TECHNISCHE UNIVERSITÄT WIEN
Vienna University of Technology

DIPLOMARBEIT

Reliability of Dynamic Causal Modelling of Brain Connectivity

Ausgeführt am Institut für
Analysis und Scientific Computing
der Technischen Universität Wien

in Kooperation mit dem Zentrum für
Medizinische Physik und Biomedizinische Technik
der Medizinischen Universität Wien

unter der Anleitung von
Ao.Univ.Prof. Dipl.-Ing. Dr.techn. Felix Breitenecker
Assoc.Prof. Priv.-Doz. Dipl.-Ing. Dr. Christian Windischberger
Dipl.-Ing. Günter Schneckenreither

durch
Nicole Geissberger, BSc
Humboldtgasse 33/44, 1100 Wien

Wien, Oktober 2017

TECHNISCHE
UNIVERSITÄT
WIEN
Vienna University of Technology

Master Thesis

Reliability of Dynamic Causal Modelling of Brain Connectivity

carried out at the Institute for
Analysis and Scientific Computing
Vienna University of Technology

in cooperation with the Center for
Medical Physics and Biomedical Engineering
Medical University of Vienna

under guidance of
Ao.Univ.Prof. Dipl.-Ing. Dr.techn. Felix Breitenecker
Assoc.Prof. Priv.-Doz. Dipl.-Ing. Dr. Christian Windischberger
Dipl.-Ing. Günter Schneckenreither

by
Nicole Geissberger, BSc
Humboldtgasse 33/44, 1100 Wien

Vienna, October 2017

Zusammenfassung

Funktionelle Magnetresonanztomographie (fMRT) ist ein Bildgebungsverfahren, welches verwendet wird, um die Funktionsweise des menschlichen Gehirns bei bestimmten Aufgabenstellungen zu untersuchen. Hierbei wird unter experimentellen Bedingungen in festen Zeitabständen ein 3D-Bild des Gehirns aufgenommen. Kleinste Änderungen im Hirnstoffwechsel erlauben Rückschlüsse über die neuronale Aktivität während kognitiver Experimente.

Für die Analyse der so gewonnenen Daten existieren verschiedene Ansätze. Üblicherweise wird ein lineares Modell (GLM) aufgestellt, in welchem ein linearer Zusammenhang zwischen Stimulus und Hirnaktivität dargestellt wird. Mit dieser robusten, sehr einfachen Methode können jedoch Zusammenhänge zwischen verschiedenen Bereichen des Gehirns nicht berücksichtigt werden. In dieser Arbeit liegt der Fokus deshalb auf Dynamic Causal Modelling (DCM), einer Methode, die für die Untersuchung von (Hypothesen zu) funktionellen Netzwerken im Gehirn verwendet wird.

Da kürzlich an verschiedenen Stellen Kritik an der Reproduzierbarkeit dieser Methoden geäußert wurden, untersucht diese Arbeit Techniken und Ansätze um die Wiederholbarkeit von Ergebnissen quantifizierbar zu machen. Dazu wurden hochaufgelöste funktionelle MRT-Daten von vierzehn gesunden Probanden mithilfe eines Ultra-Hochfeldtomographen erhoben, während die Probanden eine Gesichts- bzw. Emotionserkennungsaufgabe erfüllen mussten. Ziel der Studie war einerseits, Ergebnisse aus früheren Studien zu reproduzieren, andererseits die Stabilität der Ergebnisse bei wiederholten Messungen zu untersuchen.

Während die Resultate aus dem klassischen Ansatz des linearen Modells in dieser Studie eine sehr hohe Reproduzierbarkeit aufwiesen und sich auch mit den Ergebnissen anderer Untersuchungen deckten, konnten die DCM Ergebnisse aus einer früheren Studie nicht reproduziert werden. Ebenso waren die Ergebnisse über die wiederholten Durchläufe nicht konsistent genug, um diesbezüglich Schlüsse ziehen zu können. Weitere Untersuchungen sind nötig um herauszufinden, ob diese unschlüssigen Ergebnisse schlecht gewählten Modellhypothesen, einer zu kleinen Stichprobe oder anderen Faktoren zuzuschreiben sind.

Keywords: funktionelle Magnetresonanztomographie (fMRT), Dynamic Causal Modelling (DCM), Neuroimaging, Hirnnetzwerke, Stabilität, Reproduzierbarkeit

Abstract

Functional magnetic resonance imaging (fMRI) is a neuroimaging technique used to investigate the inner mechanisms of the human brain when presented with certain tasks. A sequence of 3D-images of the brain is acquired during certain experimental conditions to detect subtle changes in brain metabolism, which allow for inferences on neuronal activity during cognitive experiments.

In order to analyse the acquired data, numerous approaches can be used, depending on the nature of the research question. Customarily, the relationship between stimulus and brain activity is assumed linear and is thus modelled in the General Linear Model framework. However, this simple yet robust approach does not allow for modelling interrelations between different parts of the brain. In this work, we are focusing on a particular method used for investigating dynamic functional connectivity of brain regions and changes related to certain task conditions, namely Dynamic Causal Modelling (DCM).

After recent criticism on this method, we have investigated the test-retest reliability of Dynamic Causal Modelling. We acquired fMRI data on fourteen healthy subjects at ultra-high field while the subjects were repeatedly performing a facial emotion processing task. The task was repeated six times within a measurement session, the whole session was repeated after 14 days. With this study, we aimed to reproduce reported findings and assess the reproducibility of our results after repeated measurements.

While the common approach using the General Linear Model yielded results showing high reproducibility and also coinciding with the results of other groups, the DCM results of a previous study could not be reproduced using this data set. Results over repeated measurements were also inconsistent.

Further investigations are necessary to unveil the reason for these inconclusive results – whether the results may be attributed to poorly chosen model hypotheses, an undersized sample or other factors.

Key words: functional magnetic resonance imaging (fMRI), Dynamic Causal Modelling (DCM), neuroimaging, brain networks, reproducibility, stability, emotion discrimination task

Contents

1 Introduction **7**
- 1.1 Functional magnetic resonance imaging (fMRI) 7
- 1.2 Research question 8
- 1.3 Contents of this work 9

2 Models and modelling aspects **11**
- 2.1 Hemodynamic model – the Balloon model 13
- 2.2 General Linear Model (GLM) 15
  - 2.2.1 Combination of GLM and the Balloon model 16
- 2.3 Models in Dynamic Causal Modelling (DCM) 16
  - 2.3.1 Combination of DCM and the Balloon model 18

3 Methods for fMRI data processing **23**
- 3.1 Pre-processing the data 24
- 3.2 Statistical Parametric Mapping (SPM) 24
  - 3.2.1 Statistical inference 26
  - 3.2.2 Group inferences 29
  - 3.2.3 Criticism 30
- 3.3 Dynamic Causal Modelling 30
  - 3.3.1 Biophysical constraints 31
  - 3.3.2 Maximizing model evidence 33
  - 3.3.3 Parameter estimation 35
  - 3.3.4 Inference strategies 41
  - 3.3.5 Criticism 41
- 3.4 Reliability and repeatability 43

4 Reliability study **47**
- 4.1 Motivation 47
- 4.2 Experimental setup 48
- 4.3 Reliability of GLM results 51
  - 4.3.1 Results 52
- 4.4 Reliability of DCM of emotion processing 63
  - 4.4.1 Results 66

5 Discussion / further work **71**
5.1 On the GLM results 71
5.2 On the DCM results 72

6 Acknowledgements **75**

Chapter 1

Introduction

In this work, I will present a short introduction to functional magnetic resonance imaging and the analysis tools used to explore the inner workings of the human brain, with the main focus on a framework for investigating functional networks in the brain called Dynamic Causal Modelling.

1.1 Functional magnetic resonance imaging (fMRI)

Functional magnetic resonance imaging (fMRI) is a neuroimaging technique for detecting subtle changes in brain metabolism during neural activity. It allows for non-invasive assessment of processes inside the living brain and is therefore widely used in neuroscience. Applications of fMRI range from brain mapping (i.e. functional localization of brain regions associated with phenomena of interest) via research on brain connectivity to multimodal integration with other imaging (e.g. electroencephalography, EEG) and stimulation tools (e.g. transcranial magnetic stimulation, TMS).

The principle of fMRI is based on changes in cerebral blood supply. Neuronal activity causes an increase in local energy demand. This increase is met with an increase in inflow of highly oxygenated arterial blood, overcompensating the energy demand. Therefore, blood oxygenation increases and subsequently, deoxygenated hemoglobin is reduced. Since deoxygenated hemoglobin is more paramagnetic compared to oxygenated hemoglobin (oHb), magnetic field variations around dHb show reduced MR signal amplitudes (Ogawa et al., 1990).

This phenomenon is known as the blood oxygenation level dependent (BOLD) effect. This was first discovered as signal dropouts occurred around vials filled with venous blood (Ogawa et al., 1990). Logothetis et al. (2001) further uncovered the relationship between the measured fMRI signal and un-

derlying neural activity in monkeys, suggesting that BOLD signal changes reflect the input and intracortical processing of a given area.

Although fMRI provides us with an indirect measure based on changes in blood supply, differences in neuronal activation levels can be assessed via fMRI. As more strongly activated areas consume more oxygen, the MR signal is altered. The characteristic shape of the signal change is referred to as hemodynamic response and reaches its peak 3 to 6 seconds after neuronal activation.

Usual task-based fMRI experiments consist of several trials. During these trials, participants are usually exposed to visual or auditory stimuli. Changes in BOLD signal are assessed for each trial and then generally averaged over all trials of experimental conditions. The time span between the individual stimuli is called inter-stimulus interval (ISI) and must be chosen carefully to avoid overlapping of the hemodynamic response.

Most often, MR sequences using echo-planar imaging (EPI) are used to repeatedly acquire images of the same cortical structures at a certain time interval, which is referred to as Repetition Time (TR) - this is the time that passes between each 3D image acquisition. Using this sequence of images, changes over time can be made visible.

The temporal resolution generally ranges from less than a second to several seconds for whole-brain images with spatial resolution in millimetres, both spatial and temporal resolution increasing with field strength (in Tesla) of the MRI scanner.

1.2 Research question

Over the last years, functional MRI has become a widely applied method for non-invasive research on brain function. However, reproducibility and robustness of the activation maps obtained remain crucial issues.

Previous studies have targeted test-retest reliability of certain fMRI results, however only few long-term studies on processing of emotional faces in the amygdala exist. Plichta et al. (2012) has targeted test-retest reliability in a number of cognitive and emotive tasks, showing e.g. high inter-session reliability over all subjects, however low within-subject reliability using the emotional faces task.

However, apart from thresholding considerations, the standard analysis approaches for fMRI have their limitations, especially in research on functional

brain networks. One hypothesis-driven method for analyzing intermodulations and co-dependencies between brain regions is Dynamic Causal Modelling (DCM). For more details, see section 2.3 and 3.3.

Recent criticism on the validity of this framework per se (Lohmann et al., 2012) has emphasized the necessity for further research on reproducibility of the results obtained using this method.

Whereas Frässle et al. (2015) and Frässle et al. (2016) have investigated test-retest reliability of DCM results using a motor task and a face processing task, the networks investigated were limited to motor regions, face processing and visual stimulus processing regions.

In this work, we aimed to investigate the test-retest reliability of the facial emotion processing network in the human brain, including structures such as the amygdalae in both hemispheres, fusiform face areas and prefrontal areas. With the prerequisite of obtaining stable and repeatable neuronal activity within the emotion processing network during this task, we expected to reproduce DCM results from previous studies (Sladky et al., 2015).

1.3 Contents of this work

First, I will give a short overview of the models used for the analyses in chapter 2.

In chapter 3, I will give an introduction on various processing steps after acquiring fMRI data and go into more detail on these steps described. Besides the classic data analysis approach, which is a method unrestrained by prior knowledge and theories on brain networks (see section 3.2), I will introduce a more complex hypothesis-driven method for analysing effective connectivity between brain regions (Dynamic Causal Modelling, see section 3.3).

Although fMRI has provided valuable insights into a plethora of cognitive phenomena, recent critique questioned the validity of various results obtained. To address this issue, I have conducted an experiment on the stability of the methods presented and have investigated this issue by acquiring and analyzing data of 14 healthy volunteers performing a task involving the processing of images of human faces showing different emotions. The results of this study are presented in chapter 4.

Finally, I will discuss the results obtained and knowledge gained with the experiment, as well as further research options beyond the scope of this work

in chapter 5.

Chapter 2

Models and modelling aspects

The term modelling can be used for various meanings. In the broadest sense, scientific modelling refers to making a part of the real world accessible by selecting features most relevant for a certain purpose. Models are essential for our understanding of the world, as the world itself can not be apprehended in its whole complexity. Per definition, there are no "true" models, as it is usually impossible to consider every detail and every influence there might be, models being therefore reductionistic per se.

A reference frame for the classification of models was defined by Goldsmith (1972), where he described several notions of models. Starting with the **conceptual model**, which is the most commonplace model, being present in practically any verbal description, but also in flow diagrams etc.

The next terms are **in vivo** and **in vitro** models. In vivo models describe live organisms sufficiently similar in a certain scope, e.g. laboratory animals for studying cancer growth, ultimately with the goal of (human) cancer research. In vitro models are characterized by conditions outside of those of direct interest. This can be, e.g., scale models of machines showing similar behaviour, chemical reactions in isolated set-ups comparable to those within living organisms, etc. In vivo and in vitro models are both physical models that face certain limitations, not only by simplifications and differences, but also by the materials available.

The last models described by Goldsmith are **in silico** models – mathematical models using mathematical concepts and language to describe a system, expressed in equations, summarising relationships of quantifiable effects, i.e. properties that can be put in numbers.

As lies in the nature of models, there is no "true" model – a famous aphorism quoted in this context is "all models are wrong" by Box (1976). Particularly, it is important to assess the level of detail needed for a certain purpose.

> Since all models are wrong the scientist cannot obtain a "correct" one by excessive elaboration. On the contrary, following William of Occam, he should seek an economical description of natural phenomena. Just as the ability to devise simple but evocative models is the signature of the great scientist, so overelaboration and overparameterization is often the mark of mediocrity. (Box, 1976)

Avoiding overfitting and seeking a good trade-off between model accuracy and generalizability are the hallmark of scientific modelling, parsimony in model structure and parameters is discussed as one of the main aspects of good scientific methods, alongside flexibility, practicability and worrying selectively – i.e. focusing on most important aspects (Box, 1976).

For scientific and engineering applications, a model's purpose is usually to describe the behaviour of a highly complex system – with the goal to aid in decision making processes. Models may be used to set up hypotheses on the behaviour of a complex system; these hypotheses can be tested and researched on, revised and, if necessary, discarded (Murray-Smith, 2012).

Mathematical modelling usually refers to creating a set of (differential) equations with certain well-defined input parameters, which describe the system in a manner sufficiently accurate for the defined purpose. Being described using mathematical concepts, these can thus be classified as in silico models.

In this work, however, pre-defined equations are used. The term "modelling" refers to formulating hypotheses on the influencing variables and parameters. Parameters are then estimated combining "classic" mathematical modelling strategies with statistical methods.

In the following sections, I will introduce the different models used in the methods chapter (chapter 3) for inference on functional MRI data.

On the one hand, as we do not measure neuronal activity directly, but through the hemodynamic response, we need a model to set a stimulus and the resulting neuronal activation in relation with the measured signal. For this purpose, a hemodynamic model (the Balloon model) will be introduced in the next chapter.

On the other hand, we need to model the magnitude of the signal change during certain conditions – i.e. the brain activity – for each brain region, so

we can identify which cortical areas are involved in certain tasks. This is done by setting up a General Linear Model (GLM) for each volume element (voxel) of the 3D image.

Additionally, we want to assess intermodulations and co-dependencies between certain brain regions. For this purpose, a framework for testing hypotheses on such connectivity – Dynamic Causal Modelling (DCM) – will be introduced.

2.1 Hemodynamic model – the Balloon model

The basis of the fMRI signal is the BOLD effect (Ogawa et al., 1990), which links changes in blood oxygenation to increased neuronal activity. The Balloon model accounts for the non-linear behaviour of the stimulus-induced hemodynamic response.

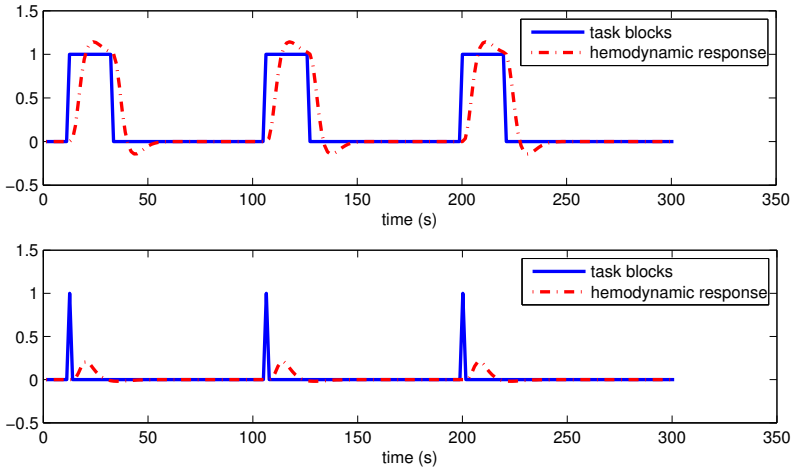

Figure 2.1: Hemodynamic response function corresponding to task blocks with duration of 20 seconds (above) or events (below). Data created using the `SPM` software package.

We need to model the hemodynamic response using only one single input – which is the stimulus. This is implemented in the so-called Balloon model, which has been shown to provide a sufficient framework for accounting for the non-linear changes of the BOLD signal following neuronal activity or stimulation (see Friston et al. (2000)). This model is described using following four differential equations and is depicted schematically in Figure 2.2.

Definition 1. The Balloon model is made up of following 4 differential equations describing the signal s, the inflow of blood f_{in}, blood volume v

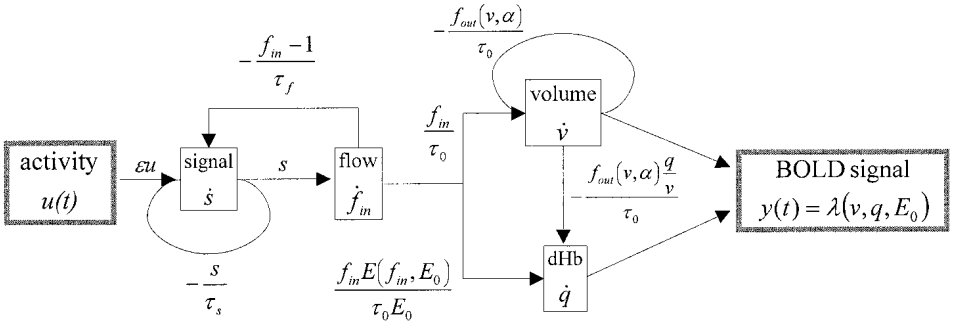

Figure 2.2: Balloon model of hemodynamic response following neuronal activity with single input (activity) and single output (BOLD signal). There are 4 state variables: s, f_{in}, v, q for the signal, inflow, volume and deoxyhemoglobin. Figure taken from Friston et al. (2000). Note that the biophysical parameters are re-named in definition 1 for consistency with section 3.3.

and deoxyhemoglobin q. It is assumed that a certain flow inducing signal s and the change in regional cerebral blood flow (rCBF) f_{in} are linked linearly,

$$\dot{f}_{\text{in}} = s. \tag{2.1}$$

The flow inducing signal is generated by neuronal responses to the stimulus function $u(t)$,

$$\dot{s} = \epsilon u(t) - \kappa s - \gamma(f_{\text{in}} - 1) \ , \tag{2.2}$$

where ϵ is a parameter that represents the efficacy the stimulus leads to signal increase; κ and γ are rate constants describing signal decay / elimination, or autoregulatory feedback from blood flow, respectively.
The venous blood volume v is the difference between inflow f_{in} and outflow f_{out} within a constant time interval (transit time) τ,

$$\dot{v} = \frac{f_{\text{in}} - f_{\text{out}}}{\tau} \tag{2.3}$$

$$f_{\text{out}} = v^{1/\alpha}. \tag{2.4}$$

The outflow function reflects the balloon-like behaviour of the venous vessels, exerting a higher outflow rate when distended. This behaviour is modelled using a single biophysical parameter, namely the stiffness exponent α which describes the dynamic flow-volume relationship, based on the so-called Windkessel model (see Mandeville et al. (1999) for details).

Resulting from inflow and volume change, the deoxyhemoglobin (dHb) content of the vessels change as well, as is modelled in following equation,

$$\dot{q} = \left(f_{\text{in}} \frac{E(f_{\text{in}}, \rho)}{\rho} - f_{\text{out}}(v) \frac{q}{v} \right) \frac{1}{\tau} \tag{2.5}$$

$$E(f_{\text{in}}, \rho) = 1 - (1 - \rho)^{1/f_{\text{in}}}. \tag{2.6}$$

Here, $E(f_{\text{in}}, \rho)$ describes oxygen extraction from the inflowing arterial blood, ρ is the constant oxygen extraction fraction. We can see that the change in deoxyhemoglobin content in the venous compartment reflects the inflow of deoxyhemoglobin (after oxygen extraction) minus what is removed by outflow.

Thus, the hemodynamic response following a stimulus is described using these 4 equations and 6 biophysical parameters: the stimulus efficacy ϵ, the rate constant for signal decay κ, the rate constant for blood flow autoregulation γ, the stiffness parameter α, the resting oxygen extraction fraction ρ and the mean transit time τ.

Assuming all biophysical parameters constant, this model can now be seen as SISO (single input single output) model. For standard fMRI analyses, these biophysical parameters have been estimated, for details see Friston et al. (2000). This allows us to view the hemodynamic response to be dependent only on the stimulus, with no other unknown variables.

We can thus write

$$x(t) = H(u(t)) \tag{2.7}$$

with H being the hemodynamic response function, translating stimulus $u(t)$ into the measured signal change $x(t)$. In practice, the observed data is discrete and can be expressed as a vector $\boldsymbol{x}$, as well as the stimulus time course, so we also write

$$\boldsymbol{x} = H(\boldsymbol{u}). \tag{2.8}$$

2.2 General Linear Model (GLM)

The General Linear Model (GLM) is one of the most versatile linear models in statistics. Many statistical approaches, such as analysis of variance (ANOVA), t-tests, ordinary linear regression etc. can be seen as special cases of this model (Sachs, 2013).

Definition 2. For a series of measurements $\boldsymbol{y}$ and a matrix of explanatory variables $\boldsymbol{X}$ – the so-called design matrix, the General Linear Model is expressed as

$$\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\epsilon} \tag{2.9}$$

where $\boldsymbol{\beta}$ contains the parameters to be estimated, and $\boldsymbol{\epsilon}$ is the residual vector containing errors or noise.

The GLM can be used when a linear relationship between the observed data and the influencing variables is assumed, as well as independent and identically distributed residuals $\boldsymbol{\epsilon}$. For functional MRI data, the design matrix typically consists of the time courses of the stimuli.

2.2.1 Combination of GLM and the Balloon model

In the case of functional MRI data, the stimulus does not happen at the same time as the measured BOLD signal – thus we must consider the delay of the hemodynamic response in our model. This is done by incorporating the hemodynamic response function (HRF, see section 2.1) into the model,

$$\boldsymbol{X} = H(\boldsymbol{U}) \tag{2.10}$$

$$\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}. \tag{2.11}$$

Note that the design matrix usually incorporates (the time courses of) several stimuli, thus we use the matrix notation ($\boldsymbol{X}$, $\boldsymbol{U}$) instead of vector notation ($\boldsymbol{x}$,$\boldsymbol{u}$), as in Equation 2.8.

For simplicity, however, the hemodynamic response function is usually not explicitly mentioned, as the hemodynamic parameters are assumed constant and the transformation from stimulus function to corresponding hemodynamic response is quickly calculated in existing tools, by convoluting the stimulus function with the HRF.

Taken together, the GLM approach calculates a linear statistical model on the input for each and every time course recorded – thus for each voxel, which can be viewed separately. This gives us a robust and easy-to-use method for fMRI analyses. However, when we are interested in more complex questions involving interconnections and co-dependencies of brain regions, we must resort to other methods, e.g. by modelling the observed behaviour using graphs and/or differential equations. One such modelling strategy is applied in Dynamic Causal Modelling.

2.3 Models in Dynamic Causal Modelling (DCM)

Dynamic causal modelling (DCM) is a framework used for investigating hypotheses on effective brain connectivity, meaning activity changes in well-defined brain regions by stimuli and activity in other brain regions. This is done using a descriptive differential equation.

Definition 3. A Dynamic Causal Model is a model of structural hypotheses on effective connectivity between well-defined brain regions, described in

terms of a bilinear differential equation

$$\dot{\boldsymbol{z}} = F(\boldsymbol{z}, u, \theta_c) = (\boldsymbol{A} + \sum_j u_j \boldsymbol{B}^j)\boldsymbol{z} + \boldsymbol{C}u \tag{2.12}$$

where $\boldsymbol{z}$ is a vector describing the activity in a number of neuronal populations, u is the (time-dependent) input, θ_c describes the (time-independent) coupling parameters, consisting of following adjacency matrices:

$\boldsymbol{A}$ describes the condition-independent influence of activity in each neuronal population on others (task/input-independent effective connectivity), $\boldsymbol{B}^j$ describes the condition-dependent intermodulations for each condition j, and $\boldsymbol{C}$ describes the direct influence of the conditions on activity in each brain region.

This differential equation is, however, not solved in the classic sense, but entered into a model-inversion framework for estimating the connectivity parameters – i.e. the weights of the connections.

There exist further developments of DCM where the differential equation may include non-linear or stochastic terms. However, in this work we will refer to the bilinear form only.

Note that while typically, mathematical "modelling" refers to setting up descriptive (differential) equations that describe the mechanism of a system, in DCM the differential equations are fixed and the term "modelling" refers to specifying the connections – i.e. defining the non-zero entries in the parameter matrices and estimating these parameters.

In the following pages, the term **connectivity** refers to the so-called effective connectivity, which describes intermodulations of activity between brain regions.

Example 1. *An example DCM with 3 neuronal populations:*
We assume 3 neuronal populations R1, R2 and R3. Activation in population 1 leads to activation change in population 2, and vice versa (see blue connections); also, region 2 and 3 are connected in the same way. From region 1 to region 3 there are no fixed modulations in both directions (in red). There are two stimulus functions (u1, u2), which each lead to activation change in regions 1 and 3 respectively (green arrows). Input / stimulus 1 furthermore also leads to a change in the connection from region 2 to 1.

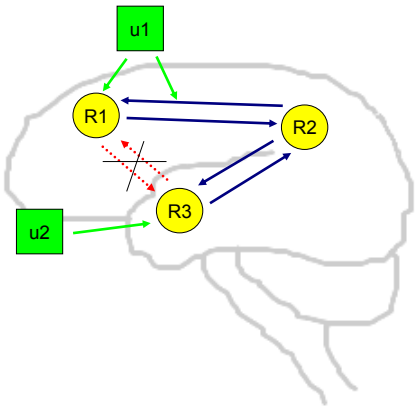

Figure 2.3: Example of DCM structural assumptions.

This DCM is translated into following $\boldsymbol{A}$*,* $\boldsymbol{B}$ *and* $\boldsymbol{C}$ *matrices:*

$$\begin{bmatrix} \dot{z}_1 \\ \dot{z}_2 \\ \dot{z}_3 \end{bmatrix} = \left(\begin{bmatrix} a_{1,1} & a_{1,2} & 0 \\ a_{2,1} & a_{2,2} & a_{2,3} \\ 0 & a_{3,2} & a_{3,3} \end{bmatrix} + u_1 \begin{bmatrix} 0 & b_{1,2} & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \right) \begin{bmatrix} z_1 \\ z_2 \\ z_3 \end{bmatrix} + \begin{bmatrix} c_{1,1} & 0 \\ 0 & 0 \\ 0 & c_{3,2} \end{bmatrix} \begin{bmatrix} u_1 \\ u_2 \end{bmatrix}$$

The indices of the matrix entries describe the direction of the modulations – the column index specifies the source, and the row indices indicate the target of the modulation.

2.3.1 Combination of DCM and the Balloon model

As in GLM, the hemodynamic response model in section 2.1 needs to be incorporated in DCM as well, as we are still relying on the measured BOLD signal for DCM analyses. However, whereas for the GLM, the biophysical parameters θ_h of the hemodynamic response function are assumed to be constant, they are being estimated with the neuronal coupling parameters $\theta_c = (\boldsymbol{A}, \boldsymbol{B}^j, \boldsymbol{C})$ – these are the matrices describing the influence of stimuli and neuronal activity on activity in other brain regions – in the model inversion framework applied for DCM. This has proven to produce better outcomes, as vascularization and local magnetization properties tend to vary across brain regions and subjects, see Stephan et al. (2007).

The hemodynamic states s, f, v and q (s being the vasodilatory signal, f describing the inflow of blood in response to the vasodilatory signal, v being the volume and q describing the deoxyhemoglobin content) can be modelled as follows (as in section 2.1). For the i-th region, the hemodynamic states

are summarized by the following equations, taking into account the neuronal state z_i:

$$\dot{s_i} = z_i - \kappa_i s_i - \gamma_i(f_i - 1) \tag{2.13}$$

$$\dot{f_i} = s_i \tag{2.14}$$

$$\tau_i \dot{v}_i = f_i - v_i^{1/\alpha} \tag{2.15}$$

$$\tau_i \dot{q_i} = f_i \frac{E(f_i, \rho_i)}{\rho_i} - v^{1/\alpha} \frac{q_i}{v_i} \tag{2.16}$$

The oxygen extraction $E(f_i, \rho_i) = 1 - (1 - \rho_i)^{1/f_i}$ is modelled as a function of flow. The hemodynamic parameters $\theta_{h_i} = \{\kappa_i, \gamma_i, \alpha_i, \tau_i, \rho_i\}$ are composed of signal decay, flow autoregulation, stiffness parameter, mean transit time, resting oxygen extraction fraction for the i-th region (see section 2.1 for further details).

The BOLD signal is subsequently modelled as a volume-weighted sum of extra- and intravascular signals,

$$\begin{aligned} y_i &= \lambda_i(q_i, v_i) \\ &= V_0(k_1(1 - q_i) + k_2(1 - q_i/v_i) + k_3(1 - v_i)) \end{aligned} \tag{2.17}$$

with the constants

$$\begin{aligned} V_0 &= 0.02 \\ k_1 &= 7\rho_i \\ k_2 &= 2 \\ k_3 &= 2\rho_i - 0.2 \ . \end{aligned}$$

where ρ_i is the resting oxygen extraction fraction. The resting blood volume fraction V_0 and the dimensionless parameters k_1, k_2 and k_3 were identified in Buxton et al. (1998) for a field strength of 1.5 Tesla. For higher field strengths, as was used in this work, it may be necessary to identify different constants.

Combining the neuronal state equation (3.13) (DCM) and hemodynamic equations (2.13)-(2.17) (Balloon model), we now have a full forward model describing the combined neuronal and hemodynamic states $x = \{z, s, f, v, q\}$,

$$\begin{aligned} \dot{x} &= f(x, u, \theta) \\ y &= \lambda(v, q) = \lambda(x) \end{aligned} \tag{2.18}$$

where $\theta = (\theta_c, \theta_h)$ are the combined neuronal (coupling) and hemodynamic (biophysical) parameters to be estimated – details in the subsequent pages

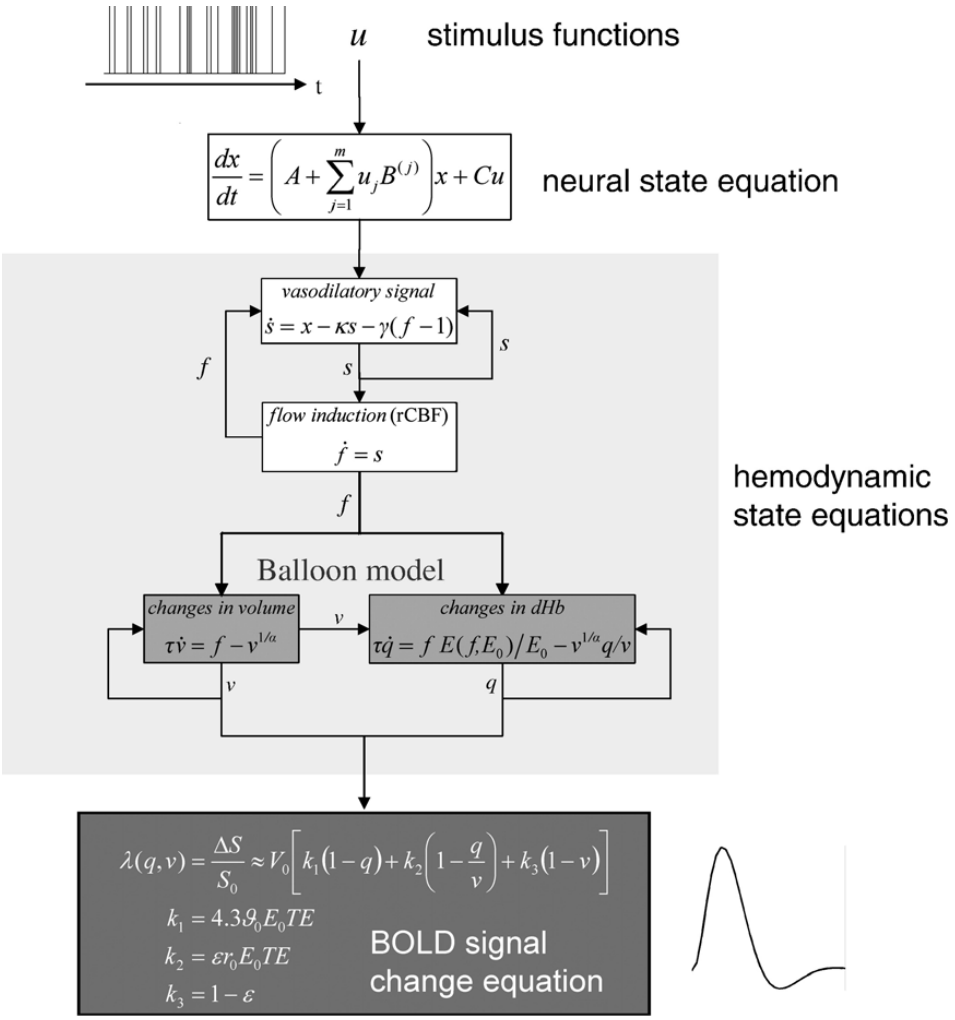

Figure 2.4: Schematic summary of the hemodynamic model, the so-called Balloon model, combined with DCM. Reproduced from Stephan et al. (2007).

(section 3.3). For simplicity, we rewrite our forward model into following non-linear function for the measured signal, dependent on the input u and the (combined) parameters θ,

$$y = h(u, \theta) \tag{2.19}$$

with Gaussian assumptions on the parameters θ. This (theoretical) forward model can be extended into a so-called "observation model" by assuming certain linear confounds to the signal, as well as noise - thus by introducing a confounding term X and an error term e,

$$y = h(u, \theta) + X\beta + e \tag{2.20}$$

where β are the unknown linear coefficients of the confounds, and we assume normally distributed errors $e \sim N(0, C_e)$.

This model can now be entered into the model inversion framework for parameter estimation - details in the methods chapter, section 3.3.

Chapter 3

Methods for fMRI data processing

The previous sections gave a short introduction into functional MRI and the BOLD effect, and an overview of the models used for fMRI analyses. To evaluate the data acquired in terms of changes in brain activity/connectivity, a number of processing steps are required. The following chapter describes procedures for fMRI data analysis.

Before the statistical analysis of the data can be conducted, several preprocessing steps should be performed, as fMRI data varies across subjects and is often noisy and possibly affected by artefacts. Some steps are listed in the following section.

The "classic" approach to fMRI data analysis is based on setting up a linear model for each volume element (voxel) acquired. This method allows for functional localization of certain brain functions and can be performed uninformed and without prior hypotheses.

For more sophisticated problems, such as connectivity analysis within the brain, other methods are applied. One of such methods is Dynamic Causal Modelling, which allows for testing hypotheses on intermodulations and dependencies between certain brain regions. Facing certain limitations of this method, very specific hypotheses must be set up beforehand.

An aspect relevant to all findings in science is the aspect of reliability of a result – i.e. how certain we can be to find a similar result within a similar setting. The last section of this chapter deals with this issue and discusses quantification methods for reliability and stability of results.

3.1 Pre-processing the data

Functional MRI data is acquired at high speed to be able to record changes in brain function through changes in the BOLD signal. However, this makes the data very sensitive to non-neural noise signals. Signal changes can occur not only through changes in blood oxygenation, which is the desired signal source, but also through movement of the subject, inhomogeneities of the magnetic field due to enclosed air, differences of the magnetic field strength over space, but also random measurement noise.

Therefore, several pre-processing steps are usually performed prior to the actual statistical analysis of the data to enhance data quality and hence yield improved results. These steps may include

- spatial re-alignment to correct for movement of the subject in 6 dimensions (3 each for translation and rotation),
- bias field correction to account for static differences in signal magnitude over the volume of interest, i.e. the brain,
- de-spiking to account for sudden, high motion artefacts,
- spatial normalization into a common coordinate system for all subjects to allow for group comparisons,
- spatial smoothing,
- slice-timing correction, to account for timing differences between the (2D-)slices of each 3D image.

The above steps have proven to result in considerably improved data quality, if not even absolutely necessary to enable comparison.

3.2 Statistical Parametric Mapping (SPM)

In fMRI, we aim to make effects of certain stimuli or other conditions over time visible in the recorded data. The easiest way to do so is by modelling a linear relationship between data and condition.

A state-of-the-art method for analysing task-based functional MRI data is the Statistical Parametric Map based on the General Linear Model (GLM; see section 2.2 and Friston et al. (1994)), which can be described using following equation:

$$\boldsymbol{y}_v = \boldsymbol{X}\boldsymbol{\beta}_v + \boldsymbol{\epsilon}_v \tag{3.1}$$

Here, $\boldsymbol{y}_v$ is the measured time course vector of one voxel v, the rows describing the values at each individual time point. $\boldsymbol{X}$ is the design matrix, containing the time courses of the conditions/stimuli, modeled as hemodynamic responses (see section 2.2.1 for details), and $\boldsymbol{\epsilon}_v$ is the error term, which is to be minimized. $\boldsymbol{\beta}_v$ is the parameter to be estimated, describing the linear signal change at the time of the stimulus. The error term is assumed to be independent and identically distributed with $\boldsymbol{\epsilon}_v \sim N(0, \sigma_v^2)$.

In practice, the design matrix consists not only of the time courses of the stimuli. Nuisance regressors from previous pre-processing steps, such as movement parameters from data realignment in 3D (usually three for translation in x, y and z-direction, and three for rotations in each direction) might also account for signal changes over time, so these effects are added into the GLM. As the MRI signal will always be greater than 0 independent of neuronal activity, the data also has to be mean-corrected, which is usually done by adding a constant vector into the design matrix.

As there are usually far more time points than unknown variables, the system is overdetermined, thus $\boldsymbol{X}$ is not invertible. In this case, the pseudoinverse is used to find the least squares estimate minimizing the sum of squares of the error term.

Theorem 1. *Let $\boldsymbol{b}_v$ be the least squares estimate of $\boldsymbol{\beta}_v$. $\boldsymbol{b}_v$ satisfies the normal equations (Scheffe, 1959),*

$$\boldsymbol{X}^\top \boldsymbol{X} \boldsymbol{b}_v = \boldsymbol{X}^\top \boldsymbol{y}_v \ . \tag{3.2}$$

Given that the columns of the Design Matrix $\boldsymbol{X}$ are linearly independent, $\boldsymbol{X}^\top \boldsymbol{X}$ is invertible and $\boldsymbol{b}$ can be uniquely calculated by

$$\boldsymbol{b}_v = (\boldsymbol{X}^\top \boldsymbol{X})^{-1} \boldsymbol{X}^\top \boldsymbol{y}_v \tag{3.3}$$

minimizing the error term ϵ_v in equation 3.1 in terms of the euclidean norm $||.||_2$ and yielding

$$\mathbb{E}(\boldsymbol{b}_v) = \boldsymbol{\beta}_v \text{ and } Cov(\boldsymbol{b}_v) = \sigma_v^2 (\boldsymbol{X}^\top \boldsymbol{X})^{-1} \ . \tag{3.4}$$

For normally distributed errors, the least squares estimates, being the maximum likelihood estimates, are themselves normally distributed (Scheffe, 1959). $Cov(\boldsymbol{b}_v)$ is the variance-covariance matrix of the parameter estimates at the voxel v.

A major advantage of this analysis method for fMRI data is that no prior knowledge of function or activation maps are required to do a standard GLM

analysis, as for each voxel time course, a linear coefficient $\boldsymbol{\beta}_v$ of the input function can be estimated.

For better readability, we are going to refer to $\boldsymbol{\beta}$, $\boldsymbol{y}$ and $\boldsymbol{\epsilon}$ instead of $\boldsymbol{\beta}_v$, $\boldsymbol{y}_v$ and $\boldsymbol{\epsilon}_v$ for the voxel-wise variables further on.

Note that $\boldsymbol{\beta}$ is a unitless parameter and can only be seen in relation to the signals recorded. For reporting and comparison purposes, it is often translated to percent signal change (PSC) by putting it in relation with the mean signal over time: $\boldsymbol{\beta}_{\text{PSC}} = \frac{\boldsymbol{\beta}}{\text{mean}_t}$.

There exist software packages with implementation of the described methods for functional neuroimaging data. The `SPM` package for `MATLAB` was developed by the Wellcome Trust Centre for Neuroimaging, at the University College London – as described in Penny et al. (2011). The software, as well as documentation and course files, is freely available at `http://www.fil.ion.ucl.ac.uk/spm/software/`.

3.2.1 Statistical inference

After calculating the parameter estimates for $\boldsymbol{\beta}$ of GLM equation 3.1, we do not know about the statistical significance of the effects of interest - thus we need to apply a statistical test on the parameter estimates.

Definition 4. In hypothesis testing, a result is statistically significant if it is very unlikely to have occurred by chance given the null hypothesis (H_0) to be true. A **significance threshold** α is defined setting the probability of rejecting the null hypothesis if it were true (type I error). The **p-value** is the probability of obtaining the observed data given the null hypothesis. The null hypothesis is rejected if $p < \alpha$ (Sachs, 2013).

For functional MRI, the null hypothesis usually states that the effects modelled using the design matrix $\boldsymbol{X}$ are not significant. This hypothesis can be tested using the t-statistic with linear compounds or contrasts of $\boldsymbol{\beta}$. We can define a row vector $\boldsymbol{c}$ which usually has a 1 as j-th component to test for the significance of the j-th effect of interest (j-th entry in $\boldsymbol{\beta}$), or contains positive and negative values for the $\text{j}_1, \ldots, \text{j}_m$-th component to compare for contrast effects between m different effects of interest $\text{j}_1, \ldots, \text{j}_m$. In that case, the vector must sum up to zero. For simplicity, we refer to the vector $\boldsymbol{c}$ as the so-called contrast vector.

For instance, to test for the 2nd condition of interest, the contrast vector would be $[0, 1, 0, \ldots]$. If our first condition is the effect to be tested, and

the second is a control condition, we would model this contrast using the vector $[1, -1, 0, 0, \dots]$.

The residual term $\boldsymbol{r}$ is the difference between the measured values and the estimation,

$$\boldsymbol{r} = \boldsymbol{y} - \boldsymbol{X} \cdot \boldsymbol{b}. \tag{3.5}$$

The standard deviation σ of the error term is estimated by

$$\hat{\sigma} = \frac{\boldsymbol{r}^\top \cdot \boldsymbol{r}}{\nu} \tag{3.6}$$

where ν is the degrees of freedom associated with $\boldsymbol{r}$. For independently distributed residuals (elements of $\boldsymbol{r}$), this would be the number of scans minus the number of effects estimated (column rank of $\boldsymbol{X}$).

However, in practice, the residuals are not distributed independently, as the stimulus functions are convoluted with the hemodynamic response function - leading to effective temporal smoothing, inserting correlation into the residual vector. This issue motivates the extension of the General Linear Model and the use of the Effective Degrees of Freedom, taking into account a smoothness parameter for the convolution. For further details see Friston et al. (1995).

The significance of a linear compound of effects is now tested with

$$t = \frac{\boldsymbol{c} \cdot \boldsymbol{b}}{e} \qquad \text{where} \tag{3.7}$$

$$e^2 = \boldsymbol{c} \cdot \hat{\sigma}^2 (\boldsymbol{X}^\top \boldsymbol{X})^{-1} \cdot \boldsymbol{c}^\top. \tag{3.8}$$

The denominator e in equation (3.7) describes the standard error of the parameter estimate. Being the standard deviation of the sampling distribution of a statistic, the expression for the standard error estimate follows from equation (3.4), $\mathrm{Cov}(\boldsymbol{c} \cdot \boldsymbol{b}) = \boldsymbol{c} \cdot \sigma_v^2 (\boldsymbol{X}^\top \boldsymbol{X})^{-1} \cdot \boldsymbol{c}^\top$.

Using the expressions above, we now obtain a t-score for each and every voxel. These t-scores now form a so-called statistical parametric map (SPM) and represent significance maps for parameter estimates in each voxel in 3D. For visualization purposes, usually a certain t-threshold is set when plotting an SPM, thus displaying only areas showing a significant effect size, i.e. when t-values exceed the t-threshold (see figure 3.1 for an example of a thresholded SPM).

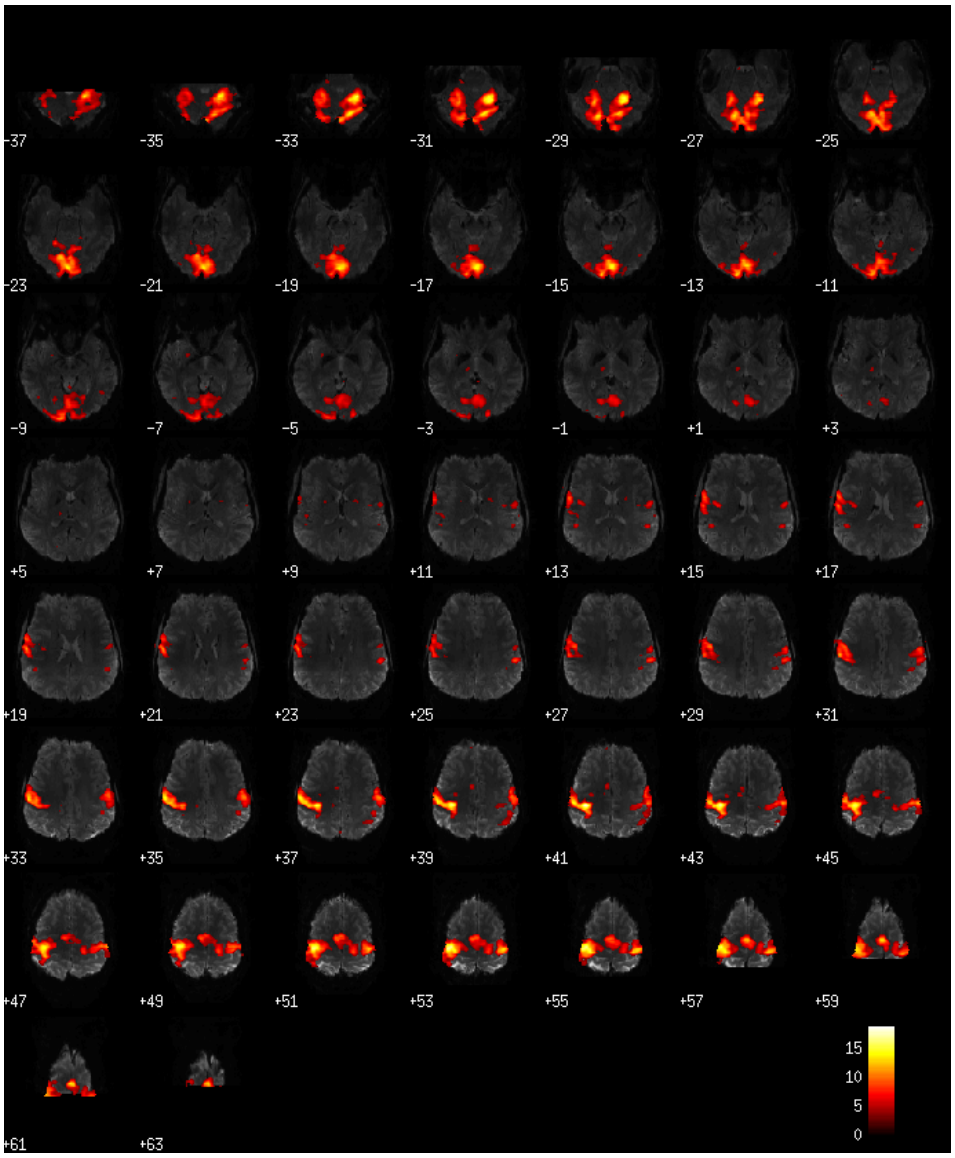

Figure 3.1: Statistical parametric map of a finger movement paradigm with visual stimulation, t-threshold at 3.182 (equivalent of $p < 0.001$), axial slices. Among others, areas in visual cortex as well as motor cortex show significant activity during the task. Data recorded and analyzed within the reliability study described in section 4. Plotted with the `SPM` software package.

Using these t-scores and the degrees of freedom ν, a corresponding p-value can be calculated, for example using the `MATLAB` function `tcdf`. In practice, usually for a given p-value, a t-threshold is calculated for retrieving the thresholded activation map.

Theorem 2. *Let X_i, $i = 1, \dots n$ be independent samples of a population $X \sim N(\mu, \sigma)$. The arithmetic mean $\overline{X}$ is thus normally distributed as well, $\overline{X} \sim N(\mu, \frac{\sigma}{\mu})$.*

For unknown standard deviation, as is usually the case, it can be approximated using the empirical standard deviation,

$$S = \sqrt{\frac{\sum_{i=1}^{n}(X_i - \overline{X})^2}{n-1}}. \tag{3.9}$$

Following term can now be used as a test statistic,

$$T = \sqrt{n}\frac{\overline{X} - \mu_0}{S}. \tag{3.10}$$

T follows a t-distribution with $n-1$ degrees of freedom under the null hypothesis that the sample was drawn from the population described above. (Sachs, 2013)

Remark 1. For n sufficiently large, the prerequisite of a normal distribution can be omitted; the random variables need only be independent and identically distributed (iid). Following the central limit theorem, their arithmetic mean will tend towards a normal distribution for $n \to \infty$.

For hypothesis testing, we must now distinguish between one-sided and two-sided tests. One-sided tests have null hypotheses of $H_0 : \mu \leq \mu_0$ (right-tailed test, $H_1 : \mu > \mu_0$) or $H_0 : \mu \geq \mu_0$ (left-tailed test, $H_1 : \mu < \mu_0$) respectively. The two-tailed variant tests for the null hypothesis of $H_0 : \mu = \mu_0$.

Lemma 1. *Let α be a given significance level. When performing a right-tailed test ($H_0 : \mu \leq \mu_0$), null hypothesis is rejected if the test statistic t lies within $(t_{1-\alpha;n-1}, \infty)$. Analogously, for the left-tailed test ($H_0 : \mu \geq \mu_0$), the criterion for rejection of the null hypothesis is if t lies within $(-\infty, -t_{1-\alpha;n-1})$. In the two-tailed test, the null hypothesis will be rejected if t lies outside of $[-t_{1-\alpha/2;n-1}, t_{1-\alpha/2;n-1}]$.*

3.2.2 Group inferences

For group inferences over a sample population, a statistical test can be performed over the individual observed effects (compounds or contrasts, respec-

tively). Analogously to section 3.2.1, a one-sample t-test is often performed.

We are testing if given linear compounds or contrasts are significantly different from zero, i.e. $\mu_0 = 0$, thus a two-tailed test must be performed.

Lemma 2. *Let $\boldsymbol{b}_i$ be the parameter estimates of $\boldsymbol{\beta}$ for n individual subjects; $\boldsymbol{b} = \frac{1}{n}\sum_{i=1}^{n} \boldsymbol{b}_i$ is the arithmetic mean of the individual parameter estimates. For a given contrast vector $\boldsymbol{c}$ (see section 3.2.1), the t-statistic is now calculated as*

$$t = \sqrt{n}\frac{\boldsymbol{c} \cdot \boldsymbol{b}}{s} \tag{3.11}$$

and follows a t_{n-1}-distribution according to Theorem 2. s is the empirical standard deviation

$$s = \sqrt{\frac{\sum_{i=1}^{n}(\boldsymbol{c} \cdot \boldsymbol{b}_i)^2}{n-1}}. \tag{3.12}$$

Analogously to section 3.2.1, we can now calculate a t-score for each voxel and thus receive a SPM over the entire sample population, allowing inferences about the whole group.

3.2.3 Criticism

GLM is a simple yet robust method for analyzing fMRI data without the need of formulating a-priori hypotheses on regions of interest (see section 4.3 below). However, while brain mapping with the GLM approach has brought insights in different brain phenomena, it is limited to taking just snapshots of brain states, neglecting brain dynamics and interplay of different cortical regions.

Therefore, various methods have been developed for analyzing dynamic connectivity of brain regions, considering cross-correlations or, as in DCM (see below), testing hypotheses on intermodulations across defined brain regions.

Critically, independent and identically (normally) distributed errors are required, in practice however independence is not given, as the error of each time point is correlated to the error of the previous time point. In addition, cross-correlation with neighboring voxels is also introduced after spatial smoothing in the pre-processing steps (see section 3.1).

3.3 Dynamic Causal Modelling

For calculating a statistical parametric map given a stimulus function, no prior hypotheses about brain dynamics and connectivity are needed, which

makes the procedure stable and very easy to perform. However, inferences can only be made on localization of areas related to brain functions, not on connections and interdependencies. For such tasks, other methods need to be employed. One method for analyzing functional connectivity is Dynamic Causal Modelling, which can be used to test hypotheses on brain networks and the effects within, as well as effects from external stimuli.

Note that in contrast to classic GLM, where analysis is performed uninformed (no prior knowledge of local function of the brain is needed), it is crucial for DCM to have one or several theory/theories on the brain connectivity which should be tested. This is why although the procedures of GLM and DCM are independent of each other, usually prior to performing Dynamic Causal Modelling, a GLM is calculated to localize the regions of interest in the brain.

For more details on the model assumptions, see section 2.3. In summary, a hypothesis on connections between several well-chosen brain regions is modeled using the differential equation

$$\dot{\boldsymbol{z}} = (\boldsymbol{A} + \sum_j u_j \boldsymbol{B}^j)z + \boldsymbol{C}u \tag{3.13}$$

where $\dot{\boldsymbol{z}}$ is a vector describing activity changes in a number of brain regions, u describes the influencing factors (conditions) over time, and the adjacency matrices $\boldsymbol{A}$, $\boldsymbol{B}$ and $\boldsymbol{C}$ describe the effective connectivity - i.e. how certain conditions or neuronal activity in brain regions influences the activity in each other brain region.

The neuronal state equation is then combined with the hemodynamic response model (see section 2.1), resulting in the model

$$y = h(u, \theta) + X\beta + e \tag{3.14}$$

where y describes each brain region , h stands for the combined neuronal state and hemodynamic model, and X and e are used to model linear confounds of the signal as well as noise. Using this forward model, we can start our parameter estimation procedure.

This is implemented within the latest `SPM` package, as described in Friston et al. (2003); available at `http://www.fil.ion.ucl.ac.uk/spm/software/`.

3.3.1 Biophysical constraints

In Bayesian parameter estimation procedures, the goal is usually to find the maximum posterior estimate (MAP) in an iterative algorithm. To form a

parameter	description	mean (η_θ)	variance (C_θ)
κ	signal decay	0.65 per s	0.015
γ	flow autoregulation	0.41 per s	0.002
α	stiffness exponent	0.32	0.0015
τ	mean transit time	0.98s	0.0568
ρ	oxygen extraction fraction	0.34	0.0024

Table 3.1: Priors on hemodynamic parameters $\theta_h = \{\kappa, \gamma, \alpha, \tau, \rho\}$, see Friston et al. (2003).

valid posterior density estimation, we need a prior density on the parameters.

The parameters θ to be estimated in DCM comprise the biophysical hemodynamic parameters θ_h, as well as the coupling parameters θ_c between neuronal populations.

There exist empirically determined estimates of the biophysical parameters for the hemodynamic model, which for other methods are often assumed constant. These estimates of the mean and covariance can well be used as a-priori estimates, see table 3.1.

For the coupling parameters, there are clear and necessary constraints due to neuronal dynamics, which need to be considered and therefore can be implemented in the prior estimates. For instance, without external stimulus, neuronal activity can not diverge to infinite values. This translates into the intrinsic coupling matrix such that the largest real eigenvalue must be negative.

For $u = 0$, equation (3.13) reduces to

$$\begin{aligned} \dot{z} &= \left[(\boldsymbol{A} + \sum_j u_j \boldsymbol{B}^j) z + \boldsymbol{C} u \right]_{u=0} \\ &= \boldsymbol{A} z. \end{aligned} \tag{3.15}$$

Being a system of homogeneous differential equations,

$$z(t) = \boldsymbol{v}_i e^{\lambda_i t} \tag{3.16}$$

is a solution to this system of ODEs for any i, where $\boldsymbol{L} = diag(\lambda_i)$ is the diagonal matrix containing the eigenvalues λ_i of $\boldsymbol{A}$ as diagonal entries; $\boldsymbol{V} = [\boldsymbol{v}_1, \boldsymbol{v}_2, .., \boldsymbol{v}_n]$ is the matrix containing the eigenvectors $\boldsymbol{v}_i$ corresponding to λ_i.

Obviously, we can see in equation 3.16 that for any eigenvalue $\lambda_i > 0$, neuronal activity z will increase exponentially without external input u and diverge for $t \to \infty$. As this is biophysically impossible, all real eigenvalues of $\boldsymbol{A}$ must be negative. For complex eigenvalues, the system will exhibit an oscillatory behaviour.

For implementation of this constraint on $\boldsymbol{A}$, we need to define the prior densities of the coupling matrices accordingly. For simplicity, all priors are assumed Gaussian. The self-connections $\boldsymbol{a}_{ii}$ have a prior expectation of 0, intrinsic coupling parameters $\boldsymbol{a}_{ij}$ have a prior expectation of 2^{-7}; both with a prior variance of 2^{-6}.

For all other connection strengths (existing connections in $\boldsymbol{B}$ and $\boldsymbol{C}$ matrices), we assume a prior distribution of $N(0, 1)$.

3.3.2 Maximizing model evidence

As in DCM, the model in mathematical terms, i.e. the differential equation is fixed, model selection refers to

1. defining non-zero entries to connectivity matrices of DCM (i.e. finding existing modulatory connections),
2. finding values for those non-zero entries (describing the connectivity strength and direction – i.e. excitatory or inhibitory effects).

Here the first point defines the structure of the model, whereas the second point is solely parameter estimation. However, both are realized based on the same optimization problem.

The goal is to maximize the model evidence: find $\arg\max_m p(y|m)$ with

$$p(y|m) = \int p(y|\theta, m)p(\theta|m)d\theta \tag{3.17}$$

where y is the measured data and m is the model.

As the model evidence can not be evaluated analytically, it is approximated. For numerical reasons, calculations are usually made based on the log-likelihood, due to numerical advantages when reaching very low values ($p(y|m) \ll 1$).

To reduce the risk of overfitting, so to find the best balance between model fit and complexity, an approximation of the model evidence is generally described using an accuracy and a complexity term.

$$\log p(y|m) \approx \text{accuracy}(m) - \text{complexity}(m). \tag{3.18}$$

For this approximation, the *Bayesian Information Criterion* (BIC) or the *Akaike Information Criterion* (AIC) are often used, as relative quality measures for models – aiming to maximize model accuracy while punishing model complexity. For model selection, AIC or BIC are minimized respectively.

These criteria are defined as follows:

$$-\text{AIC} = \log p(y|\theta, m) - n \tag{3.19}$$

$$-\text{BIC}(y, \theta, m) = \log p(y|\theta, m) - \frac{n}{2}\log(N) \tag{3.20}$$

where y is the data, m is the model with parameters θ, n is the number of parameters, and N is the number of data points. Please note that for penalization of complexity, the number of parameters is considered here, but not necessarily their information content. Thus, the so-called *negative free energy* was introduced,

$$F = \langle \log p(y|\theta, m) \rangle_q - \text{KL}[q(\theta), p(\theta|m)] \tag{3.21}$$

where q is the assumed conditional distribution, thus the left term is the expected log-likelihood under this posterior. KL denotes the Kullberg-Leibler divergence, which as complexity term also accounts for parameter interdependencies, thus taking into consideration their "information gain" instead of the sole number of parameters,

$$\text{KL}[q(\theta), p(\theta|m)] = \frac{1}{2}\left(\log|C_\theta| - \log|C_{\theta|y}| + (\mu_{\theta|y} - \mu_\theta)^T C_\theta^{-1}(\mu_{\theta|y} - \mu_\theta)\right) \tag{3.22}$$

where C_θ is the (empirical) prior covariance, $C_{\theta|y}$ is the posterior covariance, μ_θ and $\mu_{\theta|y}$ are the prior and posterior expectation. The negative free energy is a lower bound for the log-likelihood and thus is used as its approximation.

In the further sections, this (negative) free energy is used for parameter estimation, as well as for making single subject and group inferences on model structure. See Neal and Hinton (1998) for further details on the (negative) free energy function as used in statistical physics.

Remark 2. Please note that AIC, BIC as well as negative free energy are only relative measures for comparing different models – they can not be used for testing overall model quality, in the sense of null hypothesis testing. Thus, they can not indicate poor model quality on an absolute level.

3.3.3 Parameter estimation

Using the Dynamic Causal Modelling framework, both biophysical and connectivity parameters $[\theta = (\theta_h, \theta_c)]$ have to be estimated. Being a non-linear optimization problem (equations (2.19) and (2.20)), current implementations of DCM use a Gauss-Newton algorithm in an Expectation Maximization scheme (Friston et al., 2003) to calculate the maximum a-posteriori (MAP) estimates, using a fully Bayesian approach based on the conditional probability of the parameters given the data, $p(\theta|y)$.

According to Bayes' theorem, following proportionality is given:

$$p(\theta|y) \propto p(y|\theta)p(\theta) \tag{3.23}$$

Under Gaussian assumptions for the posterior density, estimating the probability density of the posterior $p(\theta|y)$ is reduced to finding its first two moments – the conditional mean $\eta_{\theta|y}$ and covariance $C_{\theta|y}$ of $\theta|y$. Analogously, their priors are estimated in terms of their expectation η_θ and C_θ.

Expanding the forward model in equation (2.19) with a working estimate of the conditional mean $\eta^{(i)}_{\theta|y}$ gives us

$$\begin{aligned} y &= h(\theta, u) + e \\ y &\approx h(\eta^{(i)}_{\theta|y}) + \frac{\partial h(\eta^{(i)}_{\theta|y})}{\partial \theta}(\theta - \eta^{(i)}_{\theta|y}) + e. \end{aligned} \tag{3.24}$$

for the i-th step of the iteration.
For better readability, we denote

$$J := \frac{\partial h(\eta^{(i)}_{\theta|y})}{\partial \theta}. \tag{3.25}$$

We now have for the residual and error term

$$r = y - h(\eta^{(i)}_{\theta|y}) \text{ and} \tag{3.26}$$

$$e \approx r - J(\theta - \eta^{(i)}_{\theta|y}). \tag{3.27}$$

Given Gaussian assumptions, we can write for the likelihood and prior probabilities:

$$p(y|\theta) \propto \exp\left[-\frac{1}{2}(r - J(\theta - \eta^{(i)}_{\theta|y})^\top \times C_e^{-1}(r - J(\theta - \eta^{(i)}_{\theta|y})\right] \tag{3.28}$$

$$p(\theta) \propto \exp\left[-\frac{1}{2}(\theta - \eta_\theta)^\top C_\theta^{-1}(\theta - \eta_\theta)\right] \tag{3.29}$$

Combining these results with equation (3.23) gives us

$$p(\theta|y) \propto \exp\left[-\frac{1}{2}(\theta - \eta_{\theta|y}^{(i+1)})^\top C_{\theta|y}^{-1}(\theta - \eta_{\theta|y}^{(i+1)})\right] \tag{3.30}$$

$$\tag{3.31}$$

with

$$C_{\theta|y} = (\bar{J}^\top \bar{C}_e^{-1} \bar{J})^{-1} \ , \tag{3.32}$$

$$\eta_{\theta|y}^{(i+1)} = \eta_{\theta|y}^{(i)} + C_{\theta|y}(\bar{J}^\top \bar{C}_e^{-1} \bar{J}) \tag{3.33}$$

$$\tag{3.34}$$

where

$$\bar{y} = \begin{bmatrix} y - h(\eta_{\theta|y}^{(i)}) \\ \eta_\theta - \eta_{\theta|y}^{(i)} \end{bmatrix}, \bar{J} = \begin{bmatrix} J \\ 1 \end{bmatrix}, \bar{C}_e = \begin{bmatrix} C_e & 0 \\ 0 & C_\theta \end{bmatrix}. \tag{3.35}$$

Iterating equation (3.30) when no priors are given can be seen as the Gauss-Newton method for parameter estimation (see Friston (2002) for further details).

Although we have so far assumed the error covariance C_e to be known, usually it is not, e.g. for temporal correlations in fMRI. However, it can be estimated using some hyperparameters λ_j, so that

$$C_e = \sum_j \lambda_j Q_j \quad \text{and} \tag{3.36}$$

$$Q_j = \frac{\partial C_e}{\partial \lambda_j} \ . \tag{3.37}$$

Q_j represents a covariance basis set that embody variant components, which can model different variances for different data blocks, or even temporal correlations within blocks. To model the error covariance using this basis set, we must now estimate its coefficients λ_j.

The posterior mean $\eta_{\theta|y}$ and covariance $C_{\theta|y}$ as well as error covariance C_e are now estimated iteratively in an expectation maximization (EM) scheme (see below), such that the model evidence – i.e. the negative free energy F – is maximized. Note that estimation and maximization of model evidence as well as parameter estimation are performed simultaneously.

The expectation maximization (EM) algorithm

The log evidence $\log p(y|\lambda) = F(\lambda)$ is maximized using Restricted Maximum likelihood (ReML) estimators, as described in Neal and Hinton (1998). Using a Fisher Scoring scheme (see below), we have following iterative procedure.

$$\lambda^{(i+1)} = \lambda^{(i)} - \left[\frac{\partial^2 F}{\partial \lambda^2}\right]^{-1} \frac{\partial F}{\partial \lambda} \text{ with} \tag{3.38}$$

$$\frac{\partial F}{\partial \lambda_i} = -\tfrac{1}{2}\mathrm{tr}(PQ_i) + \tfrac{1}{2}\bar{y}^{\mathrm{T}} P^{\mathrm{T}} Q_i P \bar{y} \tag{3.39}$$

$$\frac{\partial^2 F}{\partial \lambda_{j,k}^2} = -\tfrac{1}{2}\mathrm{tr}(PQ_j PQ_k) \tag{3.40}$$

$$P = \bar{C}_e^{-1} - \bar{C}_e^{-1} \bar{J} C_{\theta|y} \bar{J}^{\top} \bar{C}_e^{-1} \tag{3.41}$$

The covariance basis set Q_i has a sparse structure; in fact, if chosing the identity matrix, representing i.i.d. assumptions about the errors, the above iteration is equivalent to the sum of squared residual estimator.

The equations above can now be implemented recursively, resulting in an **expectation maximization (EM) algorithm**.

The EM algorithm can be split up into the **E-step**, estimating the conditional expectation of the parameters, holding the hyperparameters for the error covariance fixed (see equation 3.30 and following), and the **M-step** (maximum likelihood step), where the maximum likelihood estimates of the hyperparameters are updated, whereas the parameters are held fixed (equations 3.38 and following).

Below is a short schematic description of the Expectation Maximization (EM) algorithm, based on above equations, additionally considering the confounds X described in the observation model in Equation (2.20).

Data: priors η_θ, C_θ, $\lambda^{(0)}$; observations y; function h; confounds X
Result: MAP estimates $\eta_{\theta|y}$, $C_{\theta|y}$; neg. free energy F

`/* Initializations */`
$F = -\infty$; $\eta_{\theta|y} = \eta_\theta$; $\lambda^{(1)} = \lambda^{(0)}$, $C_{\theta|y} = C_\theta$;
repeat
  `/* E-step: update` η `and` C `*/`
  $J = \frac{\partial h\left(u,\eta_{\theta|y}\right)}{\partial\theta}$;
  $\bar{y} = \begin{bmatrix} y - h(\eta^{(i)}_{\theta|y}) \\ \eta_\theta - \eta^{(i)}_{\theta|y} \end{bmatrix}, \bar{J} = \begin{bmatrix} J & X \\ 1 & 0 \end{bmatrix}, \bar{C}_e = \begin{bmatrix} \sum \lambda^{(i)}_j Q_j & 0 \\ 0 & C_\theta \end{bmatrix}$;
  $C_{\theta|y} = \left(\bar{J}^\top \bar{C}_e^{-1} \bar{J}\right)^{-1}$;
  $\eta^{(i+1)}_{\theta|y} = \eta^{(i)}_{\theta|y} + C_{\theta|y}\left(\bar{J}^\top \bar{C}_e^{-1} \bar{J}\right)$;

  `/* M-step: update` λ `*/`
  $P = \bar{C}_e^{-1} - \bar{C}_e^{-1} \bar{J} C_{\theta|y} \bar{J}^\top \bar{C}_e^{-1}$;
  $\frac{\partial F}{\partial \lambda_i} = -\frac{1}{2}\mathrm{tr}\left(PQ_i\right) + \frac{1}{2}\bar{y}^{\mathrm{T}} P^{\mathrm{T}} Q_i P \bar{y}$;
  $\frac{\partial^2 F}{\partial \lambda^2_{j,k}} = -\frac{1}{2}\mathrm{tr}\left(PQ_j PQ_k\right)$;
  $\lambda^{(i+1)} = \lambda^{(i)} - \left[\frac{\partial^2 F}{\partial \lambda^2}\right]^{-1} \frac{\partial F}{\partial \lambda}$;
  `/* check convergence */`
  estimate F from (3.21) ;
  if *F increased* **then**
    accept current estimates
  else
    reset current estimates to priors;
    decrease log ascent rate;
  end
until *convergence, i.e.* $\left|\eta^{(i+1)}_{\theta|y} - \eta^{(i)}_{\theta|y}\right| < 10^{-6}$;

Algorithm 1: Expectation Maximization Algorithm, as currently implemented in `spm_nlsi_GN.m` within the `SPM` package, available at `http://www.fil.ion.ucl.ac.uk/spm/software`.

The EM algorithm above basically describes a Gauss-Newton scheme, giving us the MAP estimates of the parameters θ maximizing the model evidence F. For more details on maximum likelihood approaches, see Harville (1977)

and Neal and Hinton (1998).

Fisher Scoring Algorithm

The Fisher Scoring algorithm is an algorithm based on the so-called score of a Likelihood function.

The score or score function $V(\theta, X)$ with respect to a likelihood function $L(\theta; X)$ indicates the sensitivity of the likelihood function to its parameters θ. It can be viewed as the gradient of the log-likelihood with respect to θ.

Definition 5. Let X be an observation, and θ are the parameters of the likelihood function $L(\theta; X)$. Then the score V is defined as

$$\begin{aligned} V &= V(\theta, X) \\ &= \frac{\partial}{\partial \theta} \log L(\theta; X) \\ &= \frac{1}{L(\theta; X)} \frac{\partial L(\theta; X)}{\partial \theta} \end{aligned} \tag{3.42}$$

where the last equality results from the chain rule.

The following lemmata summarize some useful properties of score functions.

Lemma 3. *Let V be the score function of a likelihood function $L(\theta; X)$. Let θ^* be the true parameter. The likelihood function L can be rewritten as a probability density function: $L(\theta; x) = f(x; \theta)$. Then for the expectation of the score function, we have*

$$\begin{aligned} \mathbb{E}(V|\theta) &= \int_{-\infty}^{+\infty} f(x;\theta) \frac{\partial}{\partial \theta} \log L(\theta; X) dx \\ &= \int_{-\infty}^{+\infty} \frac{\partial}{\partial \theta} \log L(\theta; X) f(x;\theta) dx \\ &= \int_{-\infty}^{+\infty} \frac{1}{f(x;\theta)} \frac{\partial f(x;\theta)}{\partial \theta} f(x;\theta) dx \\ &= \int_{-\infty}^{+\infty} \frac{\partial}{\partial \theta} f(x;\theta) dx \\ &= \frac{\partial}{\partial \theta} \int_{-\infty}^{+\infty} f(x;\theta) dx \\ &= \frac{\partial}{\partial \theta} 1 = 0 \end{aligned} \tag{3.43}$$

This follows from (3.42) *and the Leibnitz integral rule.*

Lemma 4. *Let V be the score function of a likelihood function $L(\theta; X)$, which can be rewritten as probability density function: $L(\theta; x) = f(x; \theta)$. The variance of the score function – the so-called **Fisher Information** – is defined as*

$$\begin{aligned} I(\theta) &= var(L(\theta; X)) = \mathbb{E}(L^2(\theta; X)) \\ &= \mathbb{E}\left[\left(\frac{\partial}{\partial\theta}\log L(\theta; X)\right)^2 \Big|\theta\right] \\ &= \int \left(\frac{\partial}{\partial\theta}\log f(x; \theta)\right)^2 f(x; \theta)dx \end{aligned} \tag{3.44}$$

as the expectation of the score is 0. If the log likelihood is twice differentiable and under some regularity conditions, the Fisher information may also be written as

$$I(\theta) = -\mathbb{E}\left[\frac{\partial^2}{\partial\theta^2}\log L(\theta; X)\Big|\theta\right]. \tag{3.45}$$

As a score, especially its variance – the Fisher Information – can tell us about the quality of a parameter estimate, therefore it is often utilized as a measure for parameter estimation. A **scoring algorithm** is a form of Newton's method for finding maximum likelihood estimates for (non-linear) equations, based on the score function.

Lemma 5. *Let $Y_1, ..., Y_n$ be random i.i.d. variables with probability distribution function $f(y; \theta)$ which is twice differentiable. The maximum likelihood estimate θ^* needs to be estimated. Let $L(\theta)$ be the score function; θ_0 is given as a starting point for our algorithm. Then the Taylor approximation of our score function is given as*

$$V(\theta) \approx V(\theta_0) - J(\theta_0)(\theta - \theta_0) \tag{3.46}$$

where $J(\theta_0)$ is the observed information matrix at θ_0:

$$J(\theta_0) = -\sum_{i=1}^{n} \nabla\nabla^T|_{\theta=\theta_0} \log f(Y_i; \theta) \tag{3.47}$$

with $J(\theta_0)_{j,k} = -\frac{\partial^2}{\partial\theta_j\partial\theta_k}\log f(\theta)|_{\theta=\theta_0}$.
Setting $\theta = \theta^*$, as $V(\theta^*) = 0$, we receive

$$\theta^* \approx J^{-1}(\theta_0)V(\theta_0). \tag{3.48}$$

Thus, we have an iterative scheme, where

$$\theta_{m+1} = \theta_m + J^{-1}(\theta_m)V(\theta_m). \tag{3.49}$$

In practice however, $J(\theta)$ is usually replaced by the Fisher Information $I(\theta)$ – thus resulting in the **Fisher Scoring Algorithm**

$$\theta_{m+1} = \theta_m + I^{-1}(\theta_m)V(\theta_m). \tag{3.50}$$

For more details on scoring algorithms, see Longford (1987).

The Fisher Scoring algorithm is e.g. used within the Expectation Maximization scheme to update the hyperpriors $\lambda^{(i+1)}$ in the M-step.

3.3.4 Inference strategies

After specification and estimation of a DCM, we obtain the mean and covariance for each connectivity parameter in θ_c, as well as an estimate of the model evidence through the negative free energy F.

Regarding statistical analyses on the parameter level, significance of each connection can be tested using a one-sided t-test on the mean and covariance – analogously to section 3.2.1 and 3.2.2.

However, as the negative free energy is only a relative measure of model quality, we can not infer on the model itself, but only compare different models (different model structures) on the same data – which gives us following options for inferences on the model structural level:

- Select the model maximizing the model evidence for each subject.
- Average over models showing high posterior probability.

The first option is also referred to as Bayesian Model Selection (**BMS**). As there are often several models showing non-neglectable posterior probability, it has become fairly common to perform Bayesian Model Averaging (**BMA**) (second option), performing a weighted average over models within an Occam's window – i.e. showing a posterior probability above a certain threshold, see Penny et al. (2004) and Penny et al. (2010).

3.3.5 Criticism

Compared to classic GLM approaches, DCM inherently takes the dynamic nature of brain activity into account by implementing interconnections between brain regions. However, "effective connectivity" in the DCM does not imply direct physical connection of certain brain regions.

While in the GLM analysis no prior assumptions have to be made on spatial extend and location of activation changes, DCM depends strongly on a-priori hypotheses regarding neuronal populations and mechanisms. Within

the current DCM framework it is only possible to estimate the most plausible parameters for pre-defined modulatory connections in DCMs. There is no possibility for validating the prior assumptions (i.e. presumed connections) per se, and thus the resulting model.

Also, when defining a set of models (in the sense of presence or absence of certain connections between the pre-defined brain regions), they can only be compared against each other in a relative manner; there is no absolute measure for model fit. In other words, one model will always win, but we never know through DCM alone if it reflects the mechanism of the system. Thus, conclusions can only be as good as the prior assumptions.

For true model validation, other methods need to be used complementarily, however such evidence is rare due to practical and ethical reasons, as usually invasive methods are used (Daunizeau et al., 2011). Transcranial magnetic stimulation (TMS) may be an alternative or helpful addition for investigating functional connections between brain regions in certain settings (Tik et al., 2017).

Exploratory, i.e. non-hypothesis-driven, analysis could theoretically be conducted by setting each measurement point – i.e. voxel time course – in relation to each other, and variating the existing connections respectively. However, this is not only numerically impossible (depending on the resolution resulting in billions of models and not feasible calculation times), but also the winning model may not be the optimal (Lohmann et al., 2012).

Lohmann et al. (2012) have sparked a lively debate after their critical review on the DCM framework, stating that

> DCM currently lacks convincing model validation methods, as well as a reliable model selection procedure, so that DCM models are based on insufficient evidence.

This has evoked several responses, see Friston et al. (2013); Breakspear (2013); Lohmann et al. (2013).

Although in Friston et al. (2013), most of the aspects criticized were addressed, it was emphasised that DCM can only be regarded as a model selection framework, not for defining models per se. Model validation is thus not possible within DCM.

Therefore, reliability research in terms of reproducibility of DCM results is highly needed. This provides the motivation for our reproducibility study (see section 4), which was conducted to investigate the variation of DCM results when repeating a task several times.

3.4 Reliability and repeatability

As repeatability is one of the most important quality criteria for scientific studies, it is important to assess the relationship of results within and between subjects, as well as possible influences of other factors.

A common concept for quantifying relationships between two or more random variables is the correlation coefficient. Several different correlation measures exist for different purposes. The most commonly used correlation measure is Pearson's correlation coefficient, which describes linear relations between two variables or samples.

Definition 6. Let X and Y be quadratically integrable stochastic variables with positive standard deviation $\sigma(X) = \mathbb{E}[(X - \mathbb{E}(X))^2]$ and covariance $cov(X, Y) = \mathbb{E}[(X - \mathbb{E}(X))(Y - \mathbb{E}(Y))]$. Pearson's correlation coefficient $\rho_{X,Y}$ is defined as

$$\rho_{X,Y} = \frac{cov_{X,Y}}{\sigma_X \sigma_Y}. \tag{3.51}$$

X and Y are uncorrelated if $cov_{X,Y} = 0$.

For random samples, the empirical correlation coefficient (also called sample correlation coefficient or sample Pearson correlation coefficient) is defined analogously, by substituting the variance and covariance by their empirical estimates.

Definition 7. For two random samples $X = \{x_1, ..., x_n\}$ and $Y = \{y_1, ..., y_n\}$, the empirical correlation coefficient $r(X, Y)$ is defined as

$$r(X, Y) = \frac{cov(X, Y)}{\sigma(X)\sigma(Y)} \tag{3.52}$$

where the empirical covariance $cov(X, Y)$, the empirical standard deviation $\sigma(X)$ and the sample mean $\bar{x}$ are defined as

$$cov(X, Y) = \sum_{i=1}^{n} (x_i - \bar{x})(y_i - \bar{y}) \tag{3.53}$$

$$\sigma(X) = \sqrt{\sum_{i=1}^{n} (x_i - \bar{x})^2} \tag{3.54}$$

$$\bar{x} = \frac{1}{n} \sum_{i=1}^{n} x_i. \tag{3.55}$$

Besides connections between two different variables (samples from two different populations), in many practical cases, repeated measures within certain

groups (intra-class samples) should be assessed. Since Pearson's correlation coefficient proves to be unsuitable, the **Intraclass Correlation Coefficient** was introduced. In contrast to Pearson's correlation coefficient, intraclass correlation considers variation over subjects, as is common in clinical follow-up studies. In Shrout and Fleiss (1979) and Fleiss (2011), guidelines for applying the intraclass correlation as a reliability index for different fields of application are provided.

Depending on the nature of the data compared, there are different types of intraclass correlation coefficients. The objects of measurements (here: the subjects) are the factor representing a source of variance. If they are the only source of variance, one-way measures are used, which are denoted as ICC(1,1) or ICC(1,k) depending if single measures or average measures are being considered. If both rows and columns are seen as sources of variance, two-way measures are used: ICC(2, $\cdot$) if all targets are measured by the same raters, which are sampled from a population of raters; or ICC(3, $\cdot$) if each target is measured by the same raters, which are the only raters of interest. This describes a mixed effects model, where the sum of the column effects are assumed to be zero (unbiased raters). Subsequently we will only refer to ICC(3, $\cdot$).

Definition 8. Let $\boldsymbol{M}$ be a data matrix with $(m_{i,j}) = \boldsymbol{M} \in \mathbb{R}^{k \times n}$, containing ratings of k raters (observers) for n targets (classes).

For analyzing single measurements, the **ICC(3,1)** is used, which is defined as

$$\mathrm{ICC}(3,1) = \frac{\mathrm{MS_B} - \mathrm{MS_E}}{\mathrm{MS_B} + (k-1)\mathrm{MS_E}} \tag{3.56}$$

where $\mathrm{MS_B}$ describes the variance between the targets (between targets mean squares, with degrees of freedom $df = n - 1$)

$$\mathrm{MS}_B = \frac{k}{n-1} \sum_{i=1}^{n} (\mu_{target,i} - \mu_{total})^2, \tag{3.57}$$

$\mathrm{MS_E}$ describing the residual variance (residual mean squares, $df = (n-1)(k-1)$), and $\mu_{target,i} = \frac{1}{k} \sum_{j=1}^{k} m_{i,j}$ is the mean per target (for target i), μ_{total} is the total mean.
The residual variance MS_E is defined as

$$\mathrm{MS}_E = \frac{1}{(k-1)(n-1)} (\mathrm{SS}_W - \mathrm{SS}_R) \tag{3.58}$$

where SS_W is the within-target sum square

$$\mathrm{SS}_W = \sum_{j=1}^{k} \sum_{i=1}^{n} (m_{i,j} - \mu_{target,i})^2 \tag{3.59}$$

and SS_R is the between rater sum square

$$\mathrm{SS}_R = \sum_{j=1}^{k} (\mu_{rater,j} - \mu_{total})^2 \tag{3.60}$$

with $\mu_{rater,j} = \frac{1}{n} \sum_{i=1}^{n} m_{i,j}$ being the mean per rater (for rater j).

For analyzing a mean of several measurements over different raters, the **ICC(3,k)** is used, described as

$$\mathrm{ICC}(3,k) = \frac{\mathrm{MS_B} - \mathrm{MS_E}}{\mathrm{MS_B}}. \tag{3.61}$$

The above formulas will result in values ≤ 1, where values close to 1 show high intraclass correlation. In Fleiss (2011), ICC values below 0.4 are described as poor, values between 0.4 and 0.6 as fair, between 0.6 and 0.75 as good, and values > 0.75 are regarded as excellent.

Please note that the ICC gets higher if the targets differ more (higher variance between targets) and if the residual variance is lower. Thus, if the ratings for different targets are very similar, this will result in low ICCs.

Analyses of ICC are typically performed to assess data stability and reproducibility of results (see next chapter for an example).

Chapter 4

Reliability study

Reliability and reproducibility of fMRI study results are the key factors for assessing the validity of fMRI methods for measuring brain activity. Therefore, an fMRI study was performed, aimed at analyzing reproducibility and reliability of brain activation maps in the face and emotion processing network in the human brain, including brain structures such as the fusiform gyrus, amygdalae as well as superior temporal areas. In the subsequent pages, the experimental setup, hypotheses and results will be presented. All the presented work was performed by the author and is being prepared for publication (Geissberger et al. (2017), in preparation).

4.1 Motivation

As the amygdala has proven to play an important role for affective processing, more and more interest is placed on its function and dysfunction in health and disorder. In order to investigate amygdalar function with functional MRI, following prerequisites have to be satisfied: reproducible data quality and a reliable stimulus. During MRI data acquisition, the amygdala lies in a region where the magnetic field is very inhomogeneous and thus is suspected to rapid signal decay. However, it was shown that reducing/decreasing the voxel size leads to a significantly better signal and reduces the effects of this rapid intra-voxel dephasing (Robinson et al., 2008). Using ultra-high field fMRI at 7 Tesla, higher resolution images can be acquired and high-SNR quality images of the amygdala can be obtained (Sladky et al., 2013).

The amygdala has been shown to be involved in processing of emotional stimuli, especially emotional faces. A well-established task to stimulate and activate the amygdala is the emotion discrimination task (EDT), which has been introduced by Hariri et al. (2002). In this task, the subjects are pre-

sented with a probe face on the top and have to select one of two faces shown on the bottom which matches the emotion expressed in the face on top. This task has been used in numerous studies in healthy subjects and patients to reveal the neural underpinnings of emotion processing (Wang et al., 2017) and maladaptive changes in psychiatric disorders (Stein et al., 2007), (Sladky et al., 2012), (Paulus et al., 2005).

However, studies on the reproducibility of the EDT task are rare. Only few longitudinal studies on facial emotion processing in the amygdala exist. Plichta et al. (2012) analyzed the test-retest reliability of different cognitive and emotive tests, including an emotional face paradigm showing low intersession reproducibility of amygdalar activation. However it remains unclear if this is due to habituation of amygdalar activity or if other factors might contribute to the parameter estimate fluctuations.

Investigations on the re-test reliability of Dynamic Causal Models have been performed so far using motor tasks (Frässle et al., 2015) and a face perception task (Frässle et al., 2016), including bilateral areas in the visual cortex as well as fusiform face area, however not yet aiming for the emotion processing network.

In this study, we have investigated the intra- and intersession variability of the emotion discrimination task responses by repeatedly measuring the task responses and comparing them over runs and measurement dates. Furthermore, we have determined if modifications of the task instructions lead to different responses in the emotion processing network, and analyzed the impact of measuring at lower signal-to-noise ratio (SNR) (resulting from increasing the temporal resolution) on the effect size.

Furthermore, the data was used to perform tests on the intra- and intersession variability of Dynamic Causal Models.

4.2 Experimental setup

Fourteen right-handed, healthy volunteers with no history of neurological or psychiatric disorders were recruited (7f/7m, mean age: 25.3 $\pm$ 3.0 years). Participants were asked to refrain from caffeine, alcohol and nicotine 6 hours prior to the examination. All subjects gave written informed consent to participate in the study and were financially reimbursed. The study protocol was approved by the institutional review board of the Medical University of Vienna.

Subjects were examined using an extended version of the emotion discrimi-

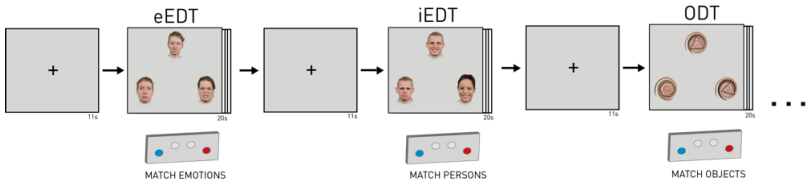

Figure 4.1: Flow chart of emotion, person and object matching trials. During the task, participants were presented with three images with the instruction to match the top image to one of the bottom images. They were asked to match emotions during the explicit emotion discrimination task (eEDT), persons during the implicit emotion discrimination task (iEDT), and objects during the control task (ODT). Each task block had a duration of 20 seconds and was followed by 11 seconds of rest, where a fixation cross was shown to the subjects for 8 seconds, and they were faced with the instruction for the next task block for 3 seconds ("match EMOTIONS", "match PERSONS", "match OBJECTS"). Each of the three task blocks was repeated three times per run, each run was repeated three times per measurement session, and the session repeated after two weeks. The order of the task blocks was randomized for each subject and session, and fixed for all runs within each session.

nation task (EDT, Hariri et al. (2002)), a paradigm known to activate brain regions associated with face recognition and processing (Sabatinelli et al., 2005). Subjects were presented with three pictures simultaneously (one on top and two on the bottom) and were instructed to match the corresponding picture on the bottom to the one on top.

In the standard EDT task (further referred to as explicit emotion discrimination task – eEDT), subjects are presented with three images showing faces, and are asked to match the facial emotional expressions, while an object discrimination task (ODT) is used as control condition – here the subjects are presented with shapes of polygons on a skin-toned background and asked to match the number of corners of the polygons. In this study, we have added another faces-stimulus based on the instructions to match persons instead of emotions – further referred to as implicit emotion discrimination task (iEDT).

The task was set up in a block design, where a fixation cross was shown for 8 seconds as a baseline condition, followed by a screen explaining the following task block ("EMOTION discrimination", "PERSON discrimination", "OBJECT discrimination") for 3 seconds and the corresponding task block with a duration of 20 seconds. The task setup is shown schematically in figure 4.1. During the task, stimuli were displayed using a video projector on a screen at the head end of the MRI scanner. Face triplets were randomly combined and presented using a script programmed in python, using the pygame library. Subjects were able to react via button press using

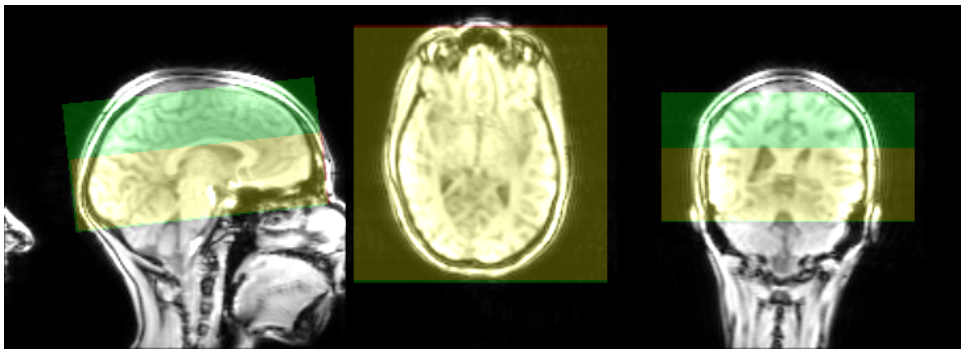

Figure 4.2: Placement of Field of View. Example of measurement area for one subject. FOV ($192 \times 192 \times 58.5\text{mm}^3$) for TR = 0.7s is shown in yellow, overlaying field of view for TR = 1.4s (green+yellow, $192 \times 192 \times 97.5\text{mm}^3$). For TR = 0.7s, the number of slices was reduced in z direction from 78 to 39 slices, and spacing between slices was increased from 1.25mm to 1.5mm.

a MRI-compatible response box. After responding (button-press) or if the subject failed to respond within 5 seconds, subjects were presented with a new triplet of images in each task block.

The images were taken from the Radboud Faces Database (Langner et al., 2010) and showed faces of caucasian males and females; the faces expressed either anger, disgust, fear, happiness, sadness, surprise or calmness (see figure 4.1).

Data acquisition was performed on a SIEMENS Magnetom 7 Tesla whole-body MR scanner, using a 32-channel head coil with the CMRR (Center for Magnetic Resonance Research, University of Minnesota, `https://www.cmrr.umn.edu`) multiband echoplanar imaging (EPI) sequence, and two different repetition times (TR). For TR = 1.4s, echo time (TE) was 23ms, 78 slices with a spatial resolution of $1.5 \times 1.5 \times 1\text{mm}^3$ (field of view (FOV) = $192 \times 192 \times 97.5\text{mm}^3$) were acquired at a flip angle of $\alpha = 62°$ and bandwidth of f = 1447 Hz/px. For TR = 0.7s, spacing between slices was increased from 1.25mm to 1.5mm, and the number of slices was reduced to 39 (FOV = $192 \times 192 \times 58.5\text{mm}^3$). An example of the recorded fields of view for one subject is depicted in figure 4.2.

In order to analyze the stability of the results, the task order was randomized for each subject; for each repetition time (TR), the task set was repeated three times per run, each run was conducted three times per measurement session, and for each subject the session was repeated once after two weeks. The order of the two acquisition protocols used was randomized as well (TR = 1.4s or TR = 0.7s first), alternating both TRs to aquire the respec-

tive first, second and third run.

Functional data pre-processing was performed using existing software packages and included de-spiking (AFNI[1]), slice-timing correction (FSL[2]), distortion correction (FSL), bias field correction (ANTs[3]), realignment (FSL), normalization to MNI space (ANTs), and smoothing with a 6mm FWHM Gaussian kernel (FSL), see section 3.1 for details on pre-processing steps. For optimal normalization results, to ensure comparability over runs, for each subject all runs were co-registered to a mean of all runs with whole-brain coverage (TR = 1.4s); the normalization parameters were calculated for all runs simultaneously using this mean image.

4.3 Reliability of GLM results in emotion processing

Analyses were conducted in `SPM12`[4]. Data was analysed using a block-design GLM (see section 3.2), where the task blocks of 20s each were used as regressors against the non-task blocks; realignment parameters, as well as white matter and cerebrospinal fluid (CSF) parameters, obtained from the previous preprocessing steps, were also included as nuisance regressors to minimize effects of movement and artifacts arising from non-grey matter. Statistical parametric maps (SPM) were calculated contrasting the explicit/implicit emotion discrimination task (eEDT/iEDT) versus the object discrimination task (ODT). Significance thresholds were set to $p < 0.05$, corrected for multiple comparisons on whole-brain level.

Parameter estimates were extracted from single-subject statistical parametric maps, coordinates taken from maxima of clusters in eEDT > ODT contrast in the first run and session, TR = 1.4s (see SPM in figure 4.3); mean parameters of sphere with radius of 2 voxels and the respective coordinates as center. These values were used for further stability analyses and to calculate means and standard errors (see figures 4.7 and 4.8).

To examine the intra- and intersession reliability of the GLM results, intraclass correlations were calculated based on the formula defined by Shrout and Fleiss (1979).

[1] Analysis of Functional NeuroImages, available at `https://afni.nimh.nih.gov/`
[2] fMRI Software Library, available at `https://fsl.fmrib.ox.ac.uk/fsl/fslwiki`
[3] Advanced Normalization Tools, available at `http://stnava.github.io/ANTs/`
[4] Statistical Parametric Mapping, `http://www.fil.ion.ucl.ac.uk/spm/software/`

As described in section 3.4, we have used

$$\mathrm{ICC}(3,1) = \frac{\mathrm{MS_B} - \mathrm{MS_E}}{\mathrm{MS_B} + (k-1)\mathrm{MS_E}} \tag{4.1}$$

for analysing single measurements, and

$$\mathrm{ICC}(3,k) = \frac{\mathrm{MS_B} - \mathrm{MS_E}}{\mathrm{MS_B}} \tag{4.2}$$

for analysing means of measurements over different raters (i.e. measurement sessions).

Here, $\mathrm{MS_B}$ is the variance between the subjects, and $\mathrm{MS_E}$ is the rest variance. We have used the ICC(3,1) to analyse intrasession stability (all 6 runs per TR, not taking into account session effects) ($n = 14, k = 6$) and ICC(3,k) on the session means for each subject to analyse intersession reliability of the parameter estimates ($n = 14, k = 2$). As the SNR differs between the sequences using different TRs, thus possibly resulting in different effect sizes, we have analysed the ICCs separately for each TR (see table 4.2).

Habituation effects were assessed by performing 2-sample t-tests on the single subject contrasts (difference of parameter estimates for eEDT and ODT) for the first and last run of each measurement session.

4.3.1 Results

Calculating the statistical parametric map (SPM) of the emotion discrimination task, contrasting the (explicit) EDT with the control task (see figure 4.3), reveals the well-established emotion processing network including bilateral amygdala, fusiform gyrus, dorsolateral prefrontal cortex (DLPFC), superior and middle temporal gyrus. The implicit EDT task (person matching) shows that activation of amygdala, fusiform gyrus, middle and superior temporal gyrus are equally present in both face paradigms independent of the task instruction (see figures 4.4 and 4.7).

DLPFC, as well as parts of cingulate gyrus and right superior and middle temporal gyrus show higher activation during the explicit emotion discrimination task compared to the implicit task (see figure 4.5), figure 4.7 shows a significant difference in activation level of the left DLPFC comparing the explicit and implicit emotion discrimination task.

For assessing stability over runs and sessions, the data was visualized by plotting results of each session against each other and inspecting the similarity of the plot to the $x = y$ graph. In figure 4.6 these plots are shown for bilateral amygdalae, fusiform gyri and DLPFC for the contrast eEDT >

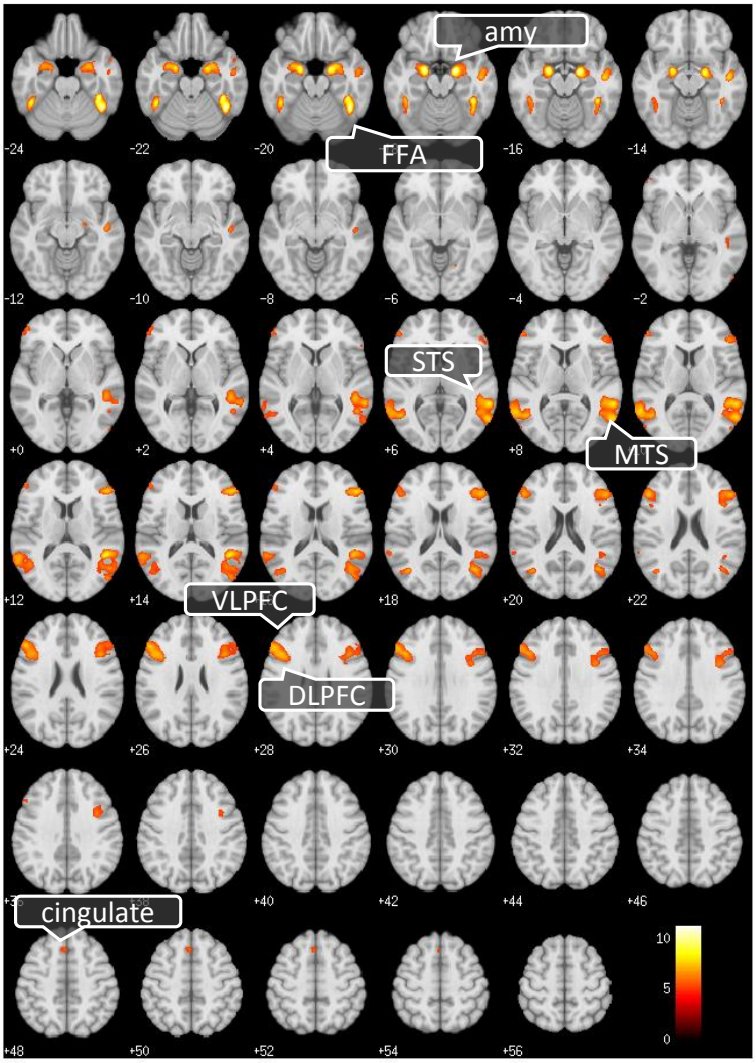

Figure 4.3: Brain activation during Emotion Discrimination (eEDT > ODT), first run of first session, TR = 1.4s. Matching emotion of faces revealed heightened activation in cortical and subcortical regions associated with face recognition and affective processing, i.e. fusiform gyrus, amygdala, DLPFC, superior temporal sulcus, middle temporal sulcus, cingulate cortex. The threshold of the t-statistics was set to $p < 0.05$, FWE whole-brain correction for multiple comparison ($T \geq 4.9220$). Axial slices plotted with the `SPM` software package.

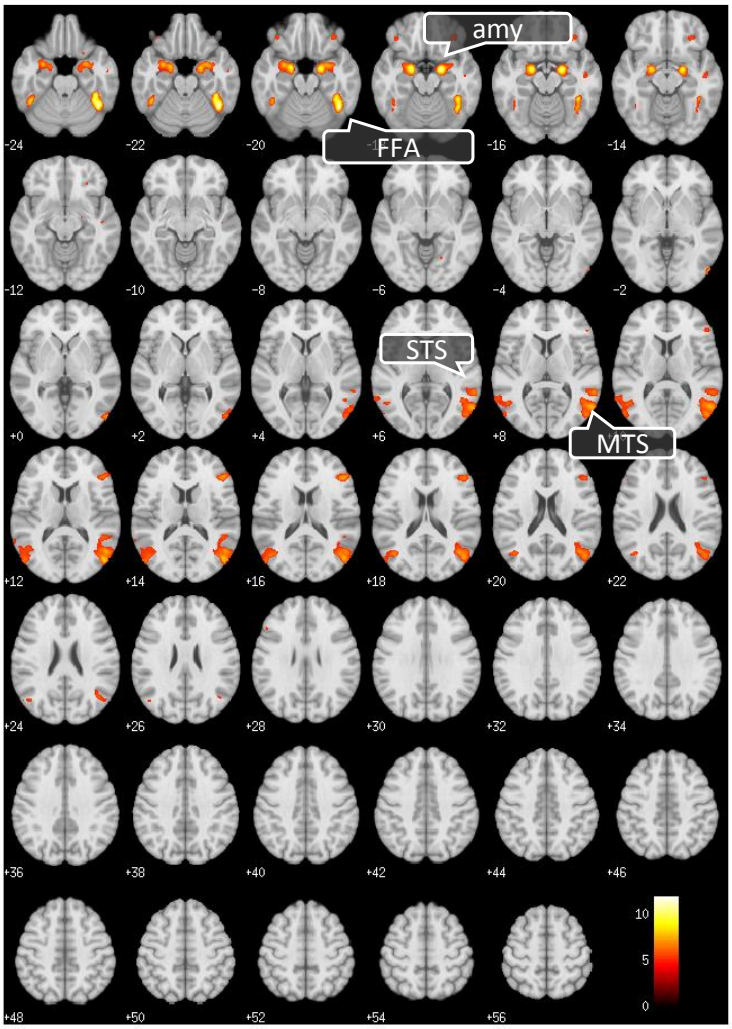

Figure 4.4: Brain activation during implicit Emotion Discrimination (iEDT > ODT), first run of first session, TR = 1.4s. Matching faces of individuals showing expressing emotions revealed comparable results to emotion matching – i.e. increased activation in fusiform gyrus, amygdala, superior temporal gyrus, middle temporal gyrus; however significantly lower activation in dorsolateral prefrontal cortex (DLPFC). The threshold of the t-statistics was set to $p < 0.05$, FWE whole-brain correction for multiple comparison ($T \geq 4.9220$). Axial slices plotted with the `SPM` software package.

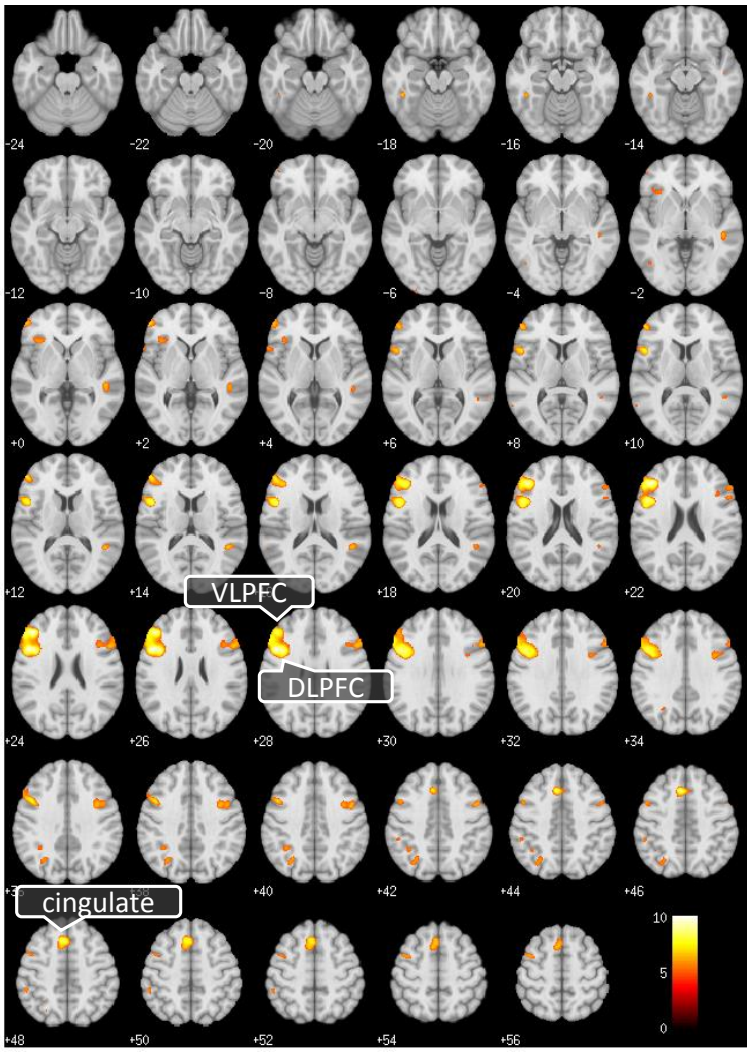

Figure 4.5: Brain activation at explicit emotion discrimination (eEDT > iEDT), first run of first session, TR = 1.4s. Matching emotion of faces instead of persons revealed higher activation in dorso- and ventrolateral prefrontal cortex, as well as cingulate cortex. The threshold of the t-statistics was set to $p < 0.05$, FWE whole-brain correction for multiple comparison ($T \geq 4.9220$). Axial slices plotted with the `SPM` software package.

Region		Session 1 Run 1	Session 1 Run 2	Session 1 Run 3	Session 2 Run 1	Session 2 Run 2	Session 2 Run 3
rAmy	coordinates	[18, -7, -18]	[18, -6, -18]	[20, -4, -19]	[18, -6, -19]	[20, -6, -18]	[20, -6, -19]
	t-value	11.09	9.17	7.77	9.95	7.92	7.07
	cluster size	1293	646	143	807	157	157
lAmy	coordinates	[-20, -6, -18]	[-20, -6, -18]	[-20, -6, -18]	[-20, -6, -18]	[-20, -4, -19]	[-20, -4, -19]
	t-value	10.6	7.79	7.61	8.67	6.48	5.61
	cluster size	755	188	128	606	69	34
rFusGyr	coordinates	[44, -52, -22]	[44, -52, -22]	[44, -52, -22]	[44, -54, -22]	[42, -54, -20]	[42, -54, -20]
	t-value	11,23	11,25	11,05	12,17	11,73	11,14
	cluster size	997	949	1010	1175	1101	1003
lFusGyr	coordinates	[-44, -46, -22]	[-44, -48, -22]	[-44, -46, -24]	[-44, -48, -22]	[-45, -48, -24]	[-45, -48, -24]
	t-value	9.07	8.8	9.51	10.21	9.89	8.82
	cluster size	542	546	706	767	690	577
rDLPFC	coordinates	[51, 32, 16]	[36, 4, 35]	[51, 28, 20]	[56, 34, 16]	[51, 29, 18]	[54, 32, 17]
	t-value	7.81	7.02	7.11	6.9	7.97	6.88
	cluster size	2078	1644	2230	1653	1927	1049
lDLPFC	coordinates	[-52, 41, 10]	[-52, 24, 26]	[-46, 14, 26]	[-46, 16, 24]	[-44, 14, 26]	[-44, 14, 26]
	t-value	6.23	6.38	7.08	7.24	7.19	7.48
	cluster size	1731	443	1060	586	1155	909
rSupTempGyr	coordinates	[48, -47, 12]	[48, -46, 14]	[50, -46, 12]	[46, -46, 14]	[50, -46, 12]	[58, -44, 6]
	t-value	8.52	8.64	8.61	8.48	7.93	6.54
	cluster size	3102*	2502*	2207*	2961*	1462*	574*
lSupTempGyr	coordinates	[-62, -55, 10]	[-58, -46, 14]	[-60, -48, 14]	[-60, -46, 12]	[-62, -55, 8]	[-60, -56, 6]
	t-value	7.86	6.47	6.27	6.47	5.9	5.48
	cluster size	1681*	739*	679*	1509*	172	31
rMidTempGyr	coordinates	[58, -64, 8]	[51, -74, 12]	[58, -64, 8]	[58, -64, 8]	[58, -46, 6]	[62, -67, 6]
	t-value	8.11	7.25	7.64	8.26	7.04	6.51
	cluster size	3102*	2502*	2207*	2961*	1462*	574*
lMidTempGyr	coordinates	[-50, -61, 8]	[-40, -68, 17]	[-38, -67, 17]	[-58, -58, 6]	[-45, -61, 10]	[-50, -59.5, 6.5]
	t-value	7.26	6.76	7.09	7.15	5.66	4.89
	cluster size	1681*	739*	679*	1509*	80	-

Table 4.1: Results for all significant regions in the first run of the first session of the eEDT > ODT contrast, TR = 1.4s, $p < 0.05$ FWE corrected ($T \geq 4.9220$), cluster size $k \geq 10$ voxels; including bilateral amygdalae, fusiform gyrus, dorsolateral prefrontal cortex (DLPFC), superior and middle temporal gyrus. The t-value is the statistic indicating significant activation over all subjects; the maximum of the t-value per region of interest is located at the coordinates in the common reference space (MNI space); the cluster size is given by the number of voxels showing significant activation (t-value above threshold).
*) Superior Temporal Gyrus and Middle Temporal Gyrus show up as one connected cluster, cluster size is the sum of both combined.

	TR = 1.4s			TR = 0.7s		
	ICC(3,1) over runs	ICC(3,k) over session means	ICC(3,1) first run per session	ICC(3,1) over runs	ICC(3,k) over session means	ICC(3,1) first run per session
rAmy	0.4337	0.9186	0.581	0.3999	0.7519	0.4955
lAmy	0.4418	0.8591	0.4843	0.393	0.714	0.3513
rFusGyr	0.7802	0.9477	0.7322	0.7183	0.9088	0.8299
lFusGyr	0.8115	0.9527	0.7891	0.6843	0.8146	0.835
rDLPFC	0.2201	0.8011	0.1355	n.a.	n.a.	n.a.
lDLPFC	0.611	0.9128	0.5937	n.a.	n.a.	n.a.
rSupTempGyr	0.6697	0.8736	0.6293	0.5455	0.8553	0.5199
lSupTempGyr	0.7383	0.8755	0.7843	0.6931	0.9374	0.7518
rMidTempGyr	0.7574	0.9524	0.8026	0.5788	0.8654	0.7489
lMidTempGyr	0.777	0.9097	0.7993	0.7246	0.8772	0.8214

Table 4.2: Correlation coefficients between parameter estimate contrasts (eEDT > ODT) of session 1 and session 2 on single subject data for TR = 1.4s **and TR** = 0.7**s.** Coordinates taken from cluster maxima of first run/first session, TR = 1.4s. As bilateral DLPFC lie outside of the TR = 0.7s field of view, there are no values to compare.

Habituation	TR = 1.4s			TR = 0.7s		
P-value	over runs s1	over runs s2	over sessions	over runs s1	over runs s2	over sessions
rAmy	0.0423	0.0090	0.18292457	0.3259	0.9793	0.2602
lAmy	0.0537	0.0205	0.01949925	0.2789	0.5250	0.3289
rFusGyr	0.8456	0.0501	0.55056918	0.8242	0.3108	0.5119
lFusGyr	0.6024	0.0996	0.35556927	0.2471	0.9852	0.2863
rDLPFC	0.4104	0.9231	0.89633682	n.a.	n.a.	n.a.
lDLPFC	0.3020	0.7757	0.22008288	n.a.	n.a.	n.a.
rSupTempGyr	0.8234	0.0371	0.06121936	0.3970	0.3639	0.7278
lSupTempGyr	0.0407	0.0228	0.22642355	0.7072	0.2771	0.5398
rMidTempGyr	0.5964	0.0801	0.25943028	0.1458	0.0138	0.2997
lMidTempGyr	0.0533	0.2012	0.0224687	0.0203	0.1137	0.2407

Table 4.3: p-values of t-tests on contrasts to investigate habituation effects for eEDT > ODT. For within-session habituation effects, we compared the contrasts of the first vs. the third run; for intersession differences, contrasts of all runs were compared between the two sessions. Tests were performed for each TR separately. Bonferroni-corrected $p < \frac{0.05}{10} = 0.005$. None of the p-values are lower than the corrected significance threshold, thus no significant habituation effects can be reported.

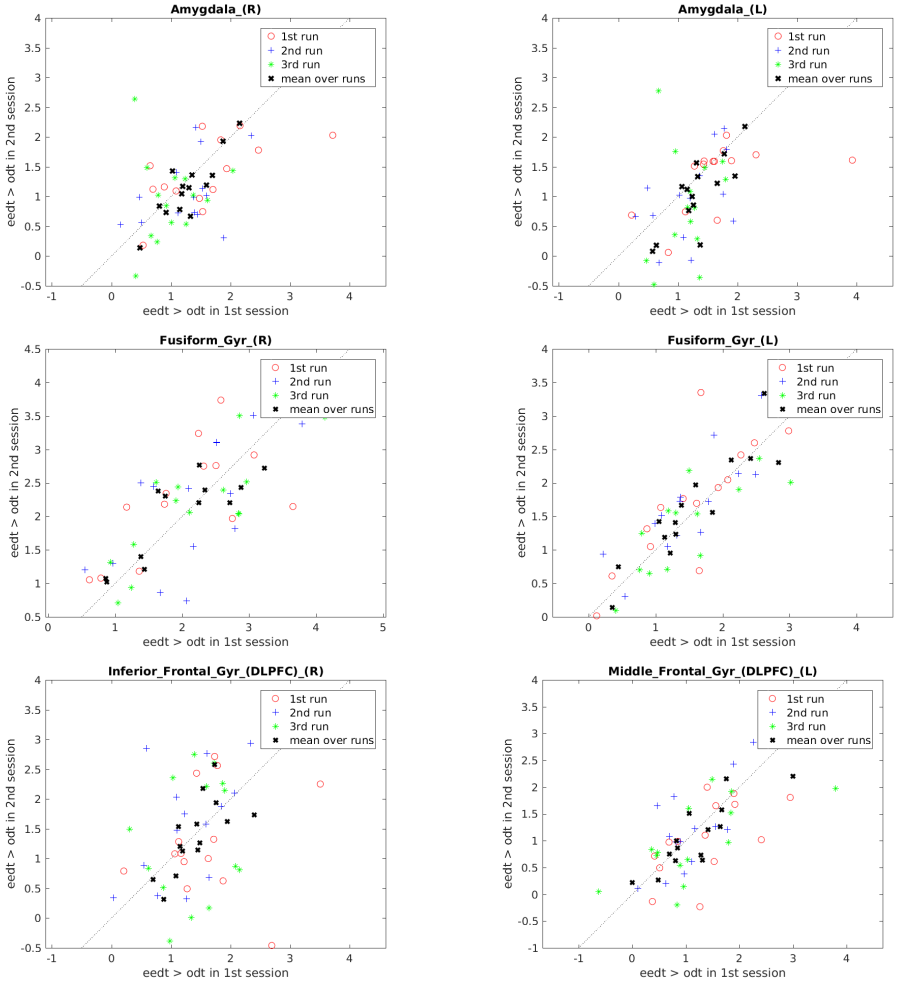

Figure 4.6: Scatter plots comparing both sessions for activation in bilateral amygdala, fusiform gyri and DLPFC, TR =1.4s. Plot of single subject contrast estimates of each run in session 1 against session 2 (contrast eEDT > ODT), as well as mean over runs; dotted line indicates perfect match $x = y$.

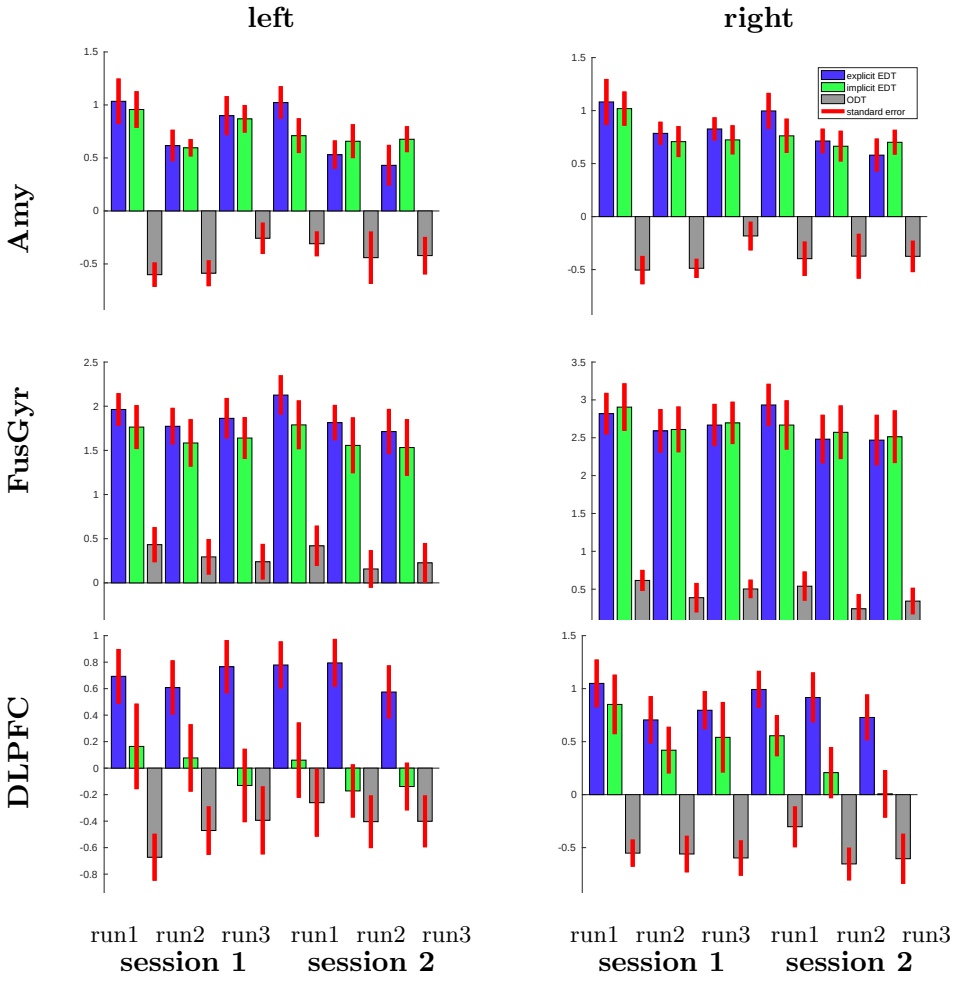

Figure 4.7: Parameter estimates and standard error bars for bilateral amygdala, fusiform gyrus and dorsolateral prefrontal cortex (DLPFC), TR =1.4s. This graph shows high stability over runs and sessions for all regions depicted. Activation does not differ for explicit and implicit task in amygdalae and fusiform gyrus, however in the DLPFC the explicit task (emotion matching) evokes much higher activation.

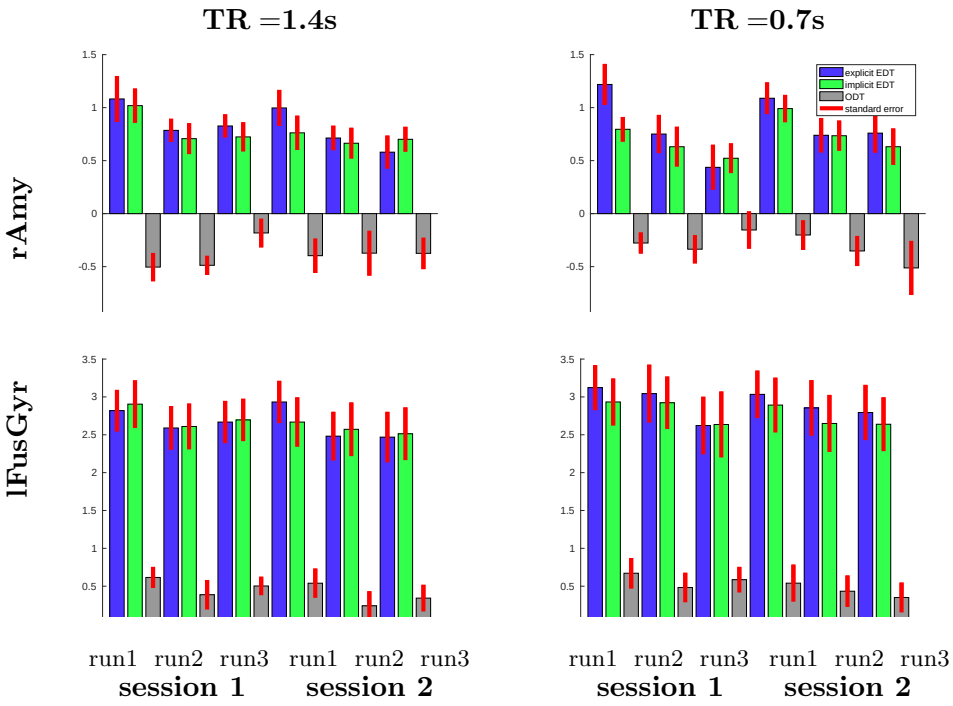

Figure 4.8: Comparison of parameter estimates for TR =1.4s and TR =0.7s. This figure shows similar results for TR = 0.7s as for TR = 1.4s for fusiform gyrus and amygdala, DLPFC exceeds FOV of TR = 0.7s.

ODT, showing visible consensus. For quantification of reliability however, other methods need to be used, e.g. ICC.

As can be seen in table 4.1, when using the (explicit) emotion discrimination task, activity in the emotion processing network remains significant for all runs and sessions, for almost all reported brain regions (only activation in left middle temporal gyrus stays below the significance threshold in the last run of the 2nd session). Activation in amygdala and fusiform gyrus, as well as DLPFC shows high stability, i.e. significant activation throughout all runs and sessions, as can be seen in figure 4.7.

In order to test for habituation effects, we have contrasted the first run against the third run of eEDT > ODT. Whole-brain comparison at standard significance level ($p < 0.05$, FWE) revealed no significant differences over time. For an explorative analysis based on pre-defined regions of interest (ROI), we additionally performed paired t-tests on contrasts between first and third run. Tests were computed for each session, TR and 10 ROIs (see table 4.3) respectively. Some uncorrected p-values fall below set significance level, i.e. when comparison is restricted to this small volumes of interest, activation differences over runs were evident (see table 4.3). However, none of these tests survive multiple comparison correction of the significance threshold (e.g. Bonferroni-adjusted $p < 0.005$). Thus, while small changes in activation levels are evident when comparing within ROIs, no activation differences were found when considering changes across the whole-brain. Therefore no conclusive habituation effect can be reported.

For bilateral amygdalae, the intraclass correlation coefficients show limited stability over runs (ICC(3,1) = 0.4337 and 0.4418 for right and left amygdala respectively), only the first run per session shows higher, but still fair intraclass correlation (ICC(3,1) = 0.5810 and 0.4843). However, the means over runs per session show excellent intersession reliability (ICC(3,k) = 0.9186 and 0.8591) (see table 2). For TR = 0.7s the results are very similar. ICCs of the fusiform gyrus show excellent reliability over runs and sessions (see table 4.2). Right DLPFC shows very poor reliability over runs, in contrast to left DLPFC, which showed good to excellent reliability over runs and sessions, as well as middle and superior temporal gyri. ICC results were classified as excellent, good, fair and limited following Fleiss (2011).

Despite the lower signal to noise ratio of measurements with lower TR, this effect apparently does not influence the GLM results. Parameter estimates for amygdala and fusiform gyrus over two different TRs do not show significant differences in effect size (see figure 4.8).

As the GLM results are fairly stable (see figure 4.7), we may hypothesize the

same for the DCM analysis. Furthermore, as the results for the data with higher temporal sampling (TR = 0.7s) are comparable in quality (see figure 4.8), we assume DCM results to benefit from the higher temporal resolution.

Remark 3. For simplicity and exemplarily, most plots and calculations inserted concern the "classic" (explicit) emotion discrimination task (eEDT) with the acquisition protocol showing full-brain coverage (TR = 1.4s), if not mentioned otherwise. Results from other task condition (implicit emotion discrimination task) and contrast constellations, as well as the second acquisition protocol (TR = 0.7s) not showing remarkable differences are not included explicitly, as they go beyond the scope of this work.

4.4 Reliability of DCM of emotion processing

For testing the reliability of an established DCM on the emotion processing network, we have aimed to reproduce the results from Sladky et al. (2015) in repeated measurements, where a simple network for processing emotional face stimuli comprising only two brain regions, namely the orbitofrontal cortex (OFC) and amygdala, was modeled, showing differences in emotional face processing for healthy subjects and patients suffering from social anxiety disorder. See figure 4.9 for a schematic view of the network.

In contrast to Sladky et al. (2015), we could not replicate significant activation in the OFC in the GLM results of present study (see above). However, significant activation in the SPM is not a prerequisite for DCM analyses. For the DCM analysis, all possible constellation for connectivity modulation ($\boldsymbol{B}$ matrix) by the three conditions eEDT, iEDT and task (sum of input functions of eEDT, iEDT and ODT) were varied, resulting in $2^6 = 64$ models:

$$\dot{\boldsymbol{z}} = \left(\boldsymbol{A} + \sum_j u_j \boldsymbol{B}^j \right) \boldsymbol{z} + \boldsymbol{C} u \tag{4.3}$$

$$\dot{\boldsymbol{z}} = \left(\begin{bmatrix} \cdot & \cdot \\ \cdot & \cdot \end{bmatrix} + \sum_{j=1}^{3} u_j \begin{bmatrix} 0 & ? \\ ? & 0 \end{bmatrix} \right) \boldsymbol{z} + \begin{bmatrix} \cdot & \cdot & \cdot \\ 0 & 0 & 0 \end{bmatrix} u \tag{4.4}$$

for the three conditions eEDT, iEDT and task (sum of eEDT, iEDT and ODT time courses). "·" stands for a non-zero entry (i.e. an existing connection) which needs to be estimated; "?" stands for a connection subject to variation (for existence or zero-entry), see figure 4.10. Bayesian model averaging was performed on all 64 models to obtain one single model per subject.

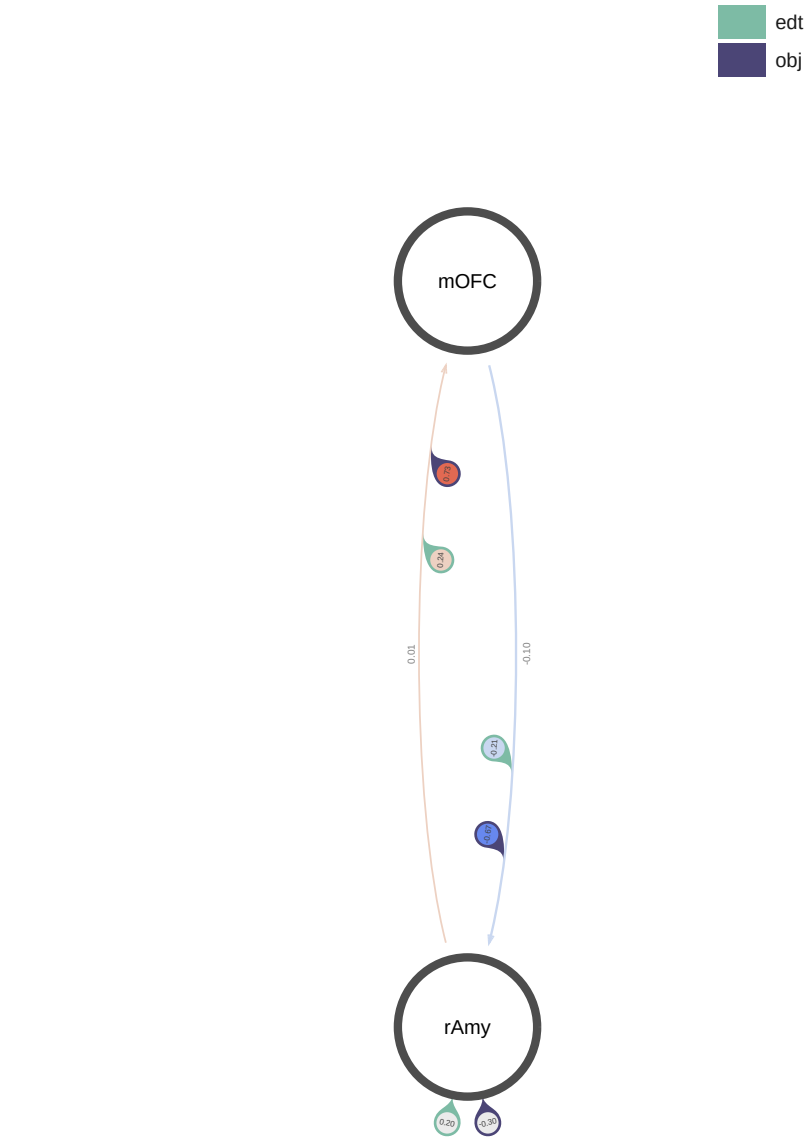

Figure 4.9: Example of specified and estimated DCM, modelling the interaction between the right amygdala (rAmy) and medial orbitofrontal cortex (OFC). The droplets indicate the influence of the conditions (explicit) emotion discrimination (edt) and object discrimination (obj). Color of connections (task-independent connectivity) as well as inside the droplets indicate direction of modulation – from blue (negative) to red (positive); grey color indicates non-significance.

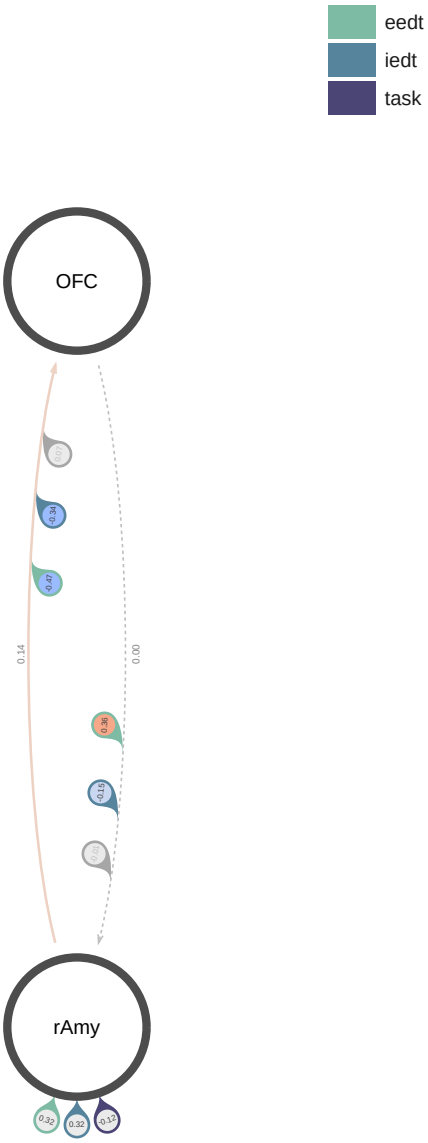

Figure 4.10: Example of specified and estimated DCM, modeling the interaction between the right amygdala (rAmy) and medial orbitofrontal cortex (OFC). The droplets indicate the influence of the conditions explicit and implicit emotion discrimination (eedt, iedt) and the image assignment task regardless of the stimulus (task). Color of connections (task-independent connectivity) as well as colors inside the droplets (task-induced connectivity changes) indicate direction of modulation – from blue (negative) to red (positive); grey color indicates non-significance.

4.4.1 Results

Our DCM analysis did not replicate the results from Sladky et al. (2015) showing down-regulation of amygdalar activity by the orbitofrontal cortex; no consistent model (consistency in modulation direction and strength) could be found over subjects and runs. Therefore this DCM analysis was not pursued further.

Due to consistent and stable activation in fusiform gyrus and amygdala, as well as visual cortex, a model was constructed including these areas – for simplicity involving only the respective brain regions from the right hemisphere. The stimulus was modeled to enter through the visual cortex (task independent of stimulus condition), and from there to be propagated to the fusiform face area, fusiform object area and amygdala; cross-modulation was assumed between amygdala and both fusiform areas. The stimulus enters only via the visual cortex, regardless of the task instruction and stimulus. This is modeled by following connectivity matrices (see figure 2.3 for a schematic view):

$$\dot{z} = \left(\begin{bmatrix} \cdot & 0 & 0 & 0 \\ \cdot & \cdot & 0 & \cdot \\ \cdot & 0 & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot \end{bmatrix} + \sum_{j=1}^{3} u_j \begin{bmatrix} 0 & 0 & 0 & 0 \\ \cdot & 0 & 0 & \cdot \\ \cdot & 0 & 0 & \cdot \\ \cdot & \cdot & \cdot & 0 \end{bmatrix} \right) z + \begin{bmatrix} 0 & 0 & 0 & \cdot \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} u \quad (4.5)$$

Here one single model was estimated per subject, run and acquisition protocol for limiting computational resources. See figure 4.12 for results (over all subjects) for each run and measurement session, TR = 1.4s.

With stable activation patterns in all involved brain regions, we expected stable modulation in strength and direction for all connections. As is evident, the effective connectivity between visual cortex and fusiform face area is reliably increased by both face-processing conditions (explicit and implicit emotion processing), whereas the control condition involving object discrimination modulates the connection to the fusiform object area in all runs.

However, despite stable activation in the amygdala over all runs and sessions for face stimuli, our DCM results in no consistent significant modulation of effective connectivity towards the amygdala (see figure 4.12).

For further assessment of DCM stability, we have calculated the intraclass correlation values for each connection strength modelled in our DCM. As the ICC can only work in 2 dimensions – the target and the rater dimension – we have assessed the intraclass correlation of type ICC(3,1) (see section 3.4 for

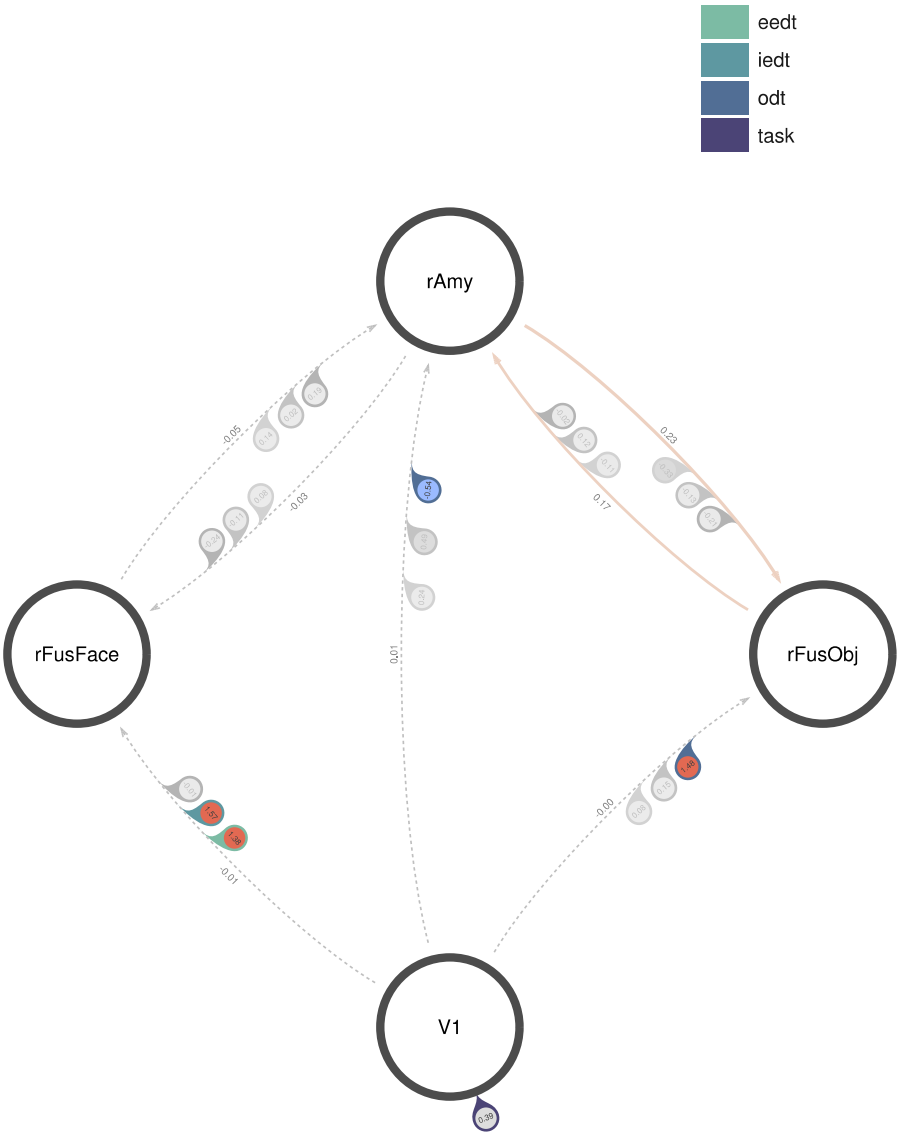

Figure 4.11: Example of specified and estimated DCM, modeling the interaction between the visual cortex (V1), right amygdala (rAmy), right fusiform face area (rFusFace), right fusiform object area (rFusObj). All conditions were modeled to enter the visual cortex with the same strength; modulation to other areas was separated for each of the three conditions (explicit, implicit emotion matching and object matching; indicated by the outer color of the droplets. Color of connections (task-independent connectivity) as well as colors inside the droplets (task-induced connectivity changes) indicate direction of modulation – from blue (negative) to red (positive); grey color indicates non-significance.

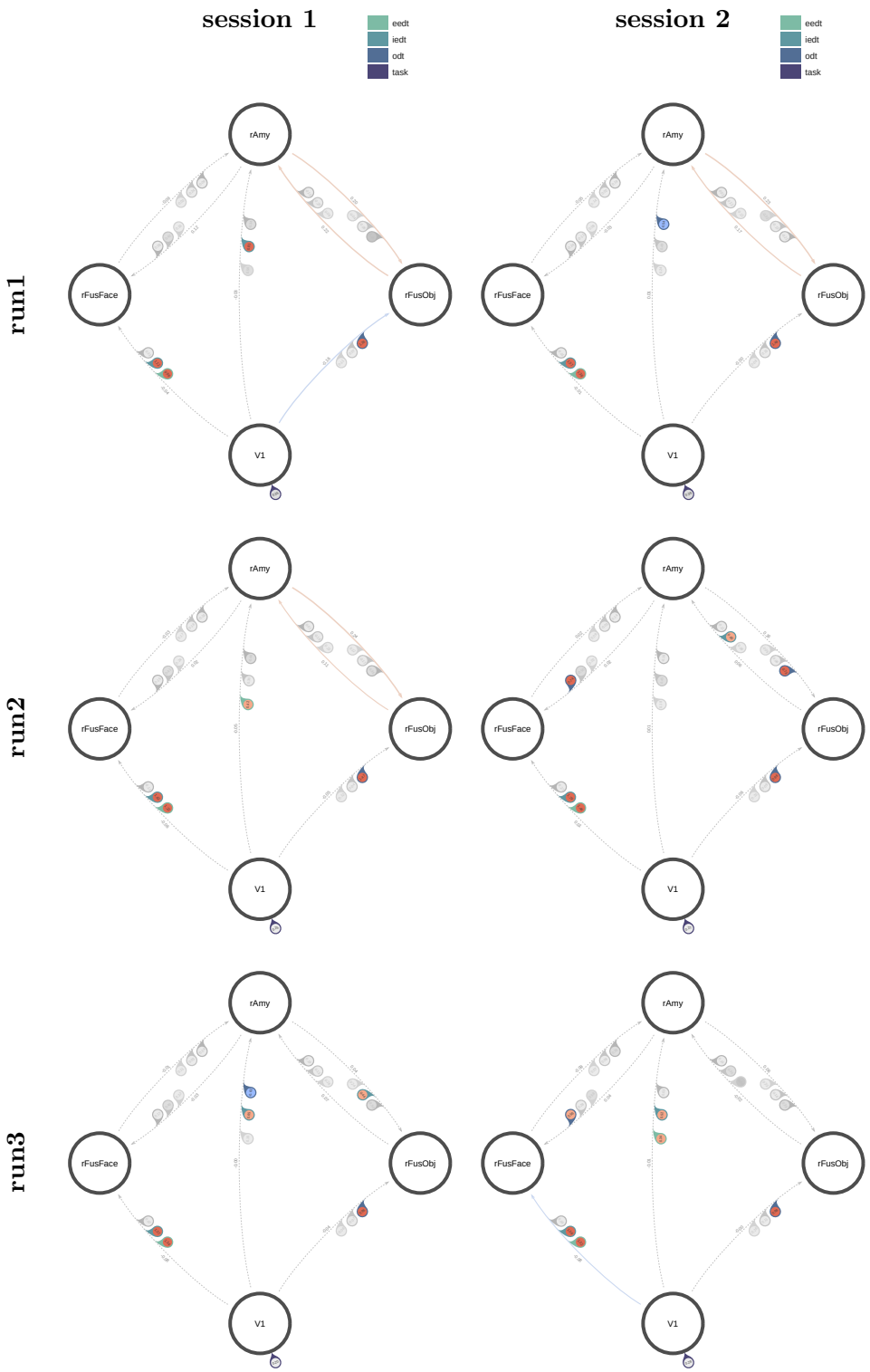

Figure 4.12: DCM of 1st, 2nd and 3rd run of each session respectively, TR =1.4s as described in figure 4.11.

	TR 1.4	TR 0.7
Session 1 / Run 1	0.1822	0.0398
Session 1 / Run 2	0.1366	0.1948
Session 1 / Run 3	0.1038	0.0771
Session 2 / Run 1	0.2158	0.0559
Session 2 / Run 2	0.2949	0.1547
Session 2 / Run 3	0.2332	0.1583
mean ± std	0.1944 ± 0.069	0.1134 ± 0.0639
median ± std	0.199 ± 0.069	0.1159 ± 0.0639

Table 4.4: ICC(3,1) over subjects (raters) and connectivity parameters (targets), for each run and TR.

	TR 1.4	TR 0.7
Sub 1	0.0816	0.0316
Sub 2	-0.0172	-0.0188
Sub 3	-0.0187	-0.0092
Sub 4	0.0166	0.0194
Sub 5	0.0157	0.0237
Sub 6	-0.0146	-0.0118
Sub 7	0.0241	0.0006
Sub 8	0.0115	0.0536
Sub 9	0.0144	-0.0047
Sub 10	0.0136	0.0306
Sub 11	0.0367	0.0234
Sub 12	0.0353	0.0191
Sub 13	0.0043	0.0185
Sub 14	0.035	0.0235
mean ± std	0.017 ± 0.0262	0.0142 ± 0.0202
median ± std	0.015 ± 0.0262	0.0193 ± 0.0202

Table 4.5: ICC(3,1) over all runs of both sessions (raters) and connectivity parameters (targets), for each subject and TR.

	TR 1.4	TR 0.7
Sub 1	0.7835	0.1871
Sub 2	0.6532	0.7662
Sub 3	-0.0484	0.5986
Sub 4	0.5839	0.3467
Sub 5	0.428	0.2285
Sub 6	0.5723	0.9287
Sub 7	0.698	-0.0371
Sub 8	0.7335	0.7267
Sub 9	0.9088	0.7859
Sub 10	0.5567	0.0433
Sub 11	-2.1974	-0.235
Sub 12	0.4485	0.6373
Sub 13	0.6835	0.793
Sub 14	0.9102	0.7159
mean ± std	0.4082 ± 0.7868	0.4633 ± 0.3682
median ± std	0.6185 ± 0.7868	0.618 ± 0.3682

Table 4.6: ICCs(3,k) over means of all runs of each sessions (raters) and connectivity parameters (targets), for each subject and TR.

details on the ICC) for each subject over runs, for each run over subjects, for each TR respectively (see tables 4.4 and 4.5); as well as the intraclass correlation of type ICC(3,k) for means of all 3 runs for each session (see table 4.6).

We can see that despite the higher temporal resolution of TR = 0.7s, the intersubject ICC in each run (table 4.4) is remarkably lower than for TR = 1.4s, however it is relatively stable over runs with no apparent outliers.

The intrasession ICCs for each subject vary very little, although they range very close to 0 – describing close to no correlation between runs. On the other hand, intersession ICCs over session means deliver fair to good intraclass correlation, according to the scale defined in Fleiss (2011) – depending on consideration of mean or median of the ICCs.

Chapter 5

Discussion / further work

The conclusion of the experiments performed can be split in 2 parts – for once, I have obtained valuable observations on stability of activation when performing an emotion discrimination task during functional MRI; on the other hand the DCM results have brought valuable insights.

5.1 On the GLM results

I have shown that for multiple sessions, the (explicit) emotion discrimination task reliably provides robust and reproducible activation patterns in the bilateral amygdalae, fusiform gyrus and the left DLPFC. Using the implicit EDT, comparable activation patterns can be obtained for the amygdalae and fusiform gyrus. Both tasks thus seem perfectly suited for large-scale and longitudinal studies and repeated measurements.

Although observing a slight decrease in parameter estimates after the first run of the emotion discrimination task, the values and significance in the subsequent runs and sessions are very similar. Taking into account that each run was performed in total 6 times per session (each 3 per time of repetition), and no further decline of the parameter estimates from the second to the third run could be observed, this indicates very good repeatability of the task without significant loss in effect magnitude.

Our findings are in line with a previous study performed on a 3 Tesla system, where amygdala habituation was observed only in social anxiety disorder patients, but not in healthy subjects (Sladky et al., 2012).

For both repetition times 1.4s and 0.7s I have found highly comparable activation magnitude, thus very similar results (see figure 4.8), despite of the lower SNR at a lower TR. This indicates that also with a lower TR, it is pos-

sible to obtain data on activation in the emotion processing network without significant effect loss, thus good suitability of this method for analysis tools relying on high temporal resolution, such as Dynamic Causal Modelling.

However, from the performed study, one important question remains unanswered – i.e. whether amygdalar activation occurs when seeing human faces alone or if there it is a reaction towards the emotions shown on the faces.

5.2 On the DCM results

We can see that our DCM results somehow reflect the GLM results in terms of the fusiform gyrus – where activity in the visual cortex (V1) leads to activity in the fusiform face area (FFA) for face stimuli, and in the fusiform object area (FOA) for the object stimuli. However, despite stable activation in the amygdala, our DCM does not result in a consistent modulatory influence on the amygdala. I assume that is because the source for amygdalar modulation is not unique here, as modulatory connections were modelled from both the visual cortex and the fusiform gyrus.

Intraclass correlation (ICC) evaluation did not reveal high stability of DCMs over subjects or runs. However, due to the nature of the ICC, this might occur because of too little intraclass variance or too high interclass variance, as these play a crucial part in the calculation. Also, using ICCs only 2 dimensions can be evaluated at once – i.e. the target and the rater.
It might be useful to turn to a measure taking into account multiple sources of variance, such as the so-called analysis of variance (ANOVA). However, this is a method for measuring significant differences instead of consistency, therefore only inferences of non-equality (or non-stability) could be made.

Further work on this topic could include

- Amygdalar modulation from FFA and V1 could be modelled separately; inferences can then be made based on Bayesian Model Selection.
- Another strategy would be concatenating several runs into one single time course and estimating a whole-session DCM on this data, for intersession comparison.
- As there are several more brain regions in the GLM results, other model hypotheses could be formulated.
- Stability can be analysed by performing tests on the approximation on the model evidence, i.e. the negative free energy.

- Last but not least, one possibility for obtaining more significant results might be the inclusion of more subjects in the study.

Also, when performing further analyses, the limitations of DCM must be understood: The major point of criticism concerning DCM is the absence of a model validation tool. Inferences are only possible in terms of deciding between several hypotheses and based on these, the most probable effective connectivity parameters can be estimated. Still, it is not possible to determine whether the winning model of this procedure is roughly right. For finding a better fitting model, one would need exhaustive hypotheses including all possibilities – which is next to impossible due to the complex nature and perhaps lack of understanding of the human brain; let alone the computational limitations. Therefore, DCM is a tool to be used with caution; users must have excellent knowledge on possible mechanisms to be modelled and understand the limitations that come with this method.

Chapter 6

Acknowledgements

First of all, I would like to thank Christian Windischberger and his team including Martin, Ronny, Michi, Allan, André and Jürgen for making me feel welcome in their lab and for helping me dive into the exciting world of functional MRI.

My warmest thanks go to Günter Schneckenreither and Felix Breitenecker for guiding me through this thesis, for all the fruitful discussions and valuable feedback.

Especially, I would like to express my gratitute to my family for always standing by me since the very beginning – to my parents, siblings, grandparents, aunts, uncles, cousins, and Corinne.

One person particularly contributed to sparking off my fascination for mathematics and has thus played a considerable role in setting me on my path to this point. To my mathematics teacher from secondary school: It's all your fault!

Last but not least, I have to mention all the joy and adventures I had over the last years with the family that I chose, including (alphabetically arranged) Claudia, Dominik, Miriam, Phil, Verena and many others who will hopefully forgive me for not mentioning them.

Bibliography

Box, G. E. (1976). Science and statistics. *Journal of the American Statistical Association*, 71(356):791–799.

Breakspear, M. (2013). Dynamic and stochastic models of neuroimaging data: A comment on lohmann et al. *Neuroimage*, 75:270–274.

Buxton, R. B., Wong, E. C., and Frank, L. R. (1998). Dynamics of blood flow and oxygenation changes during brain activation: the balloon model. *Magnetic resonance in medicine*, 39(6):855–864.

Daunizeau, J., David, O., and Stephan, K. E. (2011). Dynamic causal modelling: a critical review of the biophysical and statistical foundations. *Neuroimage*, 58(2):312–322.

Fleiss, J. L. (2011). *Design and analysis of clinical experiments*, volume 73. John Wiley & Sons.

Frässle, S., Paulus, F. M., Krach, S., and Jansen, A. (2016). Test-retest reliability of effective connectivity in the face perception network. *Human brain mapping*, 37(2):730–744.

Frässle, S., Stephan, K. E., Friston, K. J., Steup, M., Krach, S., Paulus, F. M., and Jansen, A. (2015). Test-retest reliability of dynamic causal modeling for fmri. *Neuroimage*, 117:56–66.

Friston, K., Daunizeau, J., and Stephan, K. E. (2013). Model selection and gobbledygook: Response to Lohmann et al. *Neuroimage*, 75:275–278.

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

Friston, K. J. (2002). Bayesian estimation of dynamical systems: an application to fMRI. *NeuroImage*, 16(2):513–530.

Friston, K. J., Holmes, A. P., Poline, J., Grasby, P., Williams, S., Frackowiak, R. S., and Turner, R. (1995). Analysis of fMRI time-series revisited. *Neuroimage*, 2(1):45–53.

Friston, K. J., Holmes, A. P., Worsley, K. J., Poline, J.-P., Frith, C. D., and Frackowiak, R. S. (1994). Statistical parametric maps in functional imaging: a general linear approach. *Human brain mapping*, 2(4):189–210.

Friston, K. J., Mechelli, A., Turner, R., and Price, C. J. (2000). Nonlinear responses in fMRI: the balloon model, volterra kernels, and other hemodynamics. *NeuroImage*, 12(4):466–477.

Geissberger, N., Sladky, R., Tik, M., Woletz, M., Hoffmann, A., Willinger, D., Robinson, S., and Windischberger, C. (2017). Reliability of Amygdala fMRI at Ultra-High Field. *in preparation*.

Goldsmith, E. (1972). Words and models (systems approach to linguistics). *Kybernetes*, 1(4):243–249.

Hariri, A. R., Mattay, V. S., Tessitore, A., Kolachana, B., Fera, F., Goldman, D., Egan, M. F., and Weinberger, D. R. (2002). Serotonin transporter genetic variation and the response of the human amygdala. *Science*, 297(5580):400–403.

Harville, D. A. (1977). Maximum likelihood approaches to variance component estimation and to related problems. *Journal of the American Statistical Association*, 72(358):320–338.

Langner, O., Dotsch, R., Bijlstra, G., Wigboldus, D. H., Hawk, S. T., and van Knippenberg, A. (2010). Presentation and validation of the radboud faces database. *Cognition and emotion*, 24(8):1377–1388.

Logothetis, N. K., Pauls, J., Augath, M., Trinath, T., and Oeltermann, A. (2001). Neurophysiological investigation of the basis of the fMRI signal. *Nature*, 412(6843):150–157.

Lohmann, G., Erfurth, K., Müller, K., and Turner, R. (2012). Critical comments on dynamic causal modelling. *Neuroimage*, 59(3):2322–2329.

Lohmann, G., Müller, K., and Turner, R. (2013). Response to commentaries on our paper: critical comments on dynamic causal modelling. *Neuroimage*, 75:279–281.

Longford, N. (1987). A fast scoring algorithm for maximum likelihood estimation in unbalanced mixed models with nested random effects. *ETS Research Report Series*, 1987(1).

Mandeville, J. B., Marota, J. J., Ayata, C., Zaharchuk, G., Moskowitz, M. A., Rosen, B. R., and Weisskoff, R. M. (1999). Evidence of a cerebrovascular postarteriole windkessel with delayed compliance. *Journal of Cerebral Blood Flow & Metabolism*, 19(6):679–689.

Murray-Smith, D. (2012). *Modelling and simulation of integrated systems in engineering: issues of methodology, quality, testing and application.* Woodhead Publishing, Cambridge, UK.

Neal, R. M. and Hinton, G. E. (1998). A view of the EM algorithm that justifies incremental, sparse, and other variants. In *Learning in graphical models*, pages 355–368. Springer.

Ogawa, S., Lee, T.-M., Kay, A. R., and Tank, D. W. (1990). Brain magnetic resonance imaging with contrast dependent on blood oxygenation. *Proceedings of the National Academy of Sciences*, 87(24):9868–9872.

Paulus, M. P., Feinstein, J. S., Castillo, G., Simmons, A. N., and Stein, M. B. (2005). Dose-dependent decrease of activation in bilateral amygdala and insula by lorazepam during emotion processing. *Archives of General Psychiatry*, 62(3):282–288.

Penny, W. D., Friston, K. J., Ashburner, J. T., Kiebel, S. J., and Nichols, T. E. (2011). *Statistical parametric mapping: the analysis of functional brain images*. Academic press.

Penny, W. D., Stephan, K. E., Daunizeau, J., Rosa, M. J., Friston, K. J., Schofield, T. M., and Leff, A. P. (2010). Comparing families of dynamic causal models. *PLoS Comput Biol*, 6(3):e1000709.

Penny, W. D., Stephan, K. E., Mechelli, A., and Friston, K. J. (2004). Comparing dynamic causal models. *Neuroimage*, 22(3):1157–1172.

Plichta, M. M., Schwarz, A. J., Grimm, O., Morgen, K., Mier, D., Haddad, L., Gerdes, A. B., Sauer, C., Tost, H., Esslinger, C., Colman, P., Wilson, F., Kirsch, P., and Meyer-Lindenberg, A. (2012). Test–retest reliability of evoked BOLD signals from a cognitive–emotive fMRI test battery. *NeuroImage*, 60:1746–1758.

Robinson, S. D., Pripfl, J., Bauer, H., and Moser, E. (2008). The impact of EPI voxel size on SNR and BOLD sensitivity in the anterior medio-temporal lobe: a comparative group study of deactivation of the default mode. *Magnetic Resonance Materials in Physics, Biology and Medicine*, 21(4):279–290.

Sabatinelli, D., Bradley, M. M., Fitzsimmons, J. R., and Lang, P. J. (2005). Parallel amygdala and inferotemporal activation reflect emotional intensity and fear relevance. *Neuroimage*, 24(4):1265–1270.

Sachs, L. (2013). *Angewandte statistik: anwendung statistischer methoden.* Springer-Verlag.

Scheffe, H. (1959). The analysis of variance. 1959. *New York*, pages 331–367.

Shrout, P. E. and Fleiss, J. L. (1979). Intraclass correlations: Uses in assessing rater reliability. *Psychological Bulletin*, 86:420–428.

Sladky, R., Baldinger, P., Kranz, G. S., Tröstl, J., Höflich, A., Lanzenberger, R., Moser, E., and Windischberger, C. (2013). High-resolution functional MRI of the human amygdala at 7T. *European journal of radiology*, 82(5):728–733.

Sladky, R., Höflich, A., Atanelov, J., Kraus, C., Baldinger, P., Moser, E., Lanzenberger, R., and Windischberger, C. (2012). Increased neural habituation in the amygdala and orbitofrontal cortex in social anxiety disorder revealed by fMRI. *PloS one*, 7(11):e50050.

Sladky, R., Höflich, A., Küblböck, M., Kraus, C., Baldinger, P., Moser, E., Lanzenberger, R., and Windischberger, C. (2015). Disrupted effective connectivity between the amygdala and orbitofrontal cortex in social anxiety disorder during emotion discrimination revealed by dynamic causal modeling for fMRI. *Cerebral Cortex*, 25(4):895–903.

Stein, M. B., Simmons, A. N., Feinstein, J. S., and Paulus, M. P. (2007). Increased amygdala and insula activation during emotion processing in anxiety-prone subjects. *American Journal of Psychiatry*, 164(2):318–327.

Stephan, K. E., Weiskopf, N., Drysdale, P. M., Robinson, P. A., and Friston, K. J. (2007). Comparing hemodynamic models with DCM. *Neuroimage*, 38(3):387–401.

Tik, M., Woletz, M., de Lara, L. N., Sladky, R., Hoffmann, A., Hummer, A., and Windischberger, C. (2017). Mapping TMS local and remote immediate effects by concurrent TMS/fMRI using a dedicated high-sensitivity coil array. *Brain Stimulation*, 2(10):489–491.

Wang, S., Yu, R., Tyszka, J. M., Zhen, S., Kovach, C., Sun, S., Huang, Y., Hurlemann, R., Ross, I. B., Chung, J. M., et al. (2017). The human amygdala parametrically encodes the intensity of specific facial emotions and their categorical ambiguity. *Nature Communications*, 8.