 16(4):609–620.

Brauer, J., Neumann, J., and Friederici, A. D. (2008). Temporal dynamics of perisylvian activation during language processing in children and adults. *NeuroImage*, 41(4):1484–1492.

Büchel, C. and Friston, K. J. (1997). Modulation of connectivity in visual pathways by attention: Cortical interactions evaluated with structural equation modelling and fMRI. *Cerebral Cortex*, 7(8):768–778.

Buckner, R. L. (1998). Event-related fMRI and the hemodynamic response. *Human Brain Mapping*, 6:373–377.

Buckner, R. L., Bandettini, P. A., O'Craven, K. M., Savoy, R. L., Petersen, S. E., Raichle, M. E., and Rosen, B. R. (1996). Detection of cortical activation during averaged single trials of a cognitive task using functional magnetic resonance imaging. *Proc Natl Acad Sci USA*, 93:14878–14483.

Buckner, R. L., Koutstaal, W., Schacter, D. L., Dale, A. M., Rotte, M., and Rosen, B. R. (1998). Functional-anatomic study of episodic retrieval: II. Selective averaging of event-related fMRI trials to test the retrieval success hypothesis. *NeuroImage*, 7:163–175.

Buxton, R. B. (2002). *Introduction to Functional Magnetic Resonance Imaging: Principles and Techniques*. Cambridge University Press, Cambridge.

Buxton, R. B. (2010). Interpreting oxygenation-based neuroimaging signals: the importance and the challenge of understanding brain oxygen metabolism. *Frontiers in Neuroenergetics*, 2(8).

Celeux, G. and Govaert, G. (1995). Gaussian parsimonious clustering model. *Pattern Recognition*, 28(5):781–793.

Chaudhry, S., Neumann, J., Villringer, A., and Margulies, D. (2010). The relationship of body mass index to neural activity: A meta-analysis. 16th Annual Meeting of the Organization for Human Brain Mapping (OHBM), Barcelona.

Chein, J. M., Fissell, K., Jacobs, S., and Fiez, J. A. (2002). Functional heterogeneity within broca's area during verbal working memory. *Physiology & Behavior*, 77:635–639.

Cohen, M. S. (1997). Parametric analysis of fMRI data using linear systems methods. *NeuroImage*, 6(2):93–103.

Costafreda, S. G. (2009). Pooling fMRI data: meta-analysis, mega-analysis and multi-center studies. *Frontiers in Neuroinformatics*. doi:10.3389/neuro.11/033.2009.

Davidson, M. L. (1983). *Multidimensional scaling*. John Wiley & Sons, New York.

Daw, N. D., Niv, Y., and Dayan, P. (2005). Uncertainty-based competition between prefrontal and dorsolateral striatal systems for behavioral control. *Nature Neuroscience*, 8(12):1704–1711.

Dempster, A. P., Laird, N. M., and Rubin, D. B. (1977). Maximum likelihood from incomplete data via the em algorithm. *Journal of the Royal Statistical Society, Series B*, 39:1–38.

Derrfuss, J., Brass, M., Neumann, J., and von Cramon, D. Y. (2005). Involvement of the inferior frontal junction in cognitive control: meta-analyses of switching and stroop studies. *Human Brain Mapping*, 25(1):22–34.

Derrfuss, J., Brass, M., and von Cramon, D. Y. (2004). Cognitive control in the posterior frontolateral cortex: evidence from common activations in task coordination, interference control, and working memory. *NeuroImage*, 23(2):604–612.

Derrfuss, J. and Mar, R. A. (2009). Lost in localization: The need for a universal coordinate database. *NeuroImage*, 48(1):1–7.

D'Esposito, M., Deouell, L. Y., and Gazzaley, A. (2003). Alterations in the BOLD fMRI signal with ageing and disease: a challenge for neuroimaging. *Nature Review Neuroscience*, 4(11):863–872.

DeYoe, E. A., Bandettini, P., Neitz, J., Miller, D., and Winans, P. (1994). Functional magnetic resonance imaging (fMRI) of the human brain. *Journal of Neuroscience Methods*, 54(2):171–187.

Ding, C., He, X., Zha, H., Gu, M., and Simon, H. (2001). A min-max cut for graph partitioning and data clustering. In *Proceedings of 1st IEEE International Conference on Data Mining*, pages 107–114, San Jose, CA.

Dodel, S., Herrmann, J. M., and Geisel, T. (2002). Functional connectivity by cross-correlation clustering. *Neurocomputing*, 44–46:1065–1070.

Duann, J.-R., Jung, T.-P., Kuo, W.-J., Yeh, T.-C., Makeig, S., Hsieh, J.-C., and Sejnowski, T. J. (2002). Single-trial variability in event-related BOLD signals. *NeuroImage*, 15:823–835.

Duncan, J. and Owen, A. M. (2000). Common regions of the human frontal lobe recruited by diverse cognitive demands. *Trends in Neurosciences*, 23(10):475–483.

Eickhoff, S. B., Laird, A. R., Grefkes, C., Wang, L. E., Zilles, K., and Fox, P. T. (2009). Coordinate-based ALE meta-analysis of neuroimaging data: A random-effects approach based on empirical estimates of spatial uncertainty. *Human Brain Mapping*, 30(9):2907–2926.

Everitt, B. S., Landau, S., and Leese, M. (2001). *Cluster Analysis*. Oxford University Press, 4th edition.

Ferstl, E. C., Neumann, J., Bogler, C., and von Cramon, D. Y. (2008). The extended language network: A meta-analysis of neuroimaging studies on text comprehension. *Human Brain Mapping*, 29(5):581–593.

Fiebach, C., Schlesewsky, M., Lohmann, G., von Cramon, D., and Friederici, A. (2005). Revisiting the role of broca's area in sentence processing: syntactic integration versus syntactic working memory. *Human Brain Mapping*, 24(2):79–91.

Fiedler, M. (1973). Algebraic connectivity of graphs. *Czechoslovak Mathematical Journal*, 98:298–305.

Fiehler, K., Ullsperger, M., and von Cramon, D. Y. (2004). Neural correlates of error detection and error correction: is there a common neuroanatomical substrate? *European Journal of Neuroscience*, 19(11):3081–3087.

Fisher, R. A. (1930). *The Genetical Theory of Natural Selection*. Oxford University Press, London, UK.

Forstmann, B., Brown, S., Dutilh, G., Neumann, J., and Wagenmakers, E.-J. (2010). The neural substrate of prior information in perceptual decision making: a model-based analysis. *Frontiers in Human Neuroscience*, 4.

Forstmann, B., Dutilh, G., Brown, S., Neumann, J. von Cramon, D., Ridderinkhof, K., and Wagenmakers, E.-J. (2008). Striatum and pre-SMA facilitate decision-making under time pressure. *Proceedings of the National Academy of Sciences of the United States of Americ*, 105(45):17538–17542.

Forstmann, B. U., Brass, M., Koch, I., and von Cramon, D. Y. (2005). Internally generated and directly cued task sets: an investigation with fMRI. *Neuropsychologia*, 43(6):943–952.

Fox, M. D. and Raichle, M. E. (2007). Spontaneous fluctuations in brain activity observed with functional magnetic resonance imaging. *Nature Reviews Neuroscience 8*, 8(9):700–711.

Fox, P. T., Laird, A. R., and Lancaster, J. L. (2005a). Coordinate-based voxel-wise meta-analysis: Dividends of spatial normalization. Report of a virtual workshop. *Human Brain Mapping*, 25(1):1–5.

Fox, P. T., Laird, A. R., and Lancaster, J. L., editors (2005b). *Human Brain Mapping, Special Issue: Meta-Analysis in Functional Brain Mapping*, volume 25.

Fox, P. T. and Lancaster, J. L. (2002). Mapping context and content: the BrainMap model. *Nature Reviews Neuroscience*, 3:319–321.

Fox, P. T., Parsons, L. M., and Lancaster, J. L. (1998). Beyond the single study: functional/location metanalysis in cognitive neuroimaging. *Current Opinions in Neurobiology*, 8(2):178–187.

Fraley, C. (1998). Algorithms for model-based gaussian hierarchical clustering. *Journal on Scientific Computing*, 20(1):270–281.

Fraley, C. and Raftery, A. E. (1998). How many clusters? Which clustering method? Answers via model-based cluster analysis. *The Computer Journal*, 41:578–588.

Fraley, C. and Raftery, A. E. (1999). MCLUST: Software for model-based cluster analysis. *Journal of Classification*, 16:297–206.

Fraley, C. and Raftery, A. E. (2002). Model-based clustering, discriminant analysis, and density estimation. *Journal of the American Statistical Association*, 97(458):611–631.

Fraley, C. and Raftery, A. E. (2003). Enhanced software for model-based clustering, discriminant analysis, and density estimation: MCLUST. *Journal of Classification*, 20:263–286.

Frank, L. R., Buxton, R. B., and Wong, E. C. (1998). Probabilistic analysis of functional magnetic resonance imaging data. *Magnetic Resonance in Medicine*, 39:132–148.

Frison, L. and Pocock, S. J. (1992). Repeated measures in clinical trials: Analysis using mean summary statistics and its implications for design. *Statistics in Medicine*, 11:1685–1704.

Friston, K. J. (1994). Statistical parametric maps in functional imaging: A general linear approach. *Human Brain Mapping*, 2:189–210.

Friston, K. J., Fletcher, P., Joseph, O., Holmes, A. P., Rugg, M. D., and Turner, R. (1998). Event-related responses in fMRI: Characterising differential responses. *NeuroImage*, 7:30–40.

Friston, K. J., Frith, C. D., Fletcher, P., Liddle, P. F., and Frackowiak, R. S. J. (1996). Functional topography: Multidimensional scaling and functional connectivity in the brain. *Cerebral Cortex*, 6:165–164.

Friston, K. J., Glaser, D. E., Henson, R. N. A., Kiebel, S., Phillips, C., and Ashburner, J. (2002a). Classical and Bayesian inference in neuroimaging: Applications. *NeuroImage*, 16:483–512.

Friston, K. J., Harrison, L., and Penny, W. (2003). Dynamic causal modelling. *NeuroImage*, 19(4):1273–1302.

Friston, K. J., Holmes, A. P., and Worsley, K. J. (1999). Comments and controversies: How many subjects constitute a study? *NeuroImage*, 10:1–5.

Friston, K. J., Jezzard, P., and Turner, R. (1994). Analysis of functional MRI time series. *Human Brain Mapping*, 1:153–171.

Friston, K. J., Penny, W., Phillips, C., Kiebel, S., Hinton, G., and Ashburner, J. (2002b). Classical and Bayesian inference in neuroimaging: Theory. *NeuroImage*, 16:465–483.

Frith, C. D., Friston, K., Liddle, P. F., and Frackowiak, R. S. J. (1991). Willed action and the prefrontal cortex in man: A study with PET. *Proceedings of the Royal Society London: Biological Sciences*, 244(1311):241–246.

Gelman, A., Carlin, J., Stern, H. S., and Rubin, D. B. (2000). *Bayesian Data Analysis*. Chapman and Hall-CRC.

Genovese, C., Lazar, N., and Nichols, T. (2002). Thresholding of statistical maps in functional neuroimaging using the false discovery rate. *NeuroImage*, pages 870–878.

Genovese, C. R. (2000). A Bayesian time-course model for functional magnetic resonance imaging data. *Journal of the American Statistical Association*, 95(451):691–703.

Gibbons, L. E., Hearn, D. W., Pardalos, P. M., and Ramana, M. V. (1997). Continuous characterizations of the maximum clique problem. *Mathematics of Operations Research*, 22(3):754–768.

Gigerenzer, G. (1993). The superego, the ego and the id in statistical reasoning. In Keren, G. and Lewis, C., editors, *A Handbook for Data Analysis in the Behavioural Sciences: Methodological Issues*, pages 311–339. Hillsdale, NJ: Erlbaum.

Glover, G. H. (1999). Deconvolution of impulse response in event-related BOLD fMRI. *NeuroImage*, 9:416–429.

Gössl, C., Auer, D. P., and Fahrmeir, L. (2001a). Bayesian spatiotemporal inference in functional magnetic resonance imaging. *Biometrics*, 57:554–562.

Gössl, C., Fahrmeir, L., and Auer, D. P. (2001b). Bayesian modeling of the hemodynamic response function in BOLD fMRI. *NeuroImage*, 14:140–148.

Halchenko, Y. O., Hanson, S. J., and Pearlmutter, B. A. (2005). Multimodal integration: fMRI, MRI, EEG, MEG. pages 223–265. Taylor & Francis.

Halkidi, M., Batistakis, Y., and Vazirgiannis, M. (2001). On clustering validation techniques. *Journal of Intelligent Information Systems*, 17(2-3):107–145.

Hartley, H. (1958). Maximum likelihood estimation from incomplete data. *Biometrics*, 14:174–194.

Haxby, J. V., Gobbini, M. I., Furey, M. L., Ishai, A., Schouten, J. L., and Pietrini, P. (2001). Distributed and overlapping representations of faces and objects in ventral temporal cortex. *Science*, 293(5539):2425–2430.

Heckerman, D., Geiger, D., and Chickering, D. M. (1995). Learning Bayesian networks: The combination of knowledge and statistical data. *Machine Learning*, 20:197–243.

Henson, R. N. A., Price, C., Rugg, M. D., Turner, R., and Friston, K. J. (2002). Detecting latency differences in event-related BOLD responses: Application to words versus nonwords, and initial versus repeated face presentations. *NeuroImage*, 15:83–97.

Hofbauer, J. and Sigmund, K. (1988a). *Evolutionary Games and Population Dynamics*. Cambridge University Press, Cambridge.

Hofbauer, J. and Sigmund, K. (1988b). *The Theory of Evolution and Dynamical Systems*. Cambridge University Press, Cambridge.

Højen-Sørensen, P., Hansen, L., and Rasmussen, C. (2000). Bayesian modelling of fMRI time series. In Solla, S., Leen, T., and Müller, K.-R., editors, *Advances in Neural Information Processing Systems*, volume 12, pages 754–760. MIT Press.

Holmes, A. P. and Friston, K. J. (1998). Generalisability, random effects & population inference. *NeuroImage*, 7:S754.

Hopfield, J. J. (1982). Neural networks and physical systems with emergent collective computational abilities. *Proceedings of the National Academy of Science*, 79:2554–2558.

Hotelling, H. (1936). Relation between two sets of variates. *Biometrika*, 28:321–377.

Indefrey, P. and Levelt, W. (2000). The neural correlates of language production. In Gazzaniga, M., editor, *The new cognitive neurosciences*, pages 845–865. Cambridge, MA: MIT Press, 2 edition.

Jagota, A. (1995). Approximating maximum clique with a hopfield network. *IEEE Transactions on Neuralal Networks*, 6:724–735.

Jeffreys, H. (1935). Some tests of significance, treated by the theory of probability. In *Proceedings of the Cambridge Philosophical Society*, volume 31, pages 203–222.

Jeffreys, H. (1961). *Theory of Probability*. Oxford University Press, Oxford, 3rd edition.

Jezzard, P., Matthews, P., and Smith, S., editors (2002). *Functional Magnetic Resonance Imaging: An Introduction to Methods*. Oxford University Press, Oxford.

Kass, R. E. and Raftery, A. E. (1995). Bayes factors. *Journal of the American Statistical Association*, 90.

Kershaw, J., Ardekani, B. A., and Kanno, I. (1999). Application of Bayesian inference to fMRI data analysis. *IEEE Transactions on Medical Imaging*, 18(12):1138–1153.

Kiebel, S. J., Goebel, R., and Friston, K. J. (2000). Anatomically informed basis functions. *NeuroImage*, 11(6):656–667.

Kim, S.-G., Richter, W., and Ugurbil, K. (1997). Limitations of temporal resolution in functional MRI. *Magnetic Resonance in Medicine*, 37(4):631–636.

Kimura, M. (1958). On the change of population fitness by natural selection. *Heredity*, 12:145–167.

Klein, T. A., Endrass, T., Kathmann, N., Neumann, J., von Cramon, D. Y., and Ullsperger, M. (2007a). Neural correlates of error awareness. *NeuroImage*, 34(4):1774–1781.

Klein, T. A., Neumann, J., Reuter, M., Hennig, J., von Cramon, D. Y., and Ullsperger, M. (2007b). Genetically determined differences in learning from errors. *Science*, 318(5856):1642–1645.

Koechlin, E., Ody, C., and Kouneiher, F. (2003). The architecture of cognitive control in the human prefrontal cortex. *Science*, 302:1181–1185.

Krueger, J. (2001). Null hypothesis significance testing: On the survival of a flawed method. *American Psychologist*, 56(1):16–26.

Kruggel, F. and von Cramon, D. Y. (1999a). Modeling the hemodynamic response in single-trial functional MRI experiments. *Magnetic Resonance in Medicine*, 42:787–797.

Kruggel, F. and von Cramon, D. Y. (1999b). Temporal properties of the hemodynamic response in functional MRI. *Human Brain Mapping*, 8:259–271.

Krumbholz, K., Schoenwiesner, M., Ruebsamen, R., Zilles, K., Fink, G. R., and von Cramon, D. Y. (2005). Hierarchical processing of sound location and motion in the human brainstem and planum temporale. *European Journal of Neuroscience*, 21(1):230–238.

Laird, A. R., Fox, P. M., Price, C. J., Glahn, D. C., Uecker, A. M., Lancaster, J. L., Turkeltaub, P. E., Kochunov, P., and Fox, P. T. (2005a). ALE meta-analysis: Controlling the false discovery rate and performing statistical contrasts. *Human Brain Mapping*, 25(1):155–164.

Laird, A. R., Lancaster, J. L., and Fox, P. T. (2005b). BrainMap: the social evolution of a human brain mapping database. *Neuroinformatics*, 3(1):65–78.

Laird, A. R., McMillan, K. M., Lancaster, J. L., Kochunov, P., Turkeltaub, P. E., Pardo, J. V., and Fox, P. T. (2005c). A comparison of label-based and ALE meta-analysis in the Stroop task. *Human Brain Mapping*, 25(1):6–21.

Lange, N. (1997). Empirical and substantive models, the Bayesian paradigm, and meta-analysis in functional brain imaging. *Human Brain Mapping*, 5(4):259–263.

Lange, N. and Zeger, S. L. (1997). Non-linear Fourier time series analysis for human brain mapping by functional magnetic resonance imaging. *Applied Statistics*, 46:1–29.

Lee, A. T., Glover, G. H., and Meyer, C. H. (1995). Discrimination of large venous vessels in time-course spiral blood-oxygen-level-dependent magnetic-resonance functional neuroimaging. *Magnetic Resonance in Medicine*, 33:745–754.

Lee, P. M. (1997). *Bayesian Statistics: An Introduction*. Oxford University Press.

Liao, C. H., Worsley, K. J., Poline, J.-B., Aston, J. A. D., Duncan, G. H., and Evans, A. C. (2002). Estimating the delay of the response in fMRI data. *NeuroImage*, 16:593–606.

Liu, X., Banich, M. T., Jacobson, B. L., and Tanabe, J. L. (2004). Common and distinct neural substrates of attentional control in an integrated simon and spatial Stroop task as assessed by event-related fMRI. *NeuroImage*, 22(3):1097–1106.

Logothetis, N. K. (2003). The underpinnings of the BOLD functional magnetic resonance imaging signal. *The Journal of Neuroscience*, 23(10):3963–3971.

Logothetis, N. K. (2008). What we can do and what we cannot do with fMRI. *Nature*, 453:869–878.

Lohmann, G. and Bohn, S. (2002). Using replicator dynamics for analyzing fMRI data of the human brain. *IEEE Transactions on Medical Imaging*, 21(5):485–492.

Lohmann, G., Müller, K., Bosch, V., Mentzel, H., Hessler, S., Chen, L., Zysset, S., and von Cramon, D. Y. (2001). LIPSIA - A new software system for the evaluation of functional magnetic resonance images of the human brain. *Computerized Medical Imaging and Graphics*, 25(6):449–457.

Lohmann, G., Neumann, J., Mueller, K., Lepsien, J., and Turner, R. (2008). The multiple comparison problem in fMRI - a new method based on anatomical priors. In Hamarneh, G. and Abugharbieh, R., editors, *Medical Image Computing and Computer-Assisted Intervention Workshop on Analysis of Functional Medical Image Data (MICCAI functional)*.

Lohmann, G., Schubotz, R. I., and von Cramon, D. Y. (2002). Visualization of fMRI data using coronal flat maps. 8th International Conference on Functional Mapping of the Human Brain, June 2-6, 2002, Sendai, Japan. NeuroImage, Vol. 16, No. 2.

Maitra, R., Roys, S. R., and Gullapalli, R. P. (2002). Test-retest reliability estimation of functional MRI data. *Magnetic Resonance in Medicine*, 48:62–70.

Marrelec, G., Benali, H., Ciuciu, P., Plgrini-Issac, M., and Poline, J.-B. (2003). Robust Bayesian estimation of the hemodynamic response function in event-related BOLD fMRI using basic physiological information. *Human Brain Mapping*, 19(1):1–17.

McGonigle, D. J., Howseman, A. M., Athwal, B. S., Friston, K. J., Frackowiak, R. S. J., and Holmes, A. P. (2000). Variability in fMRI: An examination of intersession differences. *NeuroImage*, 11:708–734.

McIntosh, A. R. and Gonzalez-Lima, F. (1994). Structural equation modeling and its application to network analysis in functional brain imaging. *Human Brain Mapping*, 2:2–22.

McKeown, M. J., Jung, T.-P., Makeig, S., Brown, G., Kindermann, S. S., Lee, T.-W., and Sejnowski, T. J. (1998a). Spatially independent activity patterns in functional MRI data during the stroop color-naming task. *Proc Natl Acad Sci USA*, 95(3):803–810.

McKeown, M. J., Makeig, S., Brown, G. G., Jung, T.-P., Kindermann, S. S., and Sejnowski, T. J. (1998b). Analysis of fMRI by blind separation into independent spatial components. *Human Brain Mapping*, 6(3):160–188.

Menon, A., Mehrotra, K., Mohan, C. K., and Rank, S. (1995a). Optimization using replicators. In Eshelman, L., editor, *Proceedings of the Sixth International Conference on Genetic Algorithm*, pages 209–216, San Francisco, CA. Morgan Kaufmann.

Menon, R. S. and Kim, S.-G. (1999). Spatial and temporal limits in cognitive neuroimaging with fMRI. *Trends in Cognitive Science*, 3(6):207–216.

Menon, R. S., Ogawa, S., Hu, X., Strupp, J. S., Andersen, P., and Ugurbil, K. (1995b). BOLD based functional MRI at 4 Tesla includes a capillary bed contribution: Echoplanar imaging mirrors previous optical imaging using intrinsic signals. *Magnetic Resonance in Medicine*, 33:453–459.

Menz, M. M., Neumann, J., Mueller, K., and Zysset, S. (2006). Variability of the BOLD response over time: An examination of within-session differences. *NeuroImage*, 32(3):1185–1194.

Meunier, D., Achard, S., Morcom, A., and Bullmore, E. (2009). Age-related changes in modular organization of human brain functional networks. *NeuroImage*, 44(3):715–723.

Miezin, F. M., Maccotta, L., Ollinger, J. M., Petersen, S. E., and Buckner, R. L. (2000). Characterizing the hemodynamic response: Effects of presentation rate, sampling procedure, and the possibility of ordering brain activity based on relative timing. *NeuroImage*, 11:735–759.

Müller, K., Lohmann, G., Zysset, S., and von Cramon, D. Y. (2003). Wavelet statistics of functional MRI data and the general linear model. *Journal of Magnetic Resonance Imaging*, 17(1):20–30.

Neal, R. M. and Hinton, G. E. (1998). A view of the EM algorithm that justifies incremental, sparse, and other variants. In Jordan, M. I., editor, *Learning in Graphical Models*, pages 355–368. Kluwer Academic Publishers, Norwell MA.

Neumann, J., Fox, P. T., Turner, R., and Lohmann, G. (2010). Learning partially directed functional networks from meta-analysis imaging data. *NeuroImage*, 49(2):1372–1384.

Neumann, J. and Lohmann, G. (2003). Bayesian second-level analysis of functional magnetic resonance images. *NeuroImage*, 20(2):1346–1355.

Neumann, J., Lohmann, G., Derrfuss, J., and von Cramon, D. Y. (2005). The meta-analysis of functional imaging data using replicator dynamics. *Human Brain Mapping*, 25(1):165–173.

Neumann, J., Lohmann, G., Zysset, S., and von Cramon, D. Y. (2003). Within-subject variability of BOLD response dynamics. *NeuroImage*, 19(3):784–796.

Neumann, J., von Cramon, D. Y., Forstmann, B. U., Zysset, S., and Lohmann, G. (2006). The parcellation of cortical areas using replicator dynamics in fMRI. *NeuroImage*, 32(1):208–219.

Neumann, J., von Cramon, D. Y., and Lohmann., G. (2008). Model-based clustering of meta-analytic functional imaging data. *Human Brain Mapping*, 29(2):177–192.

Nielsen, F. A. and Hansen, L. K. (2002). Modeling of activation data in the BrainMap database: Detection of outliers. *Human Brain Mapping*, 15(3):146–156.

Noll, D. C., Genovese, C. R., Nystrom, L., Forman, S., Eddy, W. F., and Cohen, J. (1997). Estimating test-retest reliability in fMRI II: Application to sensory-motor and cognitive activation. *Magnetic Resonance in Medicine*, 38:508–517.

Oakes, M. (1986). *Statistical Inference: A commentary for the social and behavioral sciences*. John Wiley & Sons, New York.

Oakes, T. R., Johnstone, T., Ores Walsh, K. S., Greischar, L. L., Alexander, A. L., Fox, A. S., and Davidson, R. J. (2005). Comparison of fMRI motion correction software tools. *NeuroImage*, 28(3):529–543.

Ogawa, S., Lee, T. M., Kay, A. R., and Tank, D. W. (1990). Brain magnetic resonance imaging with contrast dependent on blood oxygenation. *Proc Natl Acad Sci USA*, 87:9868–9872.

Ogawa, S., Menon, R., Tank, D. W., Kim, S.-G., Merkle, H., Ellermann, J., and Ugurbil, K. (1993). Functional brain mapping by blood oxygen level-dependent contrast functional magnetic resonance imaging. A comparison of signal characteristics with a biophysical model. *Biophysical Journal*, 64:803–812.

Pardalos, P. M. and Vavasis, S. A. (1991). Quadratic programming with one negative eigenvalue is NP-hard. *Journal of Global Optimization*, 1(1):15–22.

Pearl, J. (2000). *Causality*. Cambridge University Press.

Pearson, K. (1901). On lines and planes of closest fit to systems of points in space. *The London, Edinburgh and Dublin Philosophical Magazine and Journal of Science*, 2:559–572.

Pelillo, M. (2001). Replicator dynamics in combinatorial optimization. In Floudas, C. A. and Pardalos, P. M., editors, *Encyclopedia of Optimization*, volume 5, pages 23–35. Kluwer Academic Publishers, Boston.

Pelillo, M., Siddiqi, K., and Zucker, S. W. (1999). Matching hierarchical structures using association graphs. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 21(11):1105–1120.

Petersson, K. M., Nichols, T. E., Poline, J.-B., and Holmes, A. P. (1999). Statistical limitations in functional neuroimaging. I. Non-inferential methods and statistical models. *Phil. Trans. R. Soc. Lond. B*, 354:1239–1260.

Phan, K., Wager, T., Taylor, S. F., and Liberzon, I. (2002). Functional neuroanatomy of emotion: A meta-analysis of emotion activation studies in PET and fMRI. *NeuroImage*, 16(2):331–348.

Pollmann, S. and Morillo, M. (2003). Left and right occipital cortices differ in their response to spatial cueing. *NeuroImage*, 18:273–283.

Posner, M. I., Petersen, S. E., Fox, P. T., and Raichle, M. E. (1988). Localization of cognitive operations in the human brain. *Science*, 240(4859):1627–1631.

Rencher, A. C. (2002). *Methods of Multivariate Analysis*. John Wiley & Sons, New York, 2nd edition.

Richardson, S. and Green, P. J. (1997). On Bayesian analysis of mixtures with unknown number of components. *Journal of the Royal Statistical Society, Series B*, 59:731–792.

Ridderinkhof, K. R., Ullsperger, M., Crone, E. A., and Nieuwenhuis, S. (2004). The role of the medial frontal cortex in cognitive control. *Science*, 306:443–447.

Rombouts, S., Barkhof, F., Hoogenraad, F., Sprenger, M., and Scheltens, P. (1998). Within-subject reproducibility of visual activation patterns with functional magnetic resonance imaging using multislice echo planar imaging. *Magnetic Resonance Imaging*, 16(2):105–113.

Ruschemeyer, S. A., Fiebach, C. J., Kempe, V., and Friederici, A. D. (2005). Processing lexical semantic and syntactic information in first and second language: fMRI evidence from German and Russian. *Human Brain Mapping*, 25(2):266–286.

Schacter, D. L., Buckner, R. L., Koutstaal, W., Dale, A. M., and Rosen, B. R. (1997). Late onset of anterior prefrontal activity during true and false recognition: An event-related fMRI study. *NeuroImage*, 6(4):259–269.

Schroeter, M. L., Raczka, K., Neumann, J., and von Cramon, D. Y. (2007). Towards a nosology for frontotemporal lobar degenerations - A meta-analysis involving 267 subjects. *NeuroImage*, 36(3):497–510.

Schroeter, M. L., Raczka, K., Neumann, J., and von Cramon, D. Y. (2008). Neural networks in frontotemporal dementia - A meta-analysis. *Neurobiology of Aging*, 29(3):418–426.

Schroeter, M. L., Stein, T., Maslowski, N., and Neumann, J. (2009). Neural correlates of alzheimer's disease and mild cognitive impairment: A systematic and quantitative meta-analysis involving 1,351 patients. *NeuroImage*, 47(4):1196–1206.

Schuster, P. and Sigmund, K. (1983). Replicator dynamics. *Journal of Theoretical Biology*, 100:533–538.

Schwarz, A. J., Gozzi, A., and Bifone, A. (2008). Community structure and modularity in networks of correlated brain activity. *Magnetic Resonance Imaging*, 26(7):914–920.

Schwarz, G. (1978). Estimating the dimension of a model. *The Annals of Statistics*, 6:461–464.

Scott, D. W. (1992). *Multivariate Density Estimation: Theory, Practice, and Visualization*. John Wiley and Sons, New York.

Searle, S. R., Casalla, G., and McCulloch, C. E. (1992). *Variance Components*. John Wiley & Sons, New York.

Seber, G. A. F. (1977). *Linear Regression Analysis*. John Wiley & Sons, New York.

Seber, G. A. F. and Wild, C. J. (1989). *Nonlinear Regression*. John Wiley & Sons, New York.

Shen, X. and Ye, J. (2002). Adaptive model selection. *Journal of the American Statistical Association*, 19:210–221.

Silverman, B. W. (1985). *Density Estimation for Statistics and Data Analysis*. Chapman and Hall, London.

Smith, S. M., Fox, P. T., Miller, K. L., Glahn, D. C., Fox, P. . M., Mackay, C. E., Filippini, N., Watkins, K. E., Toro, R., Laird, A. R., and Beckmann, C. F. (2009). Correspondence of the brain's functional architecture during activation and rest. *Proceedings of the National Academy of Sciences of the United States of America*, 106(31):13040–13045.

Spiegelhalter, D. J., Best, N. G., Carlin, B. R., and van der Linde, A. (2002). Bayesian measures of model complexity and fit. *J. of the Royal Statist. Soc. Series B*, 64:583–616.

Stone, J. V., Porrill, J., Porter, N. R., and Wilkinson, I. D. (2002). Spatiotemporal independent component analysis of event-related fMRI data using skewed probability density functions. *NeuroImage*, 15:407–421.

Stroop, J. R. (1935). Studies of inference in serial verbal reactions. *Journal of Experimental Psychology*, 18:643–662.

Svensén, M. and Bishop, C. M. (2004). Robust Bayesian mixture modelling. *Neurocomputing*, 64:235–252.

Tabelow, K., Piech, V., Polzehl, J., and Voss, H. U. (2009). High-resolution fMRI: Overcoming the signal-to-noise problem. *Journal of Neuroscience Methods*, 178(2):357–365.

Talairach, J. and Tournoux, P. (1988). *Co-planar Stereotaxic Atlas of the Human Brain*. Thieme, Stuttgart.

Taylor, P. and Jonker, L. (1978). Evolutionary stable strategies and game dynamics. *Mathematical Biosciences*, 40:145–156.

Tervaniemi, M., Szameitat, A. J., Kruck, S., Schroger, E., Alter, K., De Baene, W., and Friederici, A. D. (2006). From air oscillations to music and speech: Functional magnetic resonance imaging evidence for fine-tuned neural networks in audition. *Journal of Neuroscience*, 26(34):8647–8652.

Thierry, G., Boulanouar, K., Kherif, F., Ranjeva, J. P., and Demonet, J. F. (1999). Temporal sorting of neural components underlying phonological processing. *NeuroReport*, 10(12):2599–2603.

Tillmann, B., Koelsch, S., Escoffier, N., nd P. Lalitte, E. B., Friederici, A., and von Cramon, D. (2006). Cognitive priming in sung and instrumental music: Activation of inferior frontal cortex. *NeuroImage*, 31(4):1771–1782.

Torgerson, W. S. (1952). Multidimensional scaling. I: Theory and method. *Psychometrika*, 17:401–419.

Turkeltaub, P. E., Eden, G. F., Jones, K. M., and Zeffiro, T. A. (2002). Meta-analysis of the functional neuroanatomy of single-word reading: Method and validation. *NeuroImage*, 16:765–780.

Ullsperger, M. and von Cramon, D. Y. (2004). Neuroimaging of performance monitoring: Error detection and beyond. *Cortex*, 40(4-5):593–604.

Volz, K. G. and von Cramon, D. Y. (2006). What neuroscience can tell about intuitive processes in the context of perceptual discovery. *Journal of Cognitive Neuroscience*, 18(12):2077–2087.

Voultsidou, M., Dodel, S., and Herrmann, J. M. (2004). Emergence of functional connectivity clusters via hopfield neural networks. 10th Annual Meeting of the Organization for Human Brain Mapping (OHBM), Budapest, June 2004.

Voultsidou, M., Dodel, S., and Herrmann, J. M. (2005). Neural networks approach to clustering of activity in fmri data. *IEEE Transactions on Medical Imaging*, 24(8):987–996.

Wager, T. D., Jonides, J., and Reading, S. (2004). Neuroimaging studies of shifting attention: a meta-analysis. *NeuroImage*, 22:1679–1693.

Wager, T. D., Lindquist, M., and Kaplan, L. (2007). Meta-analysis of functional neuroimaging data: current and future directions. *Social Cognitive and Affective Neuroscience*, 2(2):150–158.

Waldvogel, D., van Gelderen, P., Immisch, I., Pfeiffer, C., and Hallett, M. (2000). The variability of serial fMRI data: Correlation between a visual and a motor task. *NeuroReport*, 11(17):3843–3847.

Weiss, Y. (1999). Segmentation using eigenvectors: a unifying view. In *Proceedings IEEE International Conference on Computer Vision*, pages 975–982.

Welchew, D. E., Honey, G. D., Sharma, T., Robbins, T. W., and Bullmore, E. T. (2002). Multidimensional scaling of integrated neurocognitive function and schizophrenia as a disconnexion disorder. *NeuroImage*, 17(3):1227–1239.

Worsley, K. J. (2001). Statistical analysis of activation images (Chap. 14). In Jezzard, P., Matthews, P. M., and Smith, S. M., editors, *Functional MRI : An Introduction to Methods*, pages 251–270. Oxford University Press.

Worsley, K. J. and Friston, K. J. (1995). Analysis of fMRI time-series revisited – again. *NeuroImage*, 2:173–181.

Worsley, K. J., Liao, C., Aston, J., Petre, V., Duncan, G. H., Morales, F., and Evans, A. C. (2002). A general statistical analysis for fMRI data. *NeuroImage*, 15:1–15.

Worsley, K. J., Taylor, J. E., Tomaiuolo, F., and Lerch, J. (2004). Unified univariate and multivariate random field theory. *NeuroImage*, 23(Supplement 1):S189–S195.

Zarahn, E., Aguirre, G. K., and D'Esposito, M. (1997). Empirical analyses of BOLD fMRI statistics. *NeuroImage*, 5:179–197.

Zysset, S., Müller, K., Lohmann, G., and von Cramon, D. Y. (2001). Color-word matching stroop task: Separating interference and response conflict. *NeuroImage*, 13:29–36.

Zysset, S., Schroeter, M. L., Neumann, J., and von Cramon, D. Y. (2007). Stroop interference, hemodynamic response and aging: An event-related fMRI study. *Neurobiology of Aging*, 28(6):937–946.

Zysset, S., Wendt, C. S., Volz, K. G., Neumann, J., Huber, O., and von Cramon, D. Y. (2006). The neural implementation of multi-attribute decision making: A parametric fMRI study with human subjects. *NeuroImage*, 31(3):1380–1388.

List of Figures

147

List of Tables

Declaration

Selbständigkeitserklärung

Hiermit erkläre ich, die vorliegende Habilitationsschrift selbständig und ohne unzulässige fremde Hilfe angefertigt zu haben. Ich habe keine anderen als die angeführten Quellen und Hilfsmittel benutzt und sämtliche Textestellen, die wörtlich oder sinngemäß aus veröffentlichten oder unveröffentlichten Schriften entnommen wurden, und alle Angaben, die auf mündlichen Auskünften beruhen, als solche kenntlich gemacht. Ebenfalls sind alle von anderen Personen bereitgestellten Materialien oder erbrachten Dienstleistungen als solche gekennzeichnet.

Leipzig, den 16. Juli 2012

MPI Series in Human Cognitive and Brain Sciences: