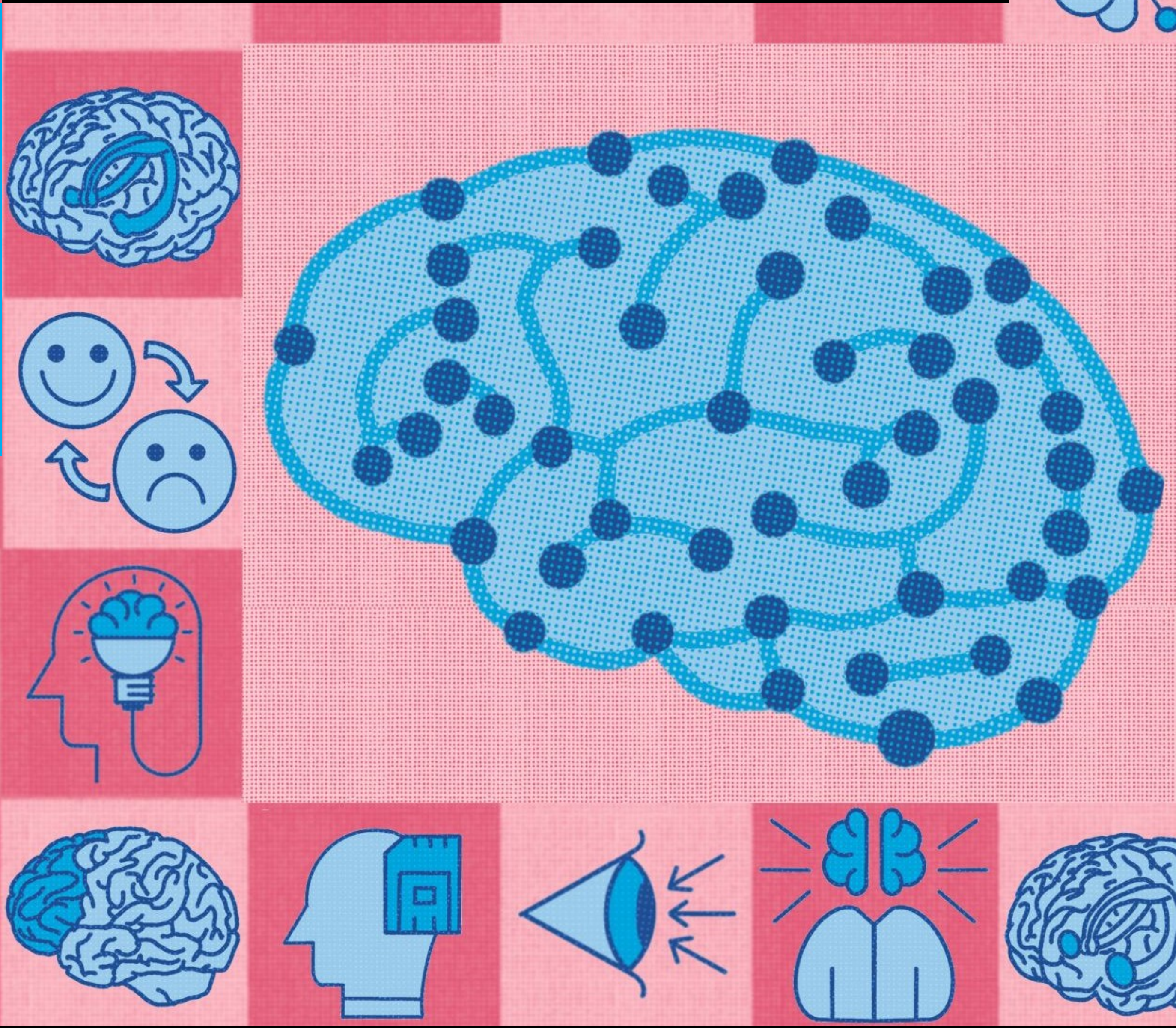
Analyzing the Neural Dynamics of Emotional Associative Memory Encoding through Data-Driven Modeling

W. Dziarnowska

Master of Science Thesis



TU Delft Delft University of Technology

Delft Center for Systems and Control

Analyzing the Neural Dynamics of Emotional Associative Memory Encoding through Data-Driven Modeling

Master of Science Thesis

For the degree of Master of Science in Systems and Control at Delft University of Technology

W. Dziarnowska

December 4, 2023

Faculty of Mechanical, Maritime and Materials Engineering (3mE) · Delft University of Technology

The work in this thesis was supported by Donders Institute for Brain, Cognition and Behaviour. Their cooperation is hereby gratefully acknowledged.

Abstract

Researchers have been interested in studying the connection between emotion and memory for decades but much remains unknown due to the elusive nature of the human brain. Furthering our understanding of the phenomenon is crucial for improving the treatment of neurological disorders associated with emotion dysregulation, as well as for enhancing autonomous systems that interact with humans, such as robotic prostheses and artificial intelligence. A promising approach to studying cognitive processes is developing computational models of brain activity, which have the potential to uncover the underlying neural mechanisms.

This master's thesis presents a novel dynamical model of the neural activity underpinning the process of emotional associative memory encoding. This type of memory is especially complex and poorly understood as it requires memorizing the relationships between various environmental stimuli that may elicit distinct emotions. To model this phenomenon, the thesis proposes a network model using the Dynamic Causal Modeling (DCM) framework, focusing on the Amygdala (Amy), the Hippocampus (Hip), and the Orbitofrontal Cortex (OFC) due to their central roles in emotional associative memory. Furthermore, the thesis provides an analysis of network state coordination and state error stability, deriving analytical bounds for state error dynamics that illustrate how the model's properties are linked to improved memory encoding.

The dataset used to estimate the DCM comes from a functional Magnetic Resonance Imaging (fMRI) study conducted by Zhu et al. at Donders Institute for Brain, Cognition, and Behaviour [1], where participants were asked to memorize pairs of two emotionally potent images (group Emotional-Emotional (EE)), two neutral images (group Neutral-Neutral (NN)), or one neutral and one emotional image (group Neutral-Emotional (NE)). The DCM model developed in this thesis reflects how the neural dynamics differ among the groups and among the modeled brain regions, corroborating many of the behavioral results of the experiment and bringing novel insights. Group NE is found to exhibit the best overall information flow, leading to the best subsequent memory, and the coupling between Amy-Hip is shown to play a key role in the improved memory encoding. Conversely, the connectivity between Hip-OFC in all groups appears to be the weakest but the most sensitive to external stimuli, indicating a key role of this connection in responding to the environment. Moreover, there is considerable evidence that the external inputs enter the network through the OFC in all groups.

Another important contribution of this thesis is the extensive overview of DCM in Chapter 2, which discusses the mathematical foundations, the system identification algorithm, criticism raised against the method, studies of its statistical validity, and advice on implementation. Additionally, Appendix B presents the first guide for DCM script development in the Statistical Parametric Mapping (SPM12) toolbox in MATLAB. Furthermore, this thesis compares the performance of four different variations of DCM, proving that Stochastic DCM achieves superior results in experimental paradigms with brief emotional stimuli. The success of this approach is believed to stem from its capacity to handle model misspecification and neuronal noise. Overall, the Stochastic DCM developed in this thesis is considered to provide the most comprehensive representation of neural dynamics governing the encoding of emotional associative memory when compared to other existing dynamical models.

Table of Contents

Preface ix

1 Introduction 1

1-1 Motivation 1

1-1-1 Understanding human behavior 2

1-1-2 Improving mental health treatment 3

1-1-3 Advancing neuroengineering and artificial intelligence 4

1-2 Background 4

1-2-1 The role of the amygdala 5

1-2-2 The role of the hippocampus 6

1-2-3 The role of the prefrontal cortex 7

1-3 Related work 8

1-4 Problem statement 10

1-5 Thesis outline 11

2 Theoretical framework 13

2-1 Data-driven models of brain activity 13

2-1-1 Brain imaging techniques 13

2-1-2 Modeling frameworks 16

2-2 Fundamentals of Dynamic Causal Modeling 18

2-2-1 Dynamical model 18

2-2-2 System identification 21

2-2-3 Practical considerations 25

2-3 Limitations and validity of Dynamic Causal Modeling 27

2-3-1 Biophysical foundations 28

2-3-2 Statistical foundations 28

2-3-3 Framework validity 31

2-4 Developments in Dynamic Causal Modeling 33
2-4-1 Nonlinear DCM 33
2-4-2 Two-state DCM 35
2-4-3 Stochastic DCM 36

3 Model development 41
3-1 Modeling approach 41
3-1-1 Experimental paradigm 41
3-1-2 Input signal 43
3-1-3 Model space 43
3-2 Model fitting 46
3-2-1 Data pre-processing 46
3-2-2 Subject-level DCM 48
3-2-3 Group-level DCM 49

4 Results and analysis 51
4-1 Model comparison 51
4-2 Analysis of the selected model 53
4-2-1 Model structure 53
4-2-2 Estimated parameters 57
4-2-3 Stability analysis 62

5 Conclusions 71
5-1 Discussion of the results 73
5-2 Future research 76

A Additional results and figures 79

B A quick guide to SPM12 DCM scripting 87
B-1 Main code 87
B-2 Time-series extraction 89
B-3 Subject-level DCM 91
B-4 Group-level DCM 95

Bibliography 99

Glossary 109
List of Acronyms 109
List of Symbols 110

List of Figures

1-1 Elements of the Medial Temporal Lobe. 5
1-2 Regions of the Prefrontal Cortex. 7
1-3 Inputs, states, and outputs in DCM. 8

2-1 Example of brain activation measured with fMRI. 14
2-2 Typical shape of the Hemodynamic Response Function. 15
2-3 Diagram of generative models of neuronal dynamics. 17
2-4 Graphical meaning of coupling matrices in DCM 19
2-5 Prior probability density functions of the temporal scaling factor in DCM. 24
2-6 Typical sequence of analysis in DCM. 27
2-7 Example of splitting data for cross-validation. 30
2-8 Predictive validity of Dynamic Causal Modeling. 32
2-9 Graphical meaning of coupling matrices in Nonlinear DCM. 34
2-10 Comparison between the classical and the two-state model in DCM. 36

3-1 Experimental design used to obtain data for this project. 42
3-2 Example of input signals used in DCM. 43
3-3 Model space for DCM. 45
3-4 Model space for Nonlinear DCM. 46
3-5 Example of a GLM observation model. 47
3-6 Definition of voxels constituting VOIs. 48
3-7 Histogram of the distribution of the number of relevant voxels in each node. . . . 49

4-1 Results of BMS for Stochastic DCM. 55
4-2 Example response of the average group model to the inputs of a subject. 57
4-3 Ranges of A matrix values across the groups. 58

4-4 Changes in A caused by the two inputs in group NN. 60
4-5 Changes in A caused by the two inputs in group NE. 60
4-6 Changes in A caused by the two inputs in group EE. 61
4-7 State and state-error evolution in group NN. 69
4-8 State and state-error evolution in group NE. 69
4-9 State and state-error evolution in group EE. 70

A-1 Average optimal model vs. data for Classical DCM 81
A-2 Average optimal model vs. data for Nonlinear DCM 82
A-3 Average optimal model vs. data for Two-State DCM 83
A-4 Average optimal model vs. data for Stochastic DCM 84
A-5 Zoomed-in average optimal model vs. data for Stochastic DCM 85

B-1 The SPM12 GUI - main menu and batch editor. 89
B-2 The SPM12 GUI - BPA menu. 97

List of Tables

2-1 Estimated time of system identification for varying sizes of Dynamic Causal Models. 28

3-1 Between-region connections in the A matrix found in similar studies. 44

4-1 The R-squared of optimal average models identified using different DCM variations. 53
4-2 Absolute log model evidence values obtained during BMS. 54
4-3 R-squared values in each node for all average Stochastic DCM models. 55
4-4 Number of significant voxels in VOIs used for BPA. 56
4-5 Algebraic connectivity of graphs and subgraphs of optimal Stochastic DCMs. . . 64

5-1 Average hippocampal coupling strength. 75

A-1 Estimated parameters in the optimal Stochastic DCM models. 79
A-2 Estimated A and A' matrices in the optimal Stochastic DCM models. 80

Preface

This thesis has been an incredible journey into the world of neuroscience, broadening my horizons and teaching me how to carry out interdisciplinary research. My heartfelt appreciation extends to my supervisors Dr. Matin Jafarian and Maria Bartzioka from the Delft Center for Systems and Control (DCSC), whose guidance, wealth of knowledge, and continued encouragement empowered me to strive for excellence. Additionally, I am profoundly thankful to Dr. Nils Kohn from the Donders Institute for Brain, Cognition and Behavior, whose generous sharing of neuroscientific expertise has enriched my project.

I would also like to thank my dear friend Zofia Borowska who has designed the beautiful cover of my thesis and has been my rock throughout the ups and downs of my university experience. Lastly, of course, I express profound appreciation to my family — my pillars of strength, Mom, Dad, and sisters Amelia and Ola — for their unwavering belief in me and support throughout this endeavor.

Delft, University of Technology
December 4, 2023

W. Dziarnowska

To my beloved parents, sisters, and friends who have always made me believe that I could achieve anything I desired.

Chapter 1

Introduction

Emotions play a crucial role in shaping our memories - we tend to remember things more vividly when they are tied to strong emotions like significant life events. While joyous or traumatic experiences are generally remembered better, emotions can also have a detrimental impact, distorting or blocking associated memories. The two brain regions that are typically considered to be the most crucially involved in emotional memory are the Amygdala (Amy) and the Hippocampus (Hip). The amygdala is responsible for the acquisition and expression of conditioned emotional responses, that is the association of stimuli from the sensory cortex with an emotional value [2]. Furthermore, the hippocampus serves a key role in declarative memory, which is the conscious, intentional recollection of past events and facts [3].

To gain insight into the exact dynamics of how emotions and memory interact, neuroimaging data can be used to create computational models of brain activity. The goal of this thesis is to construct such a model using data from an experiment where subjects were asked to memorize images that elicit emotional responses. Furthermore, the resulting model is analyzed to draw conclusions about emotional memory and its neural underpinnings.

The aim of this chapter is to provide the necessary background information and to define the problem statement for the experimental part of the thesis. The motivation for modeling emotional memory is explained in Section 1-1, the relevant concepts from neuroscience are introduced in Section 1-2, the related work is summarized in Section 1-3, the problem statement is introduced in Section 1-4, and the outline of the remainder of this thesis is described in Section 1-5.

1-1 Motivation

Emotions play a crucial role in human behavior, shaping the way we make decisions, learn, and interact in a society. Disturbance of emotional learning is at the core of disorders like depression or anxiety, which are a severe burden to the individual and to society. Modeling emotional learning can thus give researchers valuable insight into what brain regions are involved and how they interact with each other, shaping this highly complex process. Research in this field contributes to a better understanding of human behavior, which in turn finds applications in improving mental health treatment of disorders associated with memory and emotion impairment, as well as enhancing autonomous systems that interact with humans.

1-1-1 Understanding human behavior

Thanks to the numerous studies on the subject, it is now well-established that emotion significantly influences learning. As highlighted in the sections above, this interaction is highly complex and can either enhance or impair memory. Much remains unknown in this matter and studies report highly task-dependent and sometimes conflicting results. It is thus important to keep studying the phenomenon to understand exactly in what situations emotions help or disturb learning, such that findings are not contradictory but rather complementary.

One example of opposing findings is that improved learning and academic achievement have been linked to both positive and negative emotions. While positive emotions help in learning by stimulating motivation and self-satisfaction [4], negative states related to confusion around the study material aid in learning by promoting low-level anger and enhanced seeking of material explanation [5]. On the other hand, even the same emotion can enhance or impair memory, based on its intensity and duration. In particular, mild and short-term stress has been found to facilitate cognitive processes, while severe and chronic stress impairs learning and memory retrieval [6]. Emotion is thus said to influence hippocampal-dependent memory in an inverted U-shaped function [7].

Memory impairment caused by emotional arousal is especially noticeable in associative memory, which is used to learn the relationship between unrelated items. Several studies using pairs of emotional and neutral stimuli have demonstrated reduced associative memory for aversive pairs, accompanied by increased activity in the amygdala [8]. One study supporting this view has investigated the root of this phenomenon using image pairs [9]. Based on the evidence from other literature, the authors have formulated two possible hypotheses for impaired associative learning of emotional stimuli. The authors have observed increased hippocampal activity for later remembered negative image pairs but generally, associative memory for neutral pairs was better. These results are in favor of the so-called "bypassing" hypothesis, which states that hippocampal activity increases when associations can be unitized, improving associative learning of the pair. Unitization is the process by which separate elements or features of a stimulus are combined into a single, integrated representation that is easier to remember. The study has suggested that emotionally valenced pairs do not induce unitization as easily as their neutral counterparts. This is because emotional material has inherently distracting and arousing properties. Neutral items, on the other hand, have been found to easily unitize even without cognitive strategies if they form a meaningful or familiar combination [10].

The hypothesis about unitization enhancing associative learning is only one possible idea. In fact, there is evidence supporting the second suggested hypothesis as well. This other theory is called the "disruption" hypothesis. It states that associative memory for emotional stimuli is impaired because the increased amygdala activity lowers hippocampus activity and thus, disrupts hippocampal-dependent memory. The study in [11] has used the same paradigm as [9], yet the authors have found reduced hippocampal activity (opposite to the other study) and later recall of all negative image pairs, which is consistent with the disruption hypothesis. The authors of [9] have hypothesized that the contradicting results stem from different fMRI scanning resolutions, statistical approaches, or the emotional nature of the images and the emotional involvement of the participants. The differences could also be caused by the subjects as [9] has examined 84 male participants, whereas [11] has examined 11 female and 9 male participants. There is vast evidence for significant emotional memory differences between genders, with [12, 13] reporting that men and women activated different neural circuits during the encoding of emotional material. While men had more activation in the right amygdala, women had increased activation in the left amygdala. The authors have also found that females remembered the emotional stimuli better than men.

The study on associative memory of word-image pairs by [14] reports yet another conclusion wherein associative memory is actually increased for emotional material. The authors have stated that their result is consistent with previous studies but the ones they have listed examined only emotional learning for single items, not associative memory. This is not to say that these results were incorrect but this example illustrates a problematic trend in the field. Researchers sometimes make a conclusion about the impact of emotion on learning and state that it is consistent with previous findings, without acknowledging contradictory studies. All the studies discussed in this section demonstrate how much is still unknown about the nature of emotional learning and how subtle methodological differences seem to produce discrepant results.

1-1-2 Improving mental health treatment

Apart from purely advancing human knowledge of neuroscience and psychology, understanding emotional learning is vital for improving mental health treatments for numerous disorders associated with emotional dysregulation. A common treatment relies on exposure therapy, which involves gradually exposing the patient to the source of their anxiety or trauma in a safe and controlled environment [15]. This therapy relies on extinction learning, which is a decrease in response to a conditioned stimulus that occurs when the stimulus is presented without reinforcement. Understanding fear conditioning is thus crucial to improve this sort of treatment. Examples of anxiety disorders include Post-Traumatic Stress Disorder (PTSD), associated with intrusive memories of traumatic experiences, and Obsessive-Compulsive Disorder (OCD), which causes distressing thoughts and ritualistic behaviors performed to alleviate anxiety. Yet another disorder that benefits from this treatment is the Substance Use Disorder (SUD) associated with abnormal reward processing and impaired decision-making [16].

Apart from behavioral approaches, mental disorders can be treated with neuromodulation, which involves activating certain brain areas using agents that are electrical (e.g. Deep Brain Stimulation (DBS) [17] and Vagus Nerve Stimulation (VNS) [18]) or magnetic (e.g. Transcranial Magnetic Stimulation (TMS) [19]). Developing these techniques relies on knowledge of the neural circuitry and the mechanics of emotional learning. These neuromodulation techniques have been found effective in treating disorders associated with emotion dysregulation such as OCD, SUD, as well as depression, a disorder that causes people to attend to negative stimuli and interpret ambiguous information negatively.

1-1-3 Advancing neuroengineering and artificial intelligence

Models of various cognitive processes, including emotions and learning, are crucial for the development of Brain-Computer Interfaces (BCIs). These are systems that enhance the functioning of patients who are affected by various disabilities. BCIs can serve as a direct communication pathway between the brain and external devices, including robotic limbs and tools for speaking and writing. This is accomplished by measuring a biosignal and using it to estimate the person's cognitive state. Incorporating emotional models into these interfaces would allow the system to better recognize the meaning of neural activity changes induced by emotional states [20]. Moreover, it would enable the users to apply more natural strategies to control the device. Another type of BCIs is neuroprostheses, which enhance the activity of a neural system in patients who suffered a brain injury that caused cognitive impairment. As a result of the trauma, various deficits may arise, including memory loss and personality changes. The subject of detecting and incorporating emotions into BCIs is an active area of research, with some of the most notable developments summarized in reviews such as [21, 22].

Another application of emotional learning models is in enhancing Artificial Intelligence (AI) systems as outlined in several reviews [23, 24]. First of all, computational models of emotions allow AI to generate synthetic emotional responses, which makes them more believable and able to interact with people on an emotional level. As emotions shape human cognitive processes involving decision-making, attention, and memory, adding such features to AI systems can help them adapt to changing situations or make appropriate decisions based on their goals. However, incorporating emotions into AI systems is difficult, and much progress remains to be made to create highly complex models that account for varying scenarios, social and cultural contexts, and internal states.

Secondly, by understanding how the brain learns, researchers can develop bio-inspired algorithms for training AI. One of the most notable developments in this field is the Free Energy Principle (FEP), which is a unified theory of how biological systems, including the brain, learn and adjust to the environment. The principles of the FEP theory have been utilized to create various adaptive learning algorithms that efficiently combine the autonomous agent's prior knowledge with sensory inputs from the environment. In fact, the FEP is at the core of the system identification algorithm utilized in this thesis.

1-2 Background

The amygdala and the hippocampus are located in the Medial Temporal Lobe (MTL) depicted in Figure 1-1 and they both play crucial roles in governing emotional memory. A key study on the two brain regions has shown that the systems are independent, yet they interact in a complex and meaningful way [25]. In the experiment, patients with amygdalar and/or hippocampal damage were tested in a classical conditioning paradigm. The participants were shown multiple slides of different colors, including blue slides, which were followed immediately by a startling sound. After completion of the conditioning phase, the patients were asked to name the different colors they saw and indicate which color was followed by a sound. In the last experiment phase, only blue slides without sound were presented to examine whether conditioning was acquired. To measure the physiological reaction, the skin conductance response was measured. The experiment's outcome was that different brain lesions blocked

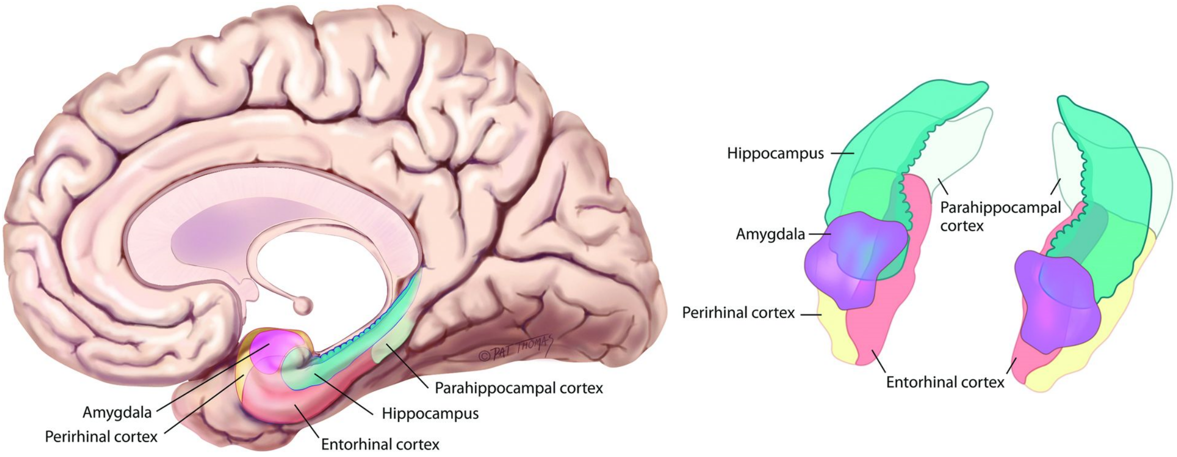


Figure 1-1: Elements of the Medial Temporal Lobe (MTL) from [26].

different aspects of memory. The patient with amygdalar damage failed to acquire a normal physiological fear response to the blue slides. She did however recall the slide colors and the fact that a sound followed the blue one. In contrast, the patient with hippocampal damage did in fact exhibit fear conditioning but failed to acquire new facts about the stimuli presented in the experiment. Finally, in the case of the third patient, damage to both the amygdala and the hippocampus blocked the acquisition of both fear conditioning and facts.

Another study involving subjects with varying degrees of amygdalar or hippocampal damage has examined brain activation during a memorization task with neutral and aversive words [27]. Brain signals were measured using a neuroimaging technique called functional Magnetic Resonance Imaging (fMRI) and described in detail in Section 2-1-1. While the level of amygdalar damage predicted memory performance for only emotional words, the severity of left hippocampal pathology predicted memory performance for both neutral and emotional phrases. Furthermore, bigger left amygdalar pathology was associated with reduced activity in the left hippocampus during the encoding of emotional words as compared to neutral words. Interestingly, bigger left hippocampal pathology predicted a decreased activity in both the left and the right amygdala. The authors have suggested that the Amy-Hip interaction is related to the strong, reciprocal anatomical connectivity between the two areas, as shown in anatomical studies in animals.

The above experiments have confirmed that the amygdala and the hippocampus govern distinct but linked memory systems that are both crucial for emotional learning. What remains to be answered is *how* these systems are coupled. The following sections explore the various functionalities of the hippocampus and the amygdala, as well as the current knowledge and hypotheses of how they interact.

1-2-1 The role of the amygdala

The amygdala is responsible for assigning emotional value to highly analyzed stimuli signals from the sensory cortex [2]. The emotional responses associated with the amygdala include happiness, fear, anxiety, anger, and aggression. The amygdala also plays a crucial role in the two stages of memory, which are encoding and consolidation.

In the encoding phase, the stimulus is first perceived and attended to. Studies have shown that emotion has an impact on attention by drawing it in and changing how quickly emotional inputs are processed upon varying levels of attention [28]. As the amygdala is strongly connected to the sensory cortices, its signals modulate the mechanisms of 'emotional attention' [29]. Studies have shown that the amygdala responds to emotional stimuli rapidly and before awareness, allowing it to enhance perception and thus the encoding of emotional events [30]. For instance, it has been shown that people have the predisposition to direct attention toward potentially threatening stimuli [31, 32].

The second memory phase is consolidation, during which the encoded memories are fragile and need time to stabilize [28]. Once memories are consolidated, their retrieval is governed by the hippocampus. The amygdala has been found to modulate consolidation by regulating brain regions such as the hippocampus and the striatum [33]. This study has suggested that the process of memory consolidation is slow to allow the emotional reaction and its physiological response to influence the memory of the event. Moreover, studies have shown that eliciting a stress response right after the encoding of a stimulus using pharmacological [34] and pain [35] stimulation also enhances the memory associated with this stimulus.

Furthermore, several brain imaging studies have shown the involvement of the amygdala in long-term memory for emotional events. For instance, it has been shown that the level of physiological response to emotional videos is significantly correlated with the number of videos recalled a few weeks later [36]. Several animal studies have also demonstrated the impact of stress on long-term memory. Experiments with rats have shown that administering an electric shock or injecting with a stress hormone before the task impairs the animals' retrieval of long-term spatial memory governed by the hippocampus [37].

1-2-2 The role of the hippocampus

The primary role of the hippocampus is forming, organizing, and retrieving declarative memories [3]. This type of memory is also called explicit memory and it pertains to the recollection of everyday facts and events. Most of the research about the Amy-Hip interaction in emotional learning and memory has focused on the influence the amygdala has on the hippocampus. There is however evidence that declarative memory governed by the hippocampus has an impact on the amygdala as well.

In classic fear conditioning experiments, subjects acquire a fear response to a neutral stimulus that was earlier presented to them together with an aversive stimulus. Yet in everyday life, humans may acquire conditioning to stimuli they have not experienced before but to which they have assigned a negative emotional meaning through, for example, verbal communication with someone else [28]. This phenomenon has been studied in an fMRI experiment with a task called instructed fear [38]. The subjects were told that they would be presented with a few different slides and that the one showing a blue square would be followed by one or more mild shocks to the wrist. However, no shocks were actually administered in the study. When shown the blue square, the subjects showed a physiological response consistent with fear and increased amygdala activity. Such learning through instruction requires acquiring declarative memories about the emotional significance of stimuli, which means it is governed by the hippocampus.

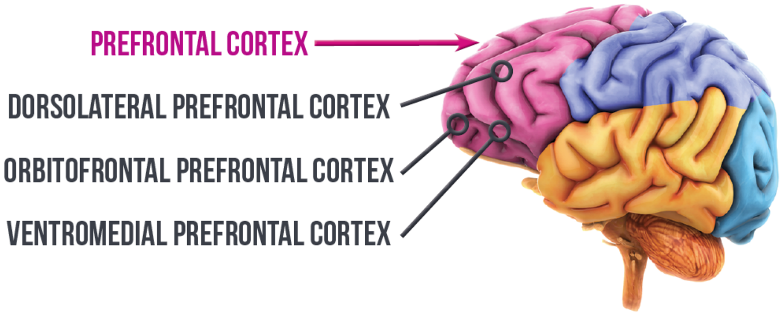

PREFRONTAL CORTEX
DORSOLATERAL PREFRONTAL CORTEX
ORBITOFRONTAL PREFRONTAL CORTEX
VENTROMEDIAL PREFRONTAL CORTEX

Figure 1-2: Regions of the Prefrontal Cortex (PFC) from [40].

1-2-3 The role of the prefrontal cortex

The cerebral cortex is the outer layer of the brain and it is where many of the higher-level cognitive processes such as language and decision-making take place. In contrast, the amygdala and the hippocampus are part of the subcortex, which is involved in the processing of more basic or involuntary functions such as emotion and attention. Several cortical regions have been found to be crucially involved in emotional and cognitive processes [2]. Firstly, the sensory cortices send their highly processed signals about the perceived stimuli to the Amy-Hip complex. They are typically not modeled directly as part of the emotional learning circuit and serve solely as a source of an exogenous input. Furthermore, the Prefrontal Cortex (PFC) is involved in a number of cognitive functions related to emotion regulation, attention, working memory, and reward processing [39]. It is thus commonly considered as a key element of the emotional learning circuit. The PFC can be subdivided into (at least) three regions depicted in Figure 1-2, namely the Dorsolateral Prefrontal Cortex (DLPFC), the Ventromedial Prefrontal Cortex (vmPFC), and the Orbitofrontal Cortex (OFC).

The PFC is crucially involved in "top-down" processing concerned with complex behavior driven by internal states and intentions. Conversely, "bottom-up" processing is responsible for simple behaviors that are "hardwired" in the brain and can be performed quickly and automatically like orienting oneself toward an unexpected noise. The "top-down" processing is illustrated well in the classic study using the Stroop task [41]. In this paradigm, the subjects are asked to read words for colors written in different colors. As the rules change, the participants have to adjust their goals and behavior. The task is especially hard in the case of conflicting stimuli like the word "red" written in blue, where goal-directed behavior has to be applied to fight the tendency to select the more automatized response (naming the font color) if it is irrelevant for the task (reading the word).

To aid the goal-oriented control, the PFC is responsible for parts of working memory concerned with the active maintenance of the goals and rules of the task [39]. As a result, the PFC plays an important role in associative learning. By keeping track of goals and performing the relevant behaviors, the PFC is capable of inhibiting associations that are no longer rewarded [42]. As the amygdala is concerned with learning from primary reinforcement, the PFC is crucial to regulate it when the reinforcement is no longer present. The OFC appears to be crucially involved in this by detecting the absence of expected rewards and driving the extinction of learning in the amygdala [2].

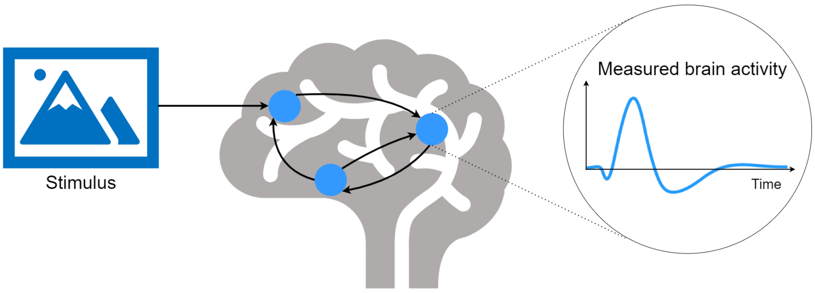


Figure 1-3: Visualization of the inputs, states, and outputs modeled by DCM.

1-3 Related work

The modeling framework utilized in this project is Dynamic Causal Modeling (DCM) [43]. The motivation for selecting this approach is explained in Section 2-1-2 and the method is introduced in detail in Section 2-2. In brief, DCM is a method for creating a dynamical model of a brain network that responds to external stimuli (e.g., sounds or images) and collectively gives rise to neural activities. These activities can be measured using various brain imaging techniques summarized in Section 2-1-1. The data used in this project was obtained using the fMRI technique. Figure 1-3 depicts a visualization of the inputs, network states, and outputs in DCM. In this project, the relevant brain regions that should constitute the network are the Amy, the Hip, and the PFC. Since the PFC is a large region, one of its subregions should be selected to focus only on the most relevant area.

Several studies have applied DCM with fMRI data to model emotional memory in different brain regions and using different experimental paradigms. A recent review includes a comprehensive list of DCM studies, divided into categories based on which brain regions were modeled [44]. Furthermore, another recent review has collected all the studies about emotional networks and synthesized their findings [45]. Based on the two reviews and additional research, four studies about Amy-Hip interactions have been found with tasks concerning emotional encoding [46], suppression of intrusive memories [47], elaboration of emotional autobiographical memories [48], and emotional associative memory retrieval [49]. Moreover, one associative memory encoding study has been found, which included the amygdala and two regions of the PFC [14].

Firstly, [46] has investigated the facilitatory effect of positive and negative pictures during memory encoding between the Amy and the Hip. The behavioral results indicated that the emotional material was better recalled than the neutral one, which is a common finding in item memory tasks. Moreover, positive images were rated on average as less arousing than their negative counterparts but they were better remembered. The DCM study revealed that the influence of the amygdala upon the hippocampus is more than ten times stronger than that of the hippocampus upon the amygdala. During the encoding of positive and negative pictures, the mutual connections between the amygdala and the hippocampus were stronger than for neutral images. Furthermore, statistical tests revealed that remembered pictures were associated with an increased connection strength between the two nodes. However, in the case of not remembered images, negative material had a bigger impact on connection strength than positive items.

Another DCM study of emotional memory encoding has studied associative learning using pairs of neutral and emotional words and pictures [14]. The network consisted of the Amy and two regions of the PFC, namely the Inferior Frontal Gyrus (IFG) and the Medial Frontal Gyrus (MFG). The study has revealed that the IFG was driving the interaction between itself and the Amy. Furthermore, bidirectional connections between the PFC regions and the Amy were found, with a higher connectivity strength from the cortical regions to the subcortical one. Furthermore, the reciprocal connection was higher during the encoding of positive and negative pairs than for neutral pairs. Additionally, the subsequent memory of the emotional stimuli was better than that of the neutral ones.

Moving on to the studies of memory retrieval, [47] has examined whether stopping the retrieval of distressing memories impairs their affective content. The authors modeled the interaction between the Amy, the Hip, the Parahippocampal Cortex (PHC) (region adjacent to Hip), and the Middle Frontal Gyrus (MidFG) (region of the PFC, not to be confused with the Medial Frontal Gyrus (MFG)). Analysis of the fMRI data revealed that suppressing distressing memories increased the activity in the MidFG and at the same time, inhibited activity in the Amy and the Hip. The results have suggested that suppressing memories not only interrupts episodic retrieval in the hippocampus and parahippocampus but also inhibits the emotional response in the amygdala. Furthermore, all these reactions are mediated by the MidFG. When incorporating the behavioral results into the analysis, it was revealed that this pattern was more pronounced for subjects who were subjectively more successful at suppressing their memories and reported reduced distress from these thoughts.

Another memory retrieval study has modeled the connectivity between the amygdala, the hippocampus, and the vmPFC during the elaboration of emotional autobiographical memories [48]. When retrieving emotions, they are first constructed (searched for and accessed) and then elaborated (memory details are integrated into a vivid construct) giving a subjective sense of reliving them. The results have suggested that during the elaboration phase, the activity in the vmPFC increases proportionally to subjective ratings of the emotional intensity of the memories. Moreover, the vmPFC has been found to serve a central role among the modeled nodes by driving the activity in the hippocampus and the amygdala. Contrary to common findings about memory encoding, the study has found that during memory elaboration, the connections to the hippocampus are much stronger than those to the amygdala. In fact, in the case of positive memories, the connection strength from the vmPFC to the amygdala was so low that it failed the statistical significance test. On the other hand, while retrieving highly emotionally intense memories (both positive and negative), the effective connectivity from the vmPFC to the hippocampus increased. These results have suggested that the emotional aspects of autobiographical memories are mostly processed and represented in the vmPFC, which then sends this information to the rest of the emotional memory network.

Finally, an associative memory retrieval study has modeled the interactions between the amygdala, the hippocampus, the OFC, and the fusiform gyrus (included to serve as a visual input area for the network) [49]. The goal was to examine the neural activity during the retrieval of emotional associative memories of previously encoded pairs of neutral and emotional objects. Most notably, the retrieval of emotional stimuli, as compared to the neutral ones, increased the strength of the connection from the fusiform gyrus to the hippocampus, and bidirectionally between the hippocampus and the amygdala. Furthermore, the influence of inputs on the OFC was also increased, enhancing the activity in both the amygdala and the hippocampus. The results have suggested that the OFC appears to modulate the activity between the amygdala and the hippocampus. This cortical region is functionally related to the vmPFC, which also appeared to serve this role in the study in [48] discussed above.

In conclusion, five related DCM studies have been identified and may serve as a reference when developing and analyzing a model in this project. One of the reasons for analyzing the related studies was to decide which of the PFC subregions to include in the model in this project. Five different PFC subregions have been modeled in the related studies but two of them, namely the vmPFC and the OFC, are closely functionally related and are sometimes treated as the same region. While the two brain areas serve similar roles, the OFC is an anatomically defined region, and the vmPFC is only functionally defined, making it harder to identify in the brain. Furthermore, the OFC has been found to be crucially involved in associative learning (see Section 1-2-3), which is the type of learning that is studied in this project. Consequently, the OFC is chosen as the third region of the modeled network, together with the amygdala and the hippocampus.

1-4 Problem statement

The goal of this thesis is to develop and analyze a dynamical model of neuronal activity involved in encoding emotional memories. Based on the literature, the most crucial brain regions responsible for this cognitive process are the Amygdala (Amy), the Hippocampus (Hip), and the Orbitofrontal Cortex (OFC). These regions shall be modeled as a network that captures how the nodes are connected, how they interact, and how they collectively give rise to the measured signal. The model shall be derived based on fMRI data from the experiment performed by Zhu et al. at Donders Institute for Brain, Cognition, and Behaviour [1] (see Section 3-1-1 for details). The experiment has studied emotional associative memory, which is the memory of relationships between concepts. In the study, participants were asked to memorize pairs of images with varying emotional values: both images were emotionally neutral (group Neutral-Neutral (NN)), both images had a negative emotional value (group Emotional-Emotional (EE)) or one image was neutral and one was negative (group Neutral-Emotional (NE)). In order to study how these different combinations of emotional stimuli affect neuronal activity, a separate model shall be derived for each group.

The three data-driven models shall be developed using the Dynamic Causal Modeling (DCM) framework, which is considered to be the most suitable for this application (see Section 2-1 for discussion). In DCM, fMRI data is modeled using a bilinear differential equation that describes the activity in the chosen brain regions (here, Amy, Hip, and OFC). There are several variations of DCM (see Section 2-4) that shall be utilized in order to identify the best-suited approach. Furthermore, after the computational models have been developed, they shall be analyzed and compared to the behavioral results found by [1]. The aim is to identify properties of the models that are aligned with the behavioral findings and that explain why and how emotional memory differs across the three groups (NN/NE/EE).

Based on these goals, the following research questions have been formulated:

1. What are the underlying neuronal dynamics in the Amy-Hip-OFC network during the encoding of emotional associative memory?

2. What are the mathematical foundations of DCM and how to apply the method in practice?
3. Which formulation of DCM reveals more detailed and complex dynamical properties of the problem at hand?
4. What are the differences between DCM models estimated for the three experimental groups: NN, NE, and EE?
5. Can analyzing the identified models explain when and why emotion either enhances or impairs memory encoding?

1-5 Thesis outline

The aim of this report is to present the work done to research the problem at hand, understand the selected modeling framework, develop a computational model using experimental data, and analyze the results. So far, Chapter 1 has motivated the need for modeling this cognitive process, introduced concepts from psychology and neuroscience, summarized related work, and presented the problem statement. Subsequently, Chapter 2 begins by explaining how brain activity is measured and what methods are available for data-driven modeling of neuronal dynamics. The most suitable approach is then selected and introduced in detail, including its mathematical foundations, practical considerations, limitations, and variations of the method. Furthermore, Chapter 3 presents the experimental paradigm and the approach for building a model based on data from the experiment. It also explains and motivates the various design decisions involved in setting up the model. Then, Chapter 4 presents and analyzes the modeling results, including parameter values, stability properties, and other dynamical properties. Finally, Chapter 5 concludes the project by discussing what the various findings tell us about the neural underpinnings of emotional memory and how they compare to the behavioral results of the experiment. Moreover, the chapter proposes several directions for future research aimed at refining the results of the thesis and obtaining new insights.

Chapter 2

Theoretical framework

This chapter introduces various data-driven models of brain activity and describes in detail the selected framework. First, Section 2-1 explains how neural activity is measured and provides an overview of generative brain modeling approaches. After considering the available options, the Dynamic Causal Modeling (DCM) framework is chosen and this decision is motivated. Furthermore, Section 2-2 provides a detailed description of DCM, including the mathematical foundations and a guide for applying the technique in practice. Then, Section 2-3 discusses the limitations and validity of the selected framework. Finally, Section 2-4 introduces several variations of DCM that may be useful in the project.

2-1 Data-driven models of brain activity

The mechanism of neuron firing was identified and consolidated into the famous Hodgkin-Huxley model [50] in the 1950s. The remaining question was then how to model the collective behavior of neurons. Researchers have taken inspiration from other phenomena caused by joint macro-scale behavior rather than individual units, namely magnetism and fluid dynamics [51]. These fields have had long-established frameworks that unify the behavior of units using the so-called "mean field" approach. From here, a multitude of data-driven brain modeling frameworks have been developed for use in various applications.

2-1-1 Brain imaging techniques

Brain imaging (or neuroimaging) techniques differ in how they measure and visualize information about the brain. They are classified into two types: structural and functional brain imaging. Structural imaging techniques are used to study the anatomical properties of the brain, which helps diagnose structural abnormalities in the brain such as tumors and lesions. Methods in this category include structural Magnetic Resonance Imaging (sMRI) and Computed Tomography (CT).

Functional imaging techniques, on the other hand, are used to study brain activity and function. These methods measure changes in the brain's activity level, blood flow, or metabolism to understand how different regions of the brain are involved in specific tasks or behaviors. Examples of functional imaging techniques include Positron Emission Tomography (PET), functional Magnetic Resonance Imaging (fMRI), Electroencephalography (EEG), and Magnetoencephalography (MEG). The purpose of this project is to model neural activity involved in emotional learning so applicable imaging techniques belong to the functional imaging category. Among the different functional neuroimaging techniques, fMRI was deemed the most suitable for this project and was used to obtain the data for developing a model. The following subsection explains the working principles of fMRI and why this technique has been selected.

Functional Magnetic Resonance Imaging

Functional Magnetic Resonance Imaging (fMRI) measures changes in blood flow and oxygenation to identify areas of the brain that are active during a specific task like reading or while at rest. For instance, Figure 2-1 depicts visual representations of brain activity measured with fMRI during reading and writing tasks [52]. The technique is safe and non-invasive, making it one of the most popular neuroimaging methods to study cognitive processes such as attention and memory, as well as the neural basis of neurological and psychiatric disorders such as Alzheimer's disease and depression.

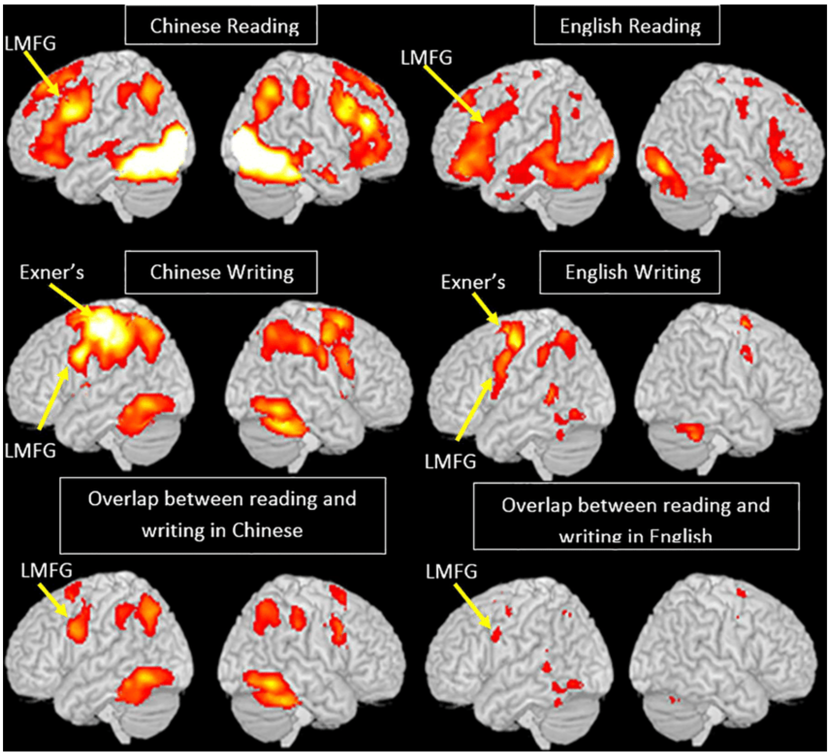


Figure 2-1: Brain activation measured with fMRI during reading and writing in English and Chinese from [52]. LMFG and Exner's are brain regions that were evidently the most active during these tasks.

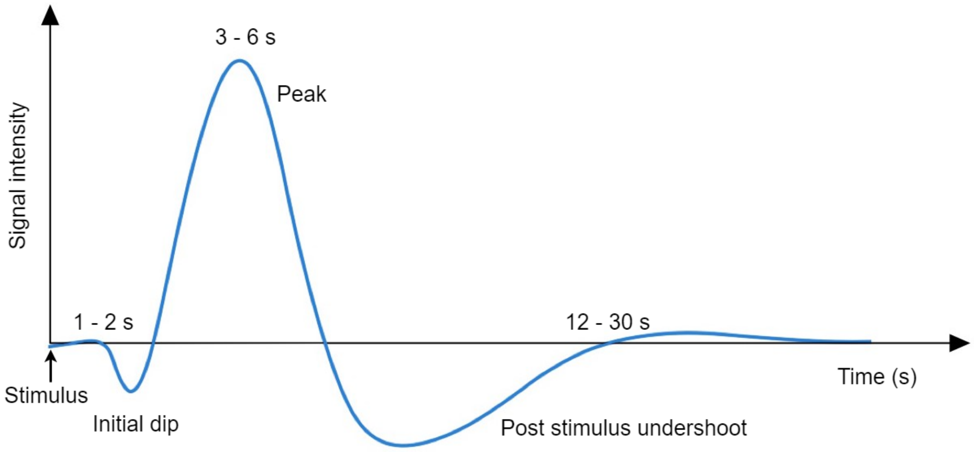


Figure 2-2: Typical shape of the Hemodynamic Response Function based on [53].

The fMRI scanner is a cylindrical tube with a powerful electromagnet that affects the magnetic nuclei of atoms in the brain. This neuroimaging technique acquires a readout of brain activity by measuring the consequences of blood flow and oxygenation on the magnetic field. The level of blood oxygenation dynamically changes in response to neuronal activity, affecting the magnetic field. Changes in the deoxyhemoglobin (hemoglobin without oxygen) content are correlated with the so-called Blood Oxygenation Level-Dependent (BOLD) signal, which is an indirect measure of brain activity.

The BOLD response is characterized by the Hemodynamic Response Function (HRF) whose typical shape is depicted in Figure 2-2. It should be noted that the HRF varies across different brain regions and can also be affected by various factors, such as age, gender, and health status. The BOLD response is based on the fact that when neurons in the brain become more active, they require more oxygen to function. This initially causes the signal intensity to drop. Then, to meet the increased demand for oxygen, blood flow to the active brain regions increases, delivering more oxygenated blood to the area. The signal intensity peaks after 3-6 seconds, after which it slowly drops back down to baseline within 12-30 seconds [53]. As a result, the fMRI measurements occur several seconds after the cognitive activity itself, which is why the imaging technique is said to have a low temporal resolution. It does however have a very good spatial resolution typically down to millimeters, which is significantly greater than other techniques such as EEG. The high spatial resolution of fMRI is one of the crucial reasons why this technique has been chosen for the experiment analyzed in this project. The goal of the study is to identify how the Amygdala (Amy), the Hippocampus (Hip), and the Orbitofrontal Cortex (OFC) interact and respond to experimental stimulation. To study the activity of these specific brain regions with high precision, high spatial resolution is required. What is more, the Amy and the Hip are located deep in the brain, making them especially hard to reach for techniques other than fMRI.

2-1-2 Modeling frameworks

In the words of one of the top scientists in the field, *"As yet, there is no broadly accepted mathematical theory for the collective activity of neuronal populations."* [51]. All the existing models come with different strengths and weaknesses and it is up to the user to pick the most suitable one. In this project, the goal is to develop a dynamical model that represents the time evolution of state variables in a network. The network consists of three brain regions, i.e., the Amy, the Hip, and the OFC (see Section 1-2 and Section 1-3). The three brain nodes interact and collectively mediate the process of learning upon emotional arousal. It should be noted that the term "dynamical model" has different meanings depending on the scientific community. In the field of control engineering, a dynamical system is one that uses differential or difference equations to describe how states and their rates of change evolve over time. Dynamical models are also generative in nature. This means that not only do they predict a response but they also generate new data that is similar to the training data [54]. They thus require leveraging prior knowledge about the process to build a representation that encompasses the underlying neuronal dynamics. After fitting the model to data, the representation can answer the scientific questions and on top of that, reveal novel insights that point to the generation of new questions.

A recent review by [54] provides an overview of generative models in neuroscience. The authors have distinguished three model categories, based on the level of detail that the model captures, the amount of necessary prior knowledge, and the size of the training dataset. The three categories and some example modeling frameworks are depicted in a Venn diagram in Figure 2-3. On one end of the spectrum are biophysical models whose aim is to capture realistic biological assumptions and processes. They require a high level of prior knowledge and can become very complex. However, when constructed well, they can provide detailed insights about the studied phenomenon. On the other end of the spectrum are agnostic computational models. These require few biological assumptions and can be powerful data-driven tools when the training dataset is large. Their drawback is that they are so-called "black-box" models, meaning that their mathematical description is difficult to interpret and is physically uninformative. In between the two extremes are phenomenological models, which require some priors on system dynamics but not detailed biological descriptions. They use tools from statistical physics and dynamical systems theory to estimate parameters that describe the overall signals observed in the data. They typically encompass neuronal dynamics into a state space model, which reveals information about collective system dynamics such as stability properties.

A natural choice for this project is to select a phenomenological model. Approaches in this category are rooted in dynamical systems theory, do not require large amounts of data (the dataset for this project includes 70 subjects), and are not based on detailed biological descriptions (which are not the main interest of this thesis). They reduce the dimensionality of the problem and leverage statistical methods to model the same collective behavior. These models usually treat neuron populations as distinct nodes in a network that jointly mediate a certain process. The influential reviews by [55] and [56] have paved the way for using neuroimaging data to model the brain as a complex network. Networks are studied using a branch of mathematics called graph theory and are defined as a set of nodes (vertices) linked by connections (edges). Most real-life networks, including the brain, are complex networks with non-trivial topological features. These properties include tightly-knit groups of nodes

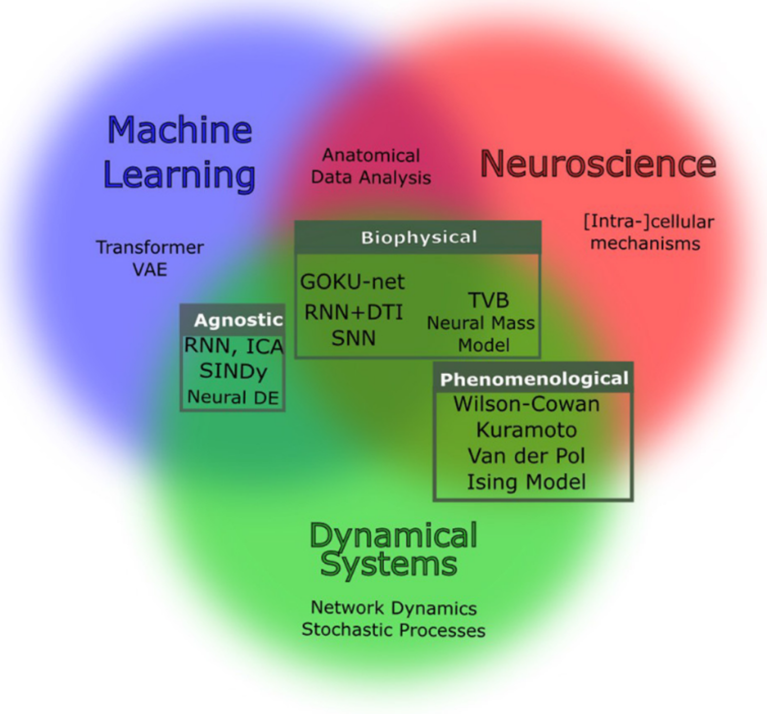


Figure 2-3: Diagram of generative models of neuronal dynamics from [54].

with dense connections (high clustering), the tendency of nodes to connect with nodes of similar properties (assortativity), and small-worldness, which is when most nodes are not neighbors of one another but they can be reached from every other node through a short path.

In large-scale brain networks, nodes typically represent brain regions, whereas links reflect different types of brain connectivity depending on the data and the application. One of the most common ways to describe brain connectivity is in terms of either functional segregation or integration [57]. Functional segregation, also called structural connectivity, refers to the anatomical segregation of functionally specialized brain areas. In turn, functional integration refers to the interaction and coupling of distinct brain regions. The two main types of functional integration are functional connectivity, which refers to the symmetrical statistical dependency between neuronal systems, and effective connectivity, which refers to the directed or causative ties between system elements. Hence, there is a fundamental difference between these two notions in that functional connectivity considers dependencies between observed signals, whereas effective connectivity is the influence one neuronal system exerts over another.

As this thesis is concerned with modeling brain activity and not its anatomical properties, applicable methods are those that model functional integration. Furthermore, this project aims to apply tools from systems and control engineering, which requires creating a dynamical model. This leads to the conclusion that the modeled network shall express effective connectivity. Among the commonly used phenomenological models are the Kuramoto [58] and the

Van Der Pol [59] oscillators, as well as DCM [43] and Granger Causality Mapping (GCM) [60]. The oscillator models have not been selected for this project because they do not appear to be as commonly used for modeling learning as some other methods. Furthermore, GCM is a method for modeling causal relationships between measured signals, not nodes, which is not the intention of this thesis. Finally, DCM is the most popular method for modeling effective connectivity between brain nodes and has been chosen as the most suitable approach for the project. The method's popularity has led to various valuable studies about its validity and limitations [61, 62, 63, 64], as well as multiple extensions to the method including stochasticity [65], a model of excitatory and inhibitory populations [66], and a nonlinear formulation [67]. There have also been numerous studies that employed DCM to model emotional learning (see Section 1-3), giving plenty of resources and comparative results for this project.

2-2 Fundamentals of Dynamic Causal Modeling

Dynamic Causal Modeling (DCM) [43] is a framework for system identification of nonlinear input-state-output systems. It is a form of systems neuroscience that uses statistical techniques to identify the underlying mechanisms that drive brain activity. DCM involves building a mathematical model of the brain that can be used to simulate how different brain regions interact with one another and how these interactions give rise to various behaviors and cognitive processes.

The approach uses Bayesian inference to estimate the hemodynamic parameters and the coupling strengths between regions of interest in the brain. Inputs are treated as known, deterministic signals, whereas the outputs are the electromagnetic or hemodynamic responses of the brain measured with techniques such as fMRI or EEG. Since the original paper was published, numerous variations of the method have been proposed. The purpose of this section is to introduce the classical DCM framework described in [43].

2-2-1 Dynamical model

The dynamic model is based on the Balloon model [68] and the Windkessel model [69]. It consists of m inputs and l outputs, with one output per region. The m inputs are the designed, experimental signals. Typically, the extrinsic effects of inputs are limited to a single input region. Furthermore, each of the l regions generates a measured output that matches the BOLD signal measured with fMRI. Every region is modeled with five state variables that involve the hemodynamic and the neuronal parameters. The former are the four variables of the hemodynamic model presented in [70], which are necessary to model the observed BOLD response but are not influenced by the states of other regions and are thus not the main interest in DCM. What is the primary concern in DCM is estimating the neuronal variables, which are equivalent to coupling parameters, with one per region.

Neuronal state equations

The underlying neuronal activity modeled with neuronal states $z = [z_1, \ldots, z_l]^\top$ evolves over time according to

$$\dot{z} = F(z, u, \alpha), \tag{2-1}$$

where F is some nonlinear function, u are the extrinsic inputs, and α are the parameters of the model. The exact function F is unknown but can be approximated using partial derivatives resulting in the bilinear form

$$\begin{aligned} \dot{z} &\approx Az + \sum u_j B^j z + Cu \\ &= \left(A + \sum u_j B^j\right) z + Cu \\ A &= \frac{\partial F}{\partial z} = \frac{\partial \dot{z}}{\partial z} \\ B^j &= \frac{\partial^2 F}{\partial z \partial u_j} = \frac{\partial}{\partial u_j}\frac{\partial \dot{z}}{\partial z} \\ C &= \frac{\partial F}{\partial u}. \end{aligned} \tag{2-2}$$

Here, matrix A represents the context-independent connectivity between brain regions resulting from anatomical connections. Matrices B^j model the change in coupling caused by the j-th input, whereas matrix C represents the direct influence of inputs on neuronal dynamics. Figure 2-4 depicts a graphical representation of the meaning of each of the matrices. Altogether, these coupling matrices form the parameter $\theta^c = \{A, B^j, C\}$ that is to be estimated.

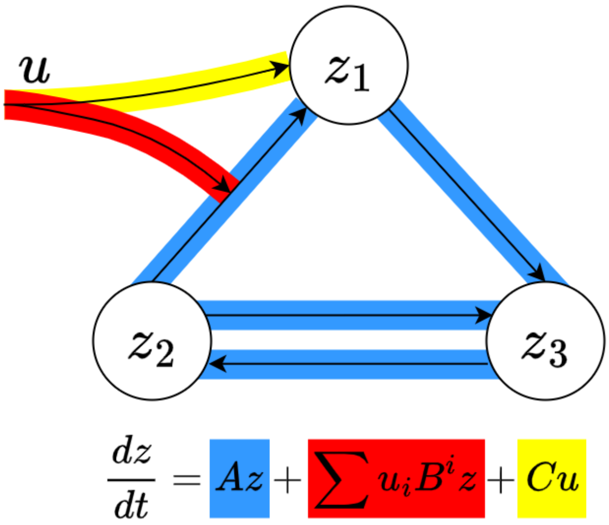


Figure 2-4: Graphical meaning of matrices in the neuronal model in DCM, with some arbitrarily drawn connections. Blue arrows correspond to influences that nodes exert on each other (matrix A), red arrows represent the influence the extrinsic input has on node connections (matrices B^j), and yellow arrows represent the direct impact the extrinsic input exerts on the neuronal states of network nodes (matrix C).

Hemodynamic state equations

The hemodynamic state variables describe how neuronal activity influences hemodynamic responses. Their dynamical equations form the Balloon-Windkessel model proposed by [68, 69, 71]. These variables correspond to the vasodilatory signal s_i, inflow f_i, normalized blood volume v_i, and normalized deoxyhemoglobin content q_i. Their dynamics are governed by

$$\begin{aligned} \dot{s}_i &= z_i - \kappa_i s_i - \gamma_i \left(f_i - 1\right) \\ \dot{f}_i &= s_i \\ \tau_i \dot{v}_i &= f_i - v_i^{1/\alpha} \\ \tau_i \dot{q}_i &= f_i E\left(f_i, \rho_i\right)/\rho_i - v_i^{1/\alpha} q_i / v_i, \end{aligned} \tag{2-3}$$

where κ_i is the rate of signal decay, γ_i is the rate of flow-dependent elimination, τ_i is the hemodynamic transit time, flow $E(f_i, \rho_i) = 1 - (1-\rho_i)^{1/f_i}$ is a function of the resting oxygen extraction fraction ρ_i, and α_i is the Grubb's exponent [72]. These comprise the biophysical parameters $\theta^h = \{\kappa, \gamma, \tau, \rho, \alpha\}$ that are estimated by the system identification algorithm.

Full model

The states of the full dynamical model are all the neuronal and hemodynamic states combined $x = \{z, s, f, v, q\}$, resulting in

$$\begin{aligned} \dot{x} &= f(x, u, \theta) \\ y &= \lambda(x), \end{aligned} \tag{2-4}$$

with parameters $\theta = \{\theta^c, \theta^h\}$. To evaluate the output y, the state equation has to be integrated and passed through some (unknown) nonlinearity λ. This is equivalent to convolving the inputs with the system's Volterra kernels, according to

$$\begin{aligned} h_i(u, \theta) &= \sum_k \int_0^t \dots \int_0^t \kappa_i^k \left(\sigma_1, \dots, \sigma_k\right) u\left(t - \sigma_l\right), \\ &\qquad \dots, u\left(t - \sigma_k\right) d\sigma_1, \dots, d\sigma_k \\ \kappa_i^k \left(\sigma_1, \dots, \sigma_k\right) &= \frac{\partial^k y_i(t)}{\partial u\left(t - \sigma_1\right), \dots, \partial u\left(t - \sigma_k\right)}, \end{aligned} \tag{2-5}$$

where $h_i(u, \theta)$ is the estimated BOLD response in region i and κ_i^k is the k-th order kernel in region i. The Volterra convolution can be evaluated either numerically or analytically through bilinear approximations explained in the appendix of [70]. This output response approximation is needed in the parameter estimation scheme discussed below. In principle, $h(u, \theta)$ has to be reevaluated at every time step but fortunately, input signals tend to change infrequently so the gradients can be reused until a change in the input occurs.

For estimation, the observation model is augmented with a term expressing error ε and confounding effects X with unknown coefficient β, resulting in

$$y = h(u, \theta) + X\beta + \varepsilon. \tag{2-6}$$

The error ε is considered to have a Gaussian distribution with a zero mean. Furthermore, designing the confounding effect signal is up to the user. The authors of the original DCM paper defined X using a low-order discrete cosine, which models low-frequency response drifts.

2-2-2 System identification

The parameters and hidden states of the model are estimated using a Bayesian approach based on the Free Energy Principle (FEP). The FEP is one of the most prominent theories of how the brain learns and adjusts to the environment [73]. The theory models how biological systems, including the brain, try to minimize their surprise (free energy), which is the difference between their internal models and the sensory input from the environment. The FEP was created by Karl Friston, who is also the creator of DCM in general, but the theory has found applications in various other fields. For instance, in the field of Artificial Intelligence (AI), FEP has been extensively utilized for developing adaptive learning algorithms that mirror the principles observed in biological organisms.

The system identification algorithm used in DCM is called Expectation Maximization (EM). It is an iterative scheme that updates prior distributions to obtain posterior distributions that minimize the free energy according to FEP. A prior distribution includes the expectation η_θ and the covariance denoted by C_θ. Priors of hemodynamic parameters θ^h are established based on empirical data, whereas priors of neural coupling parameters θ^c are designed to enforce that the parameters remain stable. The iterative scheme is based on a linear approximation of the observation model in Equation 2-6. The model is approximated by expanding the equation about a working estimate of the conditional mean $\eta_{\theta|y}$, resulting in

$$\begin{aligned} y &\approx h\left(u, \eta_{\theta|y}\right) + J(\theta - \eta_{\theta|y}) + X\beta + \varepsilon \\ &= h\left(u, \eta_{\theta|y}\right) + \begin{bmatrix} J & X \end{bmatrix} \begin{bmatrix} \theta - \eta_{\theta|y} \\ \beta \end{bmatrix} + \varepsilon \\ J &= \frac{\partial h\left(u, \eta_{\theta|y}\right)}{\partial \theta}. \end{aligned} \tag{2-7}$$

This function depends on the difference between true system parameters θ and current conditional expectations of them $\eta_{\theta|y}$, as well as the estimated confounding effect $X\beta$ and error ε. Subsequently, the linear approximation of network dynamics enters the EM scheme. This iterative process calculates the posterior distributions by repeating the so-called "E-step" and then, the "M-step" until the convergence criterion is met.

The goal of the "E-step" is to update the conditional parameter expectation $\eta_{\theta|y}$ and covariance $C_{\theta|y}$. The update equation for $p(\theta|y)$ is derived using the Bayes rule

$$p(\theta|y) \propto p(y|\theta)p(\theta). \tag{2-8}$$

Furthermore, based on Equation 2-7, the error is given by

$$\varepsilon \approx y - h\left(u, \eta_{\theta|y}\right) - \begin{bmatrix} J & X \end{bmatrix} \begin{bmatrix} \theta - \eta_{\theta|y} \\ \beta \end{bmatrix}. \tag{2-9}$$

Under the Gaussian assumption, the error and prior distributions are given by

$$\begin{aligned} p(y \mid \theta) \propto \exp & \left\{ -\frac{1}{2} \left(y - h\left(u, \eta_{\theta|y}\right) - \begin{bmatrix} J & X \end{bmatrix} \begin{bmatrix} \theta - \eta_{\theta|y} \\ \beta \end{bmatrix}) \right)^T \right. \\ & \left. \times C_\varepsilon^{-1} \left(y - h\left(u, \eta_{\theta|y}\right) - \begin{bmatrix} J & X \end{bmatrix} \begin{bmatrix} \theta - \eta_{\theta|y} \\ \beta \end{bmatrix} \right) \right\}, \\ p(\theta) \propto \exp & \left\{ -\frac{1}{2} \left(\theta - \eta_\theta\right)^T C_\theta^{-1} \left(\theta - \eta_\theta\right) \right\}. \end{aligned} \tag{2-10}$$

Now, plugging the distributions in Equation 2-10 into the Bayes rule given by Equation 2-8, results in the following equation for the posterior distributions of model parameters

$$\begin{aligned} p(\theta \mid y) \propto \exp & \left\{ -\frac{1}{2} \left(\theta - \eta_{\theta|y}\right)^T C_{\theta|y}^{-1} \left(\theta - \eta_{\theta|y}\right) \right\} \\ \eta_{\theta|y} & \leftarrow \eta_{\theta|y} + \Delta\eta_{\theta|y} \\ \begin{bmatrix} \Delta\eta_{\theta|y} \\ \eta_{\beta|y} \end{bmatrix} & = C_{\theta|y} \left(\bar{J}^\top \bar{C}_\varepsilon^{-1} \bar{y} \right) \\ \bar{J} & = \begin{bmatrix} J & X \\ 1 & 0 \end{bmatrix} \\ \bar{y} & = \begin{bmatrix} y - h\left(\eta_{\theta|y}\right) \\ \eta_\theta - \eta_{\theta|y} \end{bmatrix} \\ \bar{C}_\varepsilon & = \begin{bmatrix} \sum \lambda_i Q_i & 0 \\ 0 & C_\theta \end{bmatrix} \\ C_{\theta|y} & = \left(\bar{J}^\top \bar{C}_\varepsilon^{-1} \bar{J} \right)^{-1}. \end{aligned} \tag{2-11}$$

The above update equation simply rearranges the terms from Equation 2-10, grouping them into several augmented matrices. The equation for the augmented error covariance $\bar{C}_\varepsilon$ includes one new term, which models error covariance. It parameterizes it with error hyperparameters λ_i, which scale matrices Q_i that define the contribution of error covariance components in every i-th region. Calculating and optimizing the error hyperparameters λ_i is the goal of the "M-step".

To find the optimum error hyperparameters, the "M-step" maximizes the log-likelihood $\ln p(y|\lambda)$ using the free energy given by

$$F(\lambda) = \ln p(y \mid \lambda) - D(q(\theta) \| p(\theta \mid y, \lambda)), \tag{2-12}$$

where D denotes divergence, $q(\theta)$ is the approximate posterior distribution, and $p(\theta \mid y, \lambda)$ is the real posterior. Since divergence is always positive, F creates a lower bound for the log-likelihood

$$F(\lambda) \leq \ln p(y \mid \lambda). \tag{2-13}$$

The goal is to maximize the free energy, thus minimizing the divergence, rendering $q(\theta)$ a suitable approximation of the actual distribution. F is optimized by updating the hyperparameters λ under a Fisher-scoring scheme. This scheme quantifies how much information I about the data the estimate carries by assuming that it's proportional to the inverse of the variance. The Fisher scoring scheme updates the hyperparameters such that the estimation carries maximum information. Following this iterative scheme, the hyperparameter update equation is

$$\begin{aligned} \lambda &\leftarrow \lambda - \left\langle \frac{\partial^2 F}{\partial \lambda^2} \right\rangle^{-1} \frac{\partial F}{\partial \lambda} \\ \frac{\partial F}{\partial \lambda_j} &= -\frac{1}{2} \operatorname{tr}\{PQ_i\} + \frac{1}{2} \bar{y}^T P^T Q_i P \bar{y} \\ \left\langle \frac{\partial^2 F}{\partial \lambda_{jk}^2} \right\rangle &= -\frac{1}{2} \operatorname{tr}\{PQ_i PQ_j\} \\ P &= \bar{C}_\varepsilon^{-1} - \bar{C}_\varepsilon^{-1} \bar{J} C_{\theta|y} \bar{J}^T \bar{C}_\varepsilon^{-1}. \end{aligned} \tag{2-14}$$

The "E-step" and the "M-step" are repeated consecutively until a convergence criterion is met, typically, until the sum of the squared change in $\eta_{\theta|y}$ falls below 10^{-6}.

Prior distributions

This section begins with a detailed explanation of how the neuronal dynamics priors are designed, followed by a description of how the priors for hemodynamic equations have been obtained empirically. Then, the section discusses how all the priors are combined for use in the EM estimation algorithm.

First of all, priors on the coupling parameters are designed to ensure that neuronal activity has the correct dynamic properties. What this means is it should not diverge to infinity and it should decay to baseline with the correct time constant. In order to simplify the derivation of priors, matrices A and B^j are reparametrized according to

$$\begin{aligned} A \rightarrow \sigma A &= \sigma \begin{bmatrix} -1 & a_{12} & \cdots \\ a_{21} & -1 & \\ \vdots & & \ddots \end{bmatrix} \\ B^j \rightarrow \sigma B^j &= \sigma \begin{bmatrix} b_{11}^j & b_{12}^j & \cdots \\ b_{21}^j & \ddots & \\ \vdots & & \end{bmatrix}, \end{aligned} \tag{2-15}$$

which results in scalar σ and normalized, adimensional coupling matrices. This factorization assumes the same self-connection for all regions, which is a valid approximation according to the authors as intrinsic dynamics in each region do not differ significantly.

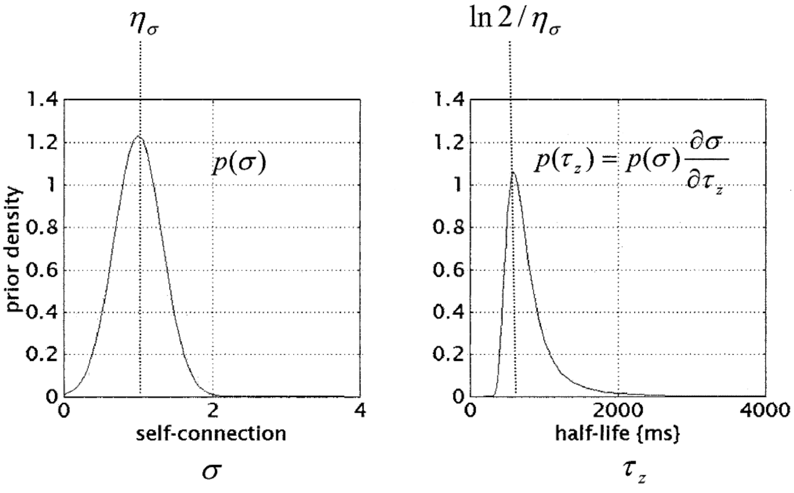


Figure 2-5: Prior probability density functions of temporal scaling factor σ from [43]. The left image shows a Gaussian distribution, which becomes a skewed distribution depicted on the right when transforming σ into the characteristic half-life.

Firstly, parameter σ corresponds to the intrinsic decay or self-inhibition, which means it controls the characteristic neuronal time constants. The authors of the paper set this parameter's expectation η_σ to 1 second, based on empirical studies. To ensure that σ is positive, its variance ν_σ is chosen to make the probability that the parameter is negative suitably small. To do so, the variance is calculated with

$$\nu_\sigma = \left(\frac{\eta_\sigma}{\phi_N^{-1}(1 - p(\sigma))} \right)^2, \tag{2-16}$$

where $p(\sigma)$ is the probability of σ being negative (10^{-3} in the paper), and ϕ_N is the cumulative normal distribution. This creates a Gaussian distribution centered around 1 with a majority of the mass falling close to the mean. To express the time constant as a function of the half-life, the mean and probability are transformed into $\tau_z(\sigma) = ln\frac{2}{\eta_\sigma}$ and $p(\tau_z) = p(\sigma)\frac{\partial\sigma}{\partial\tau_z}$, respectively. This results in a skewed distribution, which shows that time constants range from a few hundred milliseconds to several seconds. Both the Gaussian distribution and the transformed skewed distribution are depicted in Figure 2-5.

Furthermore, priors on the components of coupling matrices are set to ensure that neuronal dynamics do not diverge to infinity, i.e., they are stable. The parameters of coupling matrix C are of relatively low interest here, and they are defined as having a zero expectation and a unit variance. What is crucial for the dynamical properties of the system are the priors on the parameters of the A and B matrices. To render the system stable, these matrices should have all negative eigenvalues. To ensure this, coupling parameters a_{ij} and b_{ij}^k are assumed to be identically and independently distributed with prior expectations $\eta_a = \eta_b = 0$ and prior variances $\nu_a = \nu_b$ that make the probability of having positive eigenvalues suitably small. To achieve this goal, variances are designed according to

$$\nu_a, \nu_b = \frac{l(l-1)}{\phi_X^{-1}(1 - p(e_{\max}))}, \tag{2-17}$$

where l is the number of modeled regions, $p(e_{\max})$ is the probability of the largest eigenvalue $e_{\max}$ being positive (10^{-3} in the paper), and ϕ_χ is the cumulative $\chi^2_{l(l-1)}$ distribution. Finally, part of pre-designing the model of neuronal dynamics in DCM is deciding which connections are worth exploring and which ones should definitely be precluded. In order to remove a certain connection from the model, its variance should simply be set to zero.

Moving on to the biophysical priors, these values were calculated by the authors based on empirical data. According to the authors, the means and variances they provide in the seminal paper [43] should be sufficient for general purposes.

Finally, prior distributions of all parameters should be combined so that they can be used in the EM estimation scheme. Stacking the parameters together gives

$$\theta = \begin{bmatrix} \sigma \\ a_{ij} \\ b^k_{ij} \\ c^k_{ik} \\ \theta^h \end{bmatrix}, \tag{2-18}$$

which is distributed with mean η_θ and covariance C_θ defined as

$$\eta_\theta = \begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \\ \eta^h_\theta \end{bmatrix}, C_\theta = \begin{bmatrix} v_\sigma & & & & \\ & C_A & & & \\ & & C_B & & \\ & & & 1 & \\ & & & & C^h_\theta \end{bmatrix}, \tag{2-19}$$

where matrices C_A and C_B are diagonal matrices with values equal to corresponding variances ν_a, ν_b for connections that are allowed to vary. Similarly, matrix C^h_θ contains the biophysical parameter variances on the diagonal, in the appropriate order.

2-2-3 Practical considerations

The mathematical basis of DCM presented in the previous section has been implemented and made part of the open-source Statistical Parametric Mapping (SPM) software [74]. To aid researchers in applying the method in practice, the authors also published an article entitled "Ten simple rules for dynamic causal modeling" [75]. The purpose of this section is to summarize the parts of the paper that are relevant to this project.

First of all, researchers interested in applying DCM should establish whether the method is suitable for them, what hypothesis they want to test, and what sequence of analysis they should apply to achieve their goal. DCM was designed to model the underlying neuronal dynamics in a network that responds to various external stimuli. Due to this nature of effective connectivity models, DCM is meant for testing hypotheses about specific tasks in the presence of experimental manipulations. Conversely, the classical version of DCM is not a suitable method for datasets collected in the absence of experimental interventions (e.g., resting state or sleep, for which a variation of the method has been developed by [76]).

Furthermore, DCM can be used to obtain two different types of inference: model structure and model parameters. The former is useful when the researcher is not interested in estimating any values in the model but wants to find out some aspects of the structure of the model. For instance, the question might concern through which node the extrinsic input enters the network or whether a certain inter-region connection exists. The other form of inference is applicable when the goal is to determine brain dynamics encoded by model parameters. The identified parameters can help answer questions about which connections are stronger and which nodes are the main drivers in the network. However, even if the main goal is inference on model parameters, the first step is usually to identify (some aspect of) the model structure through Bayesian Model Selection (BMS). BMS is a well-known statistical method that computes the model evidence $p(y \mid m)$, which evaluates how well a model explains the data while having a minimal structure complexity. As less complex models tend to overfit to data less, the model evidence is also a measure of generalizability. For inference on model parameters, one typically first defines all plausible model structures and uses BMS to find the optimal one.

This leads to the next important consideration, which has to do with defining the relevant model space. When modeling networks with only a few nodes, it is feasible to compare all possible structures. However, as model complexity increases, investigating all combinations becomes increasingly hard or even computationally impossible. In such cases, it is necessary to first define a set of plausible architectures motivated by previous neuroimaging studies. Then, a comparison is done either between all individual models or between families of models. The latter is useful when models can be grouped based on some shared characteristic, for example, the extrinsic input entering the network through a given node. One might be then interested in only investigating the significance of this characteristic, for example, to see whether the presence or absence of a certain connection improves model performance. Furthermore, EEG or MEG data may also be used to infer which brain regions to include in the model. This is because when using these imaging methods in DCM, the data is always the same irrespective of the chosen set of modeled nodes. Conversely, in DCM with fMRI data, it is not possible to select the nodes through the modeling process. The reason for this is that fMRI data has to be pre-processed to include only the regions of interest for DCM.

When modeling data from multiple subjects, one must also decide between two methods. To visualize how to choose the appropriate approach, Figure 2-6 depicts a useful flowchart. The first option is Fixed-Effects (FFX), which is used when it is reasonable to assume that one model structure is optimal for all subjects. This might be the case when studying a basic physiological mechanism that is not expected to vary much across subjects, such as the mechanisms of vision. In such situations, FFX selects the optimal model by applying BMS and checks which model has the highest sum of log-evidences across subjects. When the assumption that the mechanism does not vary across subjects is not valid, the second method, namely the Random-Effects (RFX), is used. This might be necessary when studying mechanisms in subjects on a spectrum of neurological disorders such as autism. In such cases, RFX computes how likely it is that a given model generated the data of a given subject. As a result, the optimal model may vary per participant. In this project, the FFX method is used because the aim is to find one model that can be used to draw conclusions about the neural underpinnings of emotional memory.

After identifying the optimal architecture(s) across all subjects, one may want to fit the model(s) and analyze the parameters for all participants. There are again two possible choices

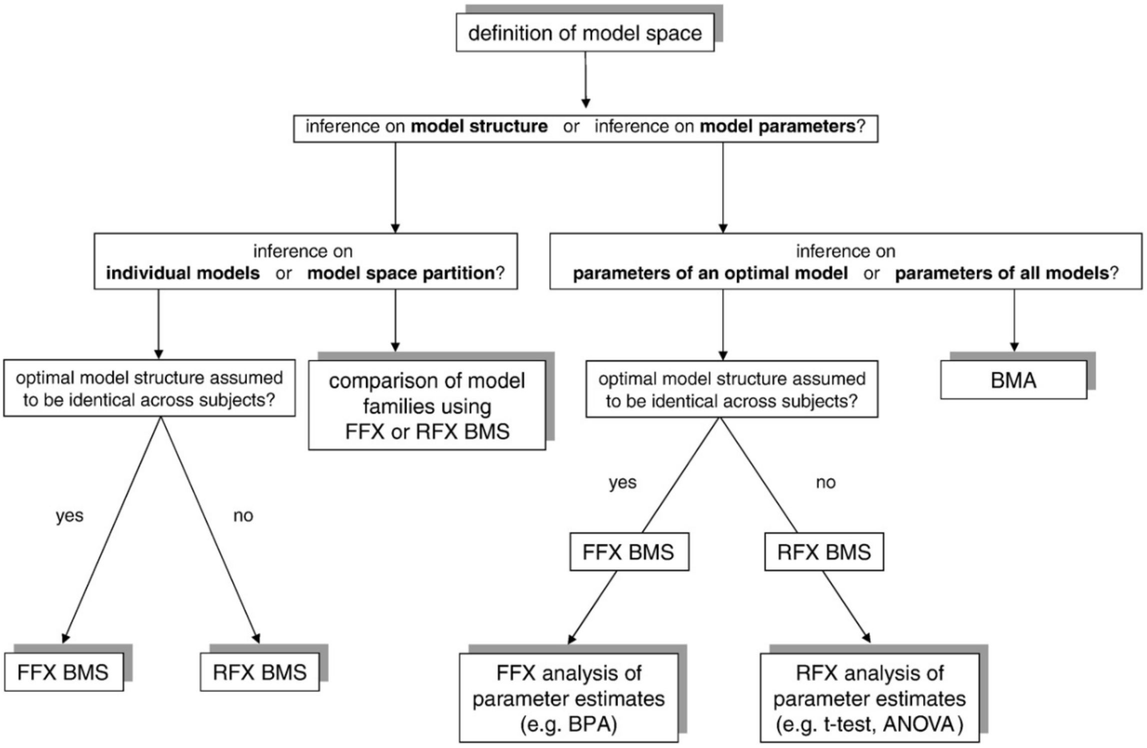

definition of model space
inference on model structure or inference on model parameters?
inference on individual models or model space partition?
inference on parameters of an optimal model or parameters of all models?
optimal model structure assumed to be identical across subjects?
comparison of model families using FFX or RFX BMS
optimal model structure assumed to be identical across subjects?
BMA
yes
no
yes
no
FFX BMS
RFX BMS
FFX BMS
RFX BMS
FFX analysis of parameter estimates (e.g. BPA)
RFX analysis of parameter estimates (e.g. t-test, ANOVA)

Figure 2-6: Typical sequence of analysis in DCM from [75].

at this stage, which are depicted in the right branch of Figure 2-6. When assuming that one model structure is optimal for all subjects (so after applying FFX), the parameters of this architecture are "averaged out" over all subjects. This is done with Bayesian Parameter Averaging (BPA), which computes the joint posterior distribution by taking the posterior of one subject and treating it as the prior for obtaining the posterior of the next subject. On the other hand, when each subject has an assigned optimal architecture (so after applying RFX), this sort of averaging is not meaningful. Instead, second-order frequentist tests (such as a t-test or Analysis of Variance (ANOVA)) may be applied to test the significance of differences in parameter estimates between subjects. As a final note, there is one more method used to infer model parameters, namely Bayesian Model Averaging (BMA), which neglects any subject- or structure-specific analysis. Instead, it computes weighted parameter averages over all the feasible model architectures. The weight is equivalent to the posterior probability of each model structure. This approach may be useful when no model architecture or family is a clear winner of the comparison. Additionally, instead of applying BMA over all possible models at once, the technique can be applied separately to families of models to compute family averages.

2-3 Limitations and validity of Dynamic Causal Modeling

DCM has gained considerable popularity within the neuroimaging community but it has also faced serious criticism, sparking heated debates over its validity. Various studies have reviewed the foundations of the approach and the practical issues associated with it. In particular, a

Nodes	Models	Seconds	Hours	Years
2	5184	2.53		
3	272 million		37	
4	1.15^{15}			17,911
5	3.98^{23}			6,168,272,683,438

Table 2-1: Estimated time of system identification for varying sizes of models from [62].

review of DCM by [62] expressed such criticism that it sparked a series of articles written in response to it [63, 64]. Nevertheless, DCM remains the most common approach for modeling the dynamics of a network of brain regions, with almost 5000 citations of the seminal paper at the time of writing this thesis. Therefore, this section aims to discuss the common critiques to identify possible pitfalls and elements of the framework that should be executed with particular care.

2-3-1 Biophysical foundations

A few years after publishing the original DCM article, some of its authors summarized the common critiques in an exhaustive review [61]. The article discusses comments pertaining to both the biophysical and statistical foundations of the method. The authors have stressed that due to the nature of fMRI, DCM analyses with this imaging technique lack detailed representations of neuronal mechanics and are thus incapable of modeling several phenomena. These unobservable effects include internal neuronal noise that may have an impact on the macro-scale, as well as the continuously changing activity-dependent efficacy of signal transmission between neurons. There are however a few details that the BOLD signal does show but the DCM approach fails to model anyway. For example, the framework ignores the role of glial cells, which support neurons and maintain their environment. It also does not account for task-related physiological influences of neuromodulators and endocrines. The extent to which these simplifications affect the model has not been studied directly. Instead, studies and critical reviews have been more concerned with the validity of the statistical methods in DCM. It is also beyond the scope of this project to examine the biophysical foundations of the method so the focus of this section is on the system identification procedure.

2-3-2 Statistical foundations

A common critique of DCM is that its number of parameters and model complexity preclude robust system identification. The critical review by [62] has provided quantitative illustrations of the problem. The authors have estimated the time necessary to identify all parameters in networks consisting of varying numbers of nodes, as shown in Table 2-1. The estimate was based on the assumption that there are around 2^{nm} possible arrangements for each of the neural state matrices (with dimensions $n \times m$) and that [77] has approximated that identifying one parameter takes around 0.0005 second.

In response to this critique, [63] have argued that the comment is based on a misconception of DCM. The critical review has calculated the time to identify the entire model space, while the aim of DCM is not to examine every single possible model to find the one "best" model. Instead, the goal is to define a much smaller subset of plausible models based on prior knowledge and to examine their common trends. This can be done through Bayesian family comparison, which is used to compare families of models with a shared characteristic, e.g., the node through which the extrinsic input enters the network. This approach is much less computationally heavy so the calculations from Table 2-1 are not applicable. What is more, in large model spaces, many models can come out as equally probable so comparing single models is in fact uninformative and thus, should be avoided.

Another critical note pertaining to parameter estimation has to do with the search strategy of the optimization algorithm. The EM scheme performs a greedy search, which means that it always selects the best direction at the current step, without considering whether the current best option is the best for the overall result. Considering that greedy search does not evaluate the entire solution space, the time estimates in Table 2-1 are again not valid as they assume an exhaustive search. The authors of [64] have argued that greedy search algorithms may end up in local optima far away from the global optimum so the solution may not be good and there is no way of knowing if it is unique. This critique is a common problem with greedy algorithms so it is not specific to DCM only. Unless the model is linear, most optimization schemes are capable of finding only the local optimum. This makes them sensitive to the starting point, which in DCM is the prior distribution of parameters. The EM scheme has also been found to have a bias toward overconfidence [61], i.e., the posterior mean is correct but the variance is too tight. Nevertheless, the algorithm has several advantages, including being easier to design and faster to compute than more exhaustive search strategies like dynamic programming.

Furthermore, model comparison relies on comparing a set of potentially plausible models against each other, meaning there is no guarantee that any of them is actually good. The authors of the critical review [62] have performed a combinatorial study of this problem as well. The experiment involved taking data from another DCM study where a winning model was identified (called model A), generating a large model set with all the combinatorial options, and checking if the previously winning model would still come out as the best one compared to the entire model space. In line with the discussion in the previous section, the authors have found that model A was outperformed by numerous other models that even included biologically implausible ones. They have concluded that the model evidence metric is not trustworthy and that the choice of the optimal model is highly sensitive to the candidate model set. In response, the authors of [63] have again pointed to the fact that model comparison in DCM is a relative measure. The approach should thus involve comparing only models that are equally plausible *a priori* rendering their prior distributions all the same. Furthermore, it should be largely focused on family comparison to avoid searching for the single best model.

Another common concern is that while the metric used for comparing models (i.e., model evidence) was designed to penalize complex models, it is not a direct measure of how well the model generalizes to data [62, 61]. To establish true generalizability, it is necessary to validate the model on a dataset that was not used for training. This is not common practice in DCM, in part because brain imaging datasets are small so all the data is used for fitting the model. When working with small datasets, cross-validation can be used instead [61]. This technique involves splitting the dataset into k blocks and fitting the model multiple times, using different subsets of the data. In each iteration, $k-1$ blocks are used to fit the model and the remaining block is used for validation. Figure 2-7 depicts an example of splitting

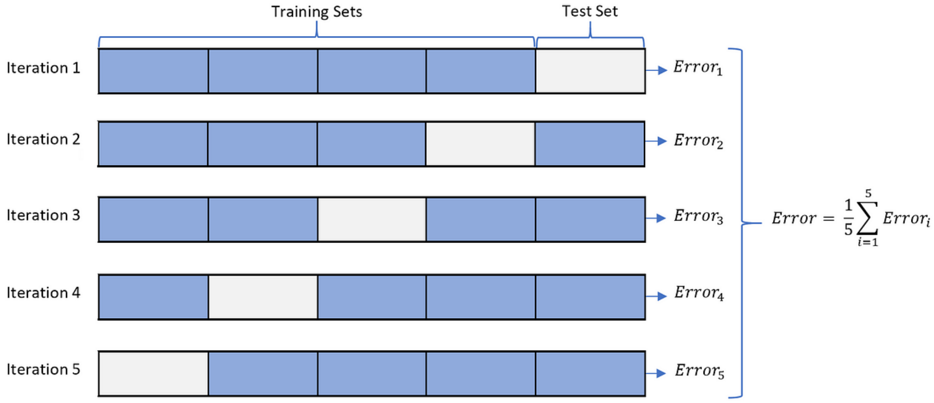


Figure 2-7: Example of splitting data for cross-validation from [78].

the dataset across several iterations. As a final step, the average error across iterations is calculated to evaluate the model.

Apart from avoiding overfitting, good models should also not underfit, meaning they should explain the training data sufficiently well. This can be quantified using the R^2 metric defined as

$$R^2 = 1 - \frac{\sum_i (y_i - f_i)^2}{\sum_i (y_i - \bar{y})^2}, \tag{2-20}$$

where y_i is the measured response at time i, $\bar{y}$ is its mean, and f_i is the predicted response at time i. The relation between the model R-squared value and the model evidence has been examined, revealing that the models selected using BMS do not always have high R^2 in all nodes [62]. A common result is that models exhibit significantly different fits among the network nodes, which can only be seen by analyzing the R-squared values for each node. A possible reason for this is that the nodes exhibited different levels and types of noise. The creators of DCM have stated that it is typical that different brain regions generate different amounts of noise [63]. In some cases, measurements may be so noisy that a given node should be excluded from the experiment. Noise in neuroimaging data is unavoidable so from the statistical point of view, the only thing that can be done is to improve how the error is modeled. In the classical DCM formulation, noise, as well as all the parameters, are assumed to have a Gaussian distribution. This is a strong assumption that was not made based on biological knowledge but rather in order to simplify computation [62]. In fact, measurement noise often has a uniform distribution, i.e., all values are equally likely, which is clearly very different from the Gaussian distribution.

To conclude, the statistical foundations of DCM have been challenged on several grounds. While some critiques seem to be excessive (like the combinatorial explosion studies) and can be salvaged by properly applying guidelines from articles like [63, 75], several limitations are unavoidable. These include the disadvantages of the local optimization algorithm, sensitivity to priors, potentially unreliable results of model comparison, and problems with underfitting and overfitting. These are serious issues but so far they have not stopped the neuroimaging community from widely applying DCM. Researchers should be simply aware that the estimated parameters are in no way "true" and the focus should be on identifying the high-level dynamical properties of the modeled network. In the words of the critical reviewers of DCM, *"the detection and analysis of networks in the human brain is still very much an open problem. None of the existing methods are as yet wholly satisfactory."* [62]. Based on the reviews of brain modeling methods discussed in Section 2-1, DCM appears to be the only dynamical systems approach for modeling effective connectivity across brain regions, suited to represent learning phenomena. Therefore, DCM remains the chosen modeling technique in this project.

2-3-3 Framework validity

To assess whether the DCM framework is sound, several studies have examined its validity. This included investigating whether the model appears to represent neural dynamics correctly (face validity), to what degree it actually models the process correctly (construct validity), and whether it can predict future outcomes (predictive validity). Face validity is the most informal and subjective measure so its findings should not be considered decisive. Construct validity aims to provide a more reliable measure of the quality of the technique. The most common way to do it is by comparing the technique to similar approaches. Finally, predictive validity is the hardest to establish, especially when data is scarce like in DCM.

Face validity has been assessed in the seminal paper [43] by modeling the system in several difficult situations. Simulations indicated that typical fMRI noise levels of $0.5 - 1.5\%$ do not compromise the accuracy and precision of system identification. This is because when the data becomes noisy, the algorithm relies more on the prior distributions. Furthermore, the authors have studied the effect of misspecification of input timing. They performed simulations using shifted input signals and found that DCMs can handle timing misspecification of up to around ± 1 second, which is acceptable for most fMRI studies with a high temporal resolution. In studies with a low temporal resolution (> 1 second), the signal may have to be realigned. DCMs are relatively insensitive to timing misalignments as these are easily absorbed into the model parameters that implicitly express timing relationships. For instance, as the input is delayed, the output response seems to evolve too early. However, this is compensated for by increasing the temporal scaling parameter σ, which accelerates the response. Finally, the model's sensitivity to incorrect priors was examined. In DCM, the normalized coupling strengths are of key interest, while the temporal scaling and the biophysical parameters serve only to complete the model. These supporting parameters are initialized using empirical values that may differ per brain region and per subject. It is thus important to establish how much misspecification of the temporal and hemodynamic priors affects the estimation of the coupling strengths. The authors have found that biophysical priors that are two standard deviations away from the correct values have a negligible effect on coupling posteriors. Moreover, varying the temporal scaling prior by up to 2 seconds from the correct value led to a poor estimation of said parameter but the coupling strengths were estimated well.

One of the most comprehensive studies of construct validity has been done by [79]. In the paper, the authors have compared DCM to another method for inferring effective connectivity from fMRI data, namely Structural Equation Modeling (SEM). This approach relies on identifying a static model of neural dynamics, meaning that it does not model the time evolution of signals like DCM does. However, SEM can be considered as a special case of DCM where the dynamics have reached an equilibrium and so the rates of change are zero.

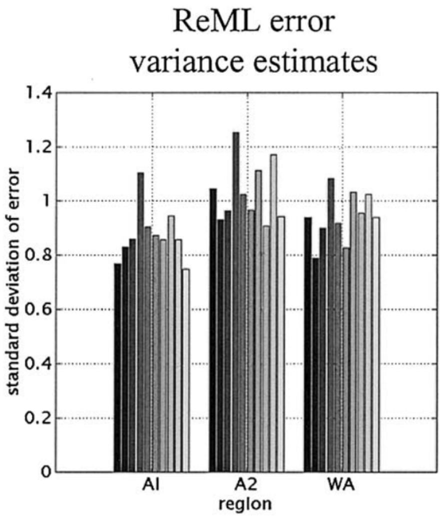


Figure 2-8: Standard deviation of the observation error across the 3 modeled regions (A1, A2, and WA) for each of the 10 datasets (shades of grey) from [43].

The authors have applied these approaches to model how attention (extrinsic input) influences the dynamics in a network of three nodes mediating the visual motion processing. The two approaches led to consistent conclusions. Therefore, the study has provided evidence for DCM's construct validity but more research would be required since SEM is not a dynamical model so the comparison is not valid when rates of change are nonzero.

Predictive validity is the hardest and most important validity notion to prove. The first study of this measure was performed in the classic DCM paper [43] by comparing models fitted on 10 independent datasets from the same experiment. The paradigm involved a passive listening task of single words used to model the connectivity in the auditory network consisting of 3 nodes. The estimated connections in the coupling matrix A were consistent among the winning models from each of the 10 training datasets. There was more discrepancy in the B matrix connections but the authors have considered them "relatively consistent" nonetheless. They have illustrated the reproducibility of the results by calculating the standard deviations of observation error across the modeled nodes for each of the 10 datasets. The estimated error variances are depicted in Figure 2-8. The figure shows that parameters were estimated with very similar degrees of error $(0.8 - 1.0\%)$.

According to the critical review by [61], the most far-reaching study of DCM's predictive validity was done by [80]. In this validation study, the authors have performed a DCM analysis using both fMRI and intracerebral EEG measurements of neural activity during epileptic seizures in rats. By comparing results using the two types of data, the study has provided supportive evidence for DCM's validity. For instance, models identified using different types of data identified the same node as the main driver in the network. However, the relevance of the study has been challenged by [62], where the authors have argued that epilepsy in rats may not be comparable to cognitive functions in humans.

2-4 Developments in Dynamic Causal Modeling

Since the influential original DCM paper, multiple variations of the method have been proposed to tackle different problems. A common concern in DCM is that the number of model parameters grows quadratically with the number of nodes, making the problem intractable. To address this issue, [81] has proposed an efficient method for modeling large networks, which rests on designing priors to constrain the number of free parameters. Another approach that tackles this problem is Regression DCM developed by [82]. This framework casts model inversion as a Bayesian linear regression problem, which accelerates system identification by several orders of magnitude. Furthermore, since different brain imaging methods offer different insights into brain activity, [83] has introduced an approach that allows DCM to generate both hemodynamic (e.g., fMRI) and electrophysiological (e.g., EEG) measurements. This enables the fusion of the two types of responses, which in turn gives access to much richer neuronal dynamics.

Undoubtedly, three of the most influential variations of DCM are Nonlinear DCM [67], Two-State DCM [66], and Stochastic DCM [65, 77, 84]. They all offer an extension to the neuronal state equations, making the model more biologically plausible. Several DCM studies of emotional memory have either used or strongly recommended using Two-State DCM [46] and Nonlinear DCM [49, 48]. While no Stochastic DCM studies of emotional memory have been found, the approach offers a more robust system identification approach, which is an attractive property. Hence, these three approaches are of particular interest in this thesis and shall all be applied to identify the one that fulfills the project's goal best.

2-4-1 Nonlinear DCM

The neuronal model in DCM is based on the assumption that the true underlying dynamics are expressed by some unknown nonlinear function F that can be approximated by the Taylor expansion,

$$F(z,u) = \frac{dz}{dt} \approx F(0,0) + \frac{\partial F}{\partial z}z + \frac{\partial F}{\partial u}u + \frac{\partial^2 F}{\partial z \partial u}zu + \frac{\partial^2 F}{\partial z^2}\frac{z^2}{2} + ..., \tag{2-21}$$

where z and u are the neuronal states and inputs, respectively. The classic version of DCM includes only first-order derivatives. Consequently, the resulting model is capable of representing the influences the nodes exert on each other, the influence the extrinsic input has on node connections, and the direct impact the extrinsic input exerts on the neuronal states (see Section 2-2-1). One important aspect that this model does not include is "neuronal gain control". This is the phenomenon by which the activity in one node directly influences the connection between two other nodes, expressed by the second-order derivative with respect to z (last term of Equation 2-21). This effect is especially crucial for learning and attentional modulation, which is the ability of the brain to selectively enhance or suppress the processing of specific sensory signals based on their relevance. To incorporate the gating mechanism, the authors of the seminal paper proposed a nonlinear extension to DCM [67], where the neuronal dynamics model is given by

$$
\begin{aligned}
\dot{z} &\approx Az + \sum u_j B^j z + Cu + \sum z_i D^i z \\
&= \left(A + \sum u_j B^j + \sum z_i D^i\right) z + Cu \\
A &= \frac{\partial F}{\partial z} = \frac{\partial \dot{z}}{\partial z} \\
B^j &= \frac{\partial^2 F}{\partial z \partial u_j} = \frac{\partial}{\partial u_j} \frac{\partial \dot{z}}{\partial z} \\
C &= \frac{\partial F}{\partial u} \\
D^i &= \frac{1}{2} \frac{\partial^2 F}{\partial z_i^2}.
\end{aligned}
\tag{2-22}
$$

Here, non-zero entries D^i_{kl} indicate that the causal connection from node k to node l depends on neural activity in node i. To visualize the meaning of the coupling matrices, Figure 2-9 depicts a graphical representation of the model.

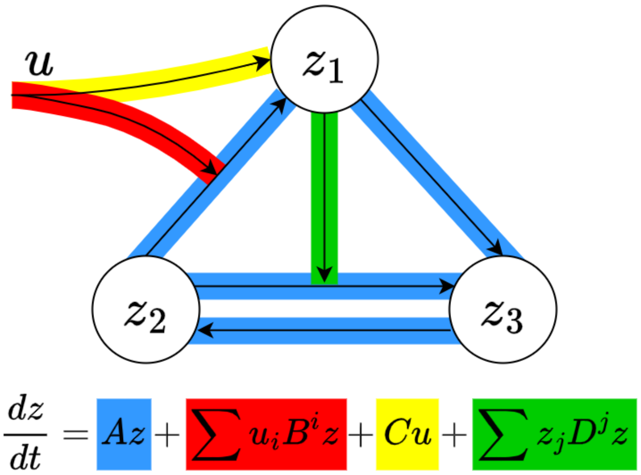


Figure 2-9: Graphical meaning of matrices in the neuronal model in Nonlinear DCM, with some arbitrarily drawn connections. Blue arrows correspond to matrix A connections, green arrows to the connections of matrices B^j, yellow arrows to matrix C connections, and red arrows to the connections of matrices D^i.

The rest of the analysis is almost identical to that in the original DCM framework. The nonlinear neural dynamics model is combined with the hemodynamic model described in Section 2-2-1 and the full model is identified using the EM estimation scheme explained in Section 2-2-2. Priors on the parameters of the D^i matrices are derived exactly the same as for the A and B^j matrices, as discussed in Section 2-2-2. There is however an added difficulty when estimating system parameters in that the D^i matrix is now a function of the states, which change continuously. The D^i matrices are needed to evaluate the state equation, which has to be integrated to obtain the state at the current time step that is then fed into the output equation. In the classic DCM formulation, all coupling matrices are dependent only on the input signals. Due to the typical sparsity of inputs in fMRI, the functions need to be reevaluated only when the inputs change. However, with the addition of the D term, the state equation continuously changes and thus has to be reevaluated at every time step.

2-4-2 Two-state DCM

The neuronal model in the seminal DCM paper involves one state per node. This means that the neuronal activity within a given brain region is pooled and considered as a single signal. This is a rather simplistic approach that neglects many important processes. One detail that it fails to model is that neurons in the brain can have either an excitatory or an inhibitory nature. Most of the brain's neurons are excitatory, meaning that when they fire, they excite other neurons and propagate signals through the brain. Some neurons are however inhibitory, which means they inhibit other neurons, making them less likely to fire. To incorporate these effects into DCM, [66] has proposed a two-state neuronal model. The authors argue that the two-state model can represent more complex dynamics than its single-state counterpart, including periodic and harmonic oscillatory modes. Furthermore, single-state DCM ensures system stability by imposing shrinkage priors (prior variances that guarantee negative real parts of Jacobian eigenvalues). This precludes identifying systems that are close to instability. However, when using the more biologically plausible two-state model, the dynamics of excitatory and inhibitory neurons inherently grant more stability.

In the approach, each brain region i is modeled with an excitatory state z_i^E and an inhibitory state z_i^I. This means that the system's Jacobian matrix $J = A + \sum u_j B^j$ changes to accommodate the extra state per region, resulting in

$$\begin{aligned}
\dot{z} &= \mathfrak{J}z + Cu \\
\mathfrak{J} &= \begin{bmatrix} \mathfrak{J}_{11}^{\mathrm{EE}} & \mathfrak{J}_{11}^{\mathrm{EI}} & \cdots & \mathfrak{J}_{1N}^{\mathrm{EE}} & 0 \\ \mathfrak{J}_{11}^{\mathrm{IE}} & \mathfrak{J}_{11}^{\mathrm{II}} & & 0 & 0 \\ \vdots & & \ddots & & \vdots \\ \mathfrak{J}_{N1}^{\mathrm{EE}} & 0 & & \mathfrak{J}_{NN}^{\mathrm{EE}} & \mathfrak{J}_{NN}^{\mathrm{EI}} \\ 0 & 0 & \cdots & \mathfrak{J}_{NN}^{\mathrm{IE}} & \mathfrak{J}_{NN}^{\mathrm{II}} \end{bmatrix}, \quad z = \begin{bmatrix} z_1^{\mathrm{E}} \\ z_1^{I} \\ \vdots \\ z_N^{\mathrm{E}} \\ z_N^{\mathrm{I}} \end{bmatrix} \\
\mathfrak{J}_{ij}^{**} &= \mu_{ij}^{**}\exp(A_{ij}^{**} + \sum u_k B_{ij}^{**(k)}) = \mu_{ij}^{**}\exp(A_{ij}^{**}) \prod \exp(u_k B_{ij}^{**(k)}), \quad * \in \{E, I\}.
\end{aligned} \tag{2-23}$$

The coupling matrix $\mathfrak{J}$ consists of 2×2 blocks corresponding to intrinsic (within-region) connections on the diagonal and extrinsic (between-region) connections off the diagonal. Regarding the intrinsic coupling, all possible connections are allowed, including self-connections $E \to E$ and $I \to I$, as well as interstate connections $E \to I$ and $I \to E$. Furthermore, connections $E \to E$, $E \to I$, and $I \to I$ are set to be negative (i.e., $\mathfrak{J}_{ii}^{EE}, \mathfrak{J}_{ii}^{EI}, \mathfrak{J}_{ii}^{II} \leq 0$) to ensure that they mediate a dampening effect on the response.

Furthermore, anatomical studies suggest that long-range connections between inhibitory populations should not be allowed. This means that extrinsic connections should be mediated only through the connections of excitatory populations $E \to E$, which are assumed to be positive (i.e., $\mathfrak{J}_{ij}^{EE} \geq 0$). The authors have set the prior mean of the extrinsic connections to $\mu_{ij}^{EE} = 0.5$. Finally, the authors have set the variance of all the parameters of the A and B

matrices to $\nu = 1/16$. This allows the prior to be scaled by up to a factor of two, which is "mildly informative", according to the authors.

The two-state variation of DCM changes the number of neuronal states but the general form of the state equation remains the same. The remaining steps of the framework are also the same as in classical DCM: the two-state neuronal model is combined with the biophysical model (see Section 2-2-1) and then the EM scheme is applied to estimate model parameters (see Section 2-2-2).

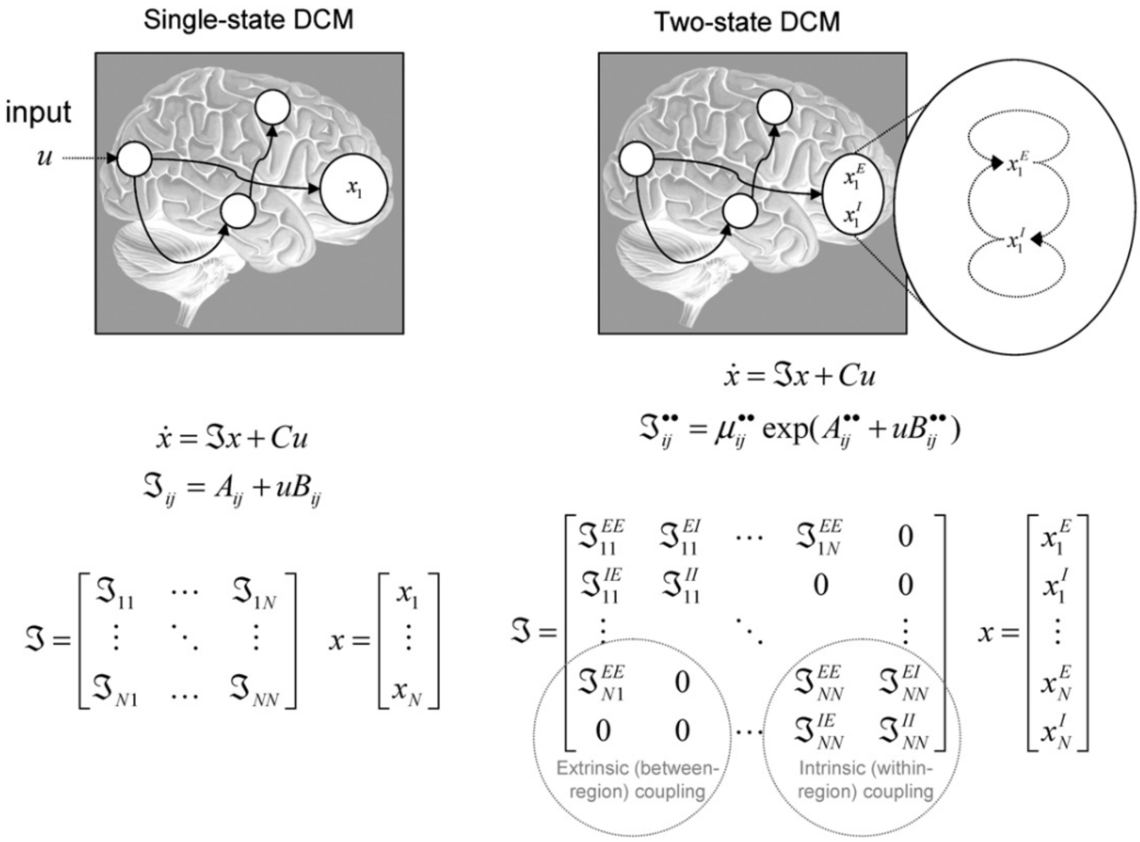

Single-state DCM
Two-state DCM
input
u
x_1
x_1^E
x_1^I
$\dot{x} = \Im x + Cu$
$\Im_{ij} = A_{ij} + uB_{ij}$
$\Im^{\bullet\bullet}_{ij} = \mu^{\bullet\bullet}_{ij} \exp(A^{\bullet\bullet}_{ij} + uB^{\bullet\bullet}_{ij})$
$\Im = \begin{bmatrix} \Im_{11} & \cdots & \Im_{1N} \\ \vdots & \ddots & \vdots \\ \Im_{N1} & \cdots & \Im_{NN} \end{bmatrix}$ $x = \begin{bmatrix} x_1 \\ \vdots \\ x_N \end{bmatrix}$
$\Im = \begin{bmatrix} \Im^{EE}_{11} & \Im^{EI}_{11} & \cdots & \Im^{EE}_{1N} & 0 \\ \Im^{IE}_{11} & \Im^{II}_{11} & & 0 & 0 \\ \vdots & & \ddots & & \vdots \\ \Im^{EE}_{N1} & 0 & & \Im^{EE}_{NN} & \Im^{EI}_{NN} \\ 0 & 0 & \cdots & \Im^{IE}_{NN} & \Im^{II}_{NN} \end{bmatrix}$ $x = \begin{bmatrix} x_1^E \\ x_1^I \\ \vdots \\ x_N^E \\ x_N^I \end{bmatrix}$
Extrinsic (between-region) coupling
Intrinsic (within-region) coupling

Figure 2-10: Comparison between the classical single-state DCM and the two-state DCM [66].

2-4-3 Stochastic DCM

The classical formulation of DCM models brain dynamics as a deterministic system, which means that it does not account for any randomness or model fluctuations. However, the brain is known to exhibit a large amount of "neuronal noise", which refers to the random firing of neurons and variations in transmission between neurons, arising from inherent stochasticity at the cellular and molecular levels [85, 86, 87, 88]. Neuronal noise plays an important role in adapting and responding to environmental cues, affecting decision-making and information processing. To account for this phenomenon, researchers have developed Stochastic DCM [65, 77, 84], which models noise in neuronal system states. Apart from rendering the model more biologically plausible, this approach also offers more robustness to model misspecification and improved convergence in system identification.

Just like in the classical formulation of DCM, the stochastic approach models observation error ε in the output equation

$$y = h(u, \theta) + \varepsilon. \tag{2-24}$$

On top of that, Stochastic DCM also models neuronal noise as additive noise ϖ in the neuronal state equation

$$\dot{z} = Az + \sum u_j B^j z + Cu + \varpi. \tag{2-25}$$

By accommodating fluctuations in neuronal states, Stochastic DCM faces several challenges. Firstly, it is self-evident that there is a considerable increase in the number of unknown variables. As a result, there is less information in the data per unknown variable, making the algorithm more sensitive to priors. Moreover, the activity in the network is now considered to be caused by both experimental input u and neuronal noise ϖ. Therefore, the system identification algorithm has to determine the extent of contributions of the two sources, which becomes increasingly challenging as the noise-to-signal ratio increases. Because of the increased difficulty in identifying the system, fitting a Stochastic DCM takes approximately 30x more time than fitting its deterministic counterpart [84]. While this makes the approach considerably more computationally expensive, the resulting models have been shown to be less overconfident, providing a less biased fit to the data [84]. They can also capture phenomena not explained by the simplistic deterministic model approximation and can deal better with incorrectly specified model structures.

To identify the parameters, hidden states, and error hyperparameters of this new model, Stochastic DCM can apply one of the following two methods: Dynamic Estimation Maximization (DEM) [89, 90] or Generalised Filtering (GF) [65, 91]. The focus of this section is on DEM as it is the original scheme used in Stochastic DCM, and it bears more resemblance to the EM algorithm applied in classical DCM, explained in Section 2-2-2. The DEM algorithm and its relation to EM are especially thoroughly discussed by [73], who have reformulated the algorithm for the control systems community and provided missing derivations. The following discussion aims to summarize the most important aspects of DEM, and the detailed proofs can be found in [73].

The two system identification algorithms are both based on the Free Energy Principle (FEP), which is a unified theory of how the brain combines prior knowledge with stimuli from the environment to learn and adapt. Using this principle, the EM and DEM algorithms find the optimal parameters by maximizing the free energy defined as

$$F = \ln p(y) - D(q(\vartheta) \| p(\vartheta \mid y)), \tag{2-26}$$

where $\vartheta = [z\ u\ \theta\ \lambda]^T$ is the vector of all hidden states, parameters, and hyperparameters. Furthermore, $\ln p(y)$ is the log-evidence, D denotes the Kullback-Leibner divergence, $q(\vartheta)$ is the approximate posterior distribution, and $p(\vartheta \mid y)$ is the real (unknown) posterior. Since divergence is always positive, F creates a lower bound for the log-likelihood

$$F \leq \ln p(y). \tag{2-27}$$

The goal is to maximize the free energy, thus minimizing the divergence, rendering $q(\theta)$ a suitable approximation of the actual distribution.

While both EM and DEM are based on the above principles, they have two significant differences. What the DEM algorithm introduces are the use of generalized coordinates and a mean-field approximation for modeling the free energy. Both of these additions are crucial for handling complexities associated with estimating parameters in systems with state noise. In the ensuing discussion, two operators are used to denote a variable in its generalized coordinates (tilde operator $\tilde{}$) and a time integral of a variable (bar operator $\bar{}$).

First of all, DEM models the time evolution of states using generalized coordinates, meaning that the states are expressed using their higher-order derivatives, i.e., $\tilde{z} = [z\ z'\ z'' \dots]^T$. As a result of using the generalized coordinates, DEM tracks the trajectories of states instead of their point estimates like EM does. This is crucial for handling state noise, as it allows DEM to model the correlation between noise derivatives, providing a more accurate estimation under state noise.

Furthermore, DEM employs a different approximation of the free energy. When the unknown variables include both time-invariant and time-varying parameters, the free action $\bar{F}$ is used as the objective function, instead of F. The free action is defined as the integral of the free energy over time,

$$\bar{F} = \bar{V}(\vartheta) + \bar{H}(\vartheta) = \int \langle U(y,\vartheta)\rangle_{q(\vartheta)} dt + \int H(\vartheta)_{q(\vartheta)} dt, \tag{2-28}$$

where $V(\vartheta) = \langle U(y,\vartheta)\rangle_{q(\vartheta)}$ is called the Variational Free Energy (VFE). The internal action U is expressed as a sum of the log-likelihoods of the posterior output distribution and prior parameter distribution, given by

$$U(y,\vartheta) = \ln p(y \mid \vartheta) + \ln p(\vartheta). \tag{2-29}$$

By assuming that the densities of the parameters constituting ϑ are independent of each other, i.e.,

$$q(\vartheta) = q(\tilde{z})q(\tilde{u})q(\theta)q(\lambda), \tag{2-30}$$

the formula for the internal action $U(y,\vartheta)$ can be simplified using the second-degree Taylor series expansion near the mean $\mu^{\vartheta} = \left\{\mu^{\tilde{z}}, \mu^{\tilde{u}}, \mu^{\theta}, \mu^{\lambda}\right\}$, resulting in,

$$U(y,\vartheta) = U\left(y,\mu^{\vartheta}\right) + \sum_{i=1}^{4}\sum_{j=1}^{4}\left(\theta^i - \mu^i\right)^T U\left(y,\mu^{\vartheta}\right)_{\theta^i\theta^j}\left(\theta^j - \mu^j\right). \tag{2-31}$$

The above approximation of the internal energy may now be substituted into the formula for the VFE, giving,

$$\begin{aligned}\bar{V} &= \bar{U}(y,\mu^{\vartheta}) + \int W dt \\ W &= \frac{1}{2}\sum_{i,j=1}^{4} \operatorname{tr}\left[\Sigma^{ij} U\left(y,\mu^{\vartheta}\right)_{\theta^i\theta^j}\right].\end{aligned} \tag{2-32}$$

Here, W is called the mean-field term and represents the second important addition of the DEM algorithm. In comparison, the EM scheme neglects the mean-field terms. Since the parameters in ϑ are independently distributed, the covariance between them is zero, simplifying the equation to

$$\begin{aligned}\bar{V} &= \bar{U}(y,\mu^{\vartheta}) + \int [W^{\tilde{z}} + W^{\tilde{u}} + W^{\theta} + W^{\lambda}] dt \\ W^{\vartheta^i} &= \frac{1}{2} \operatorname{tr}\left[\Sigma^{i} U\left(y,\mu^{\vartheta}\right)_{\theta^i\theta^i}\right].\end{aligned} \tag{2-33}$$

This formulation of the VFE may now be substituted into the formula for the free action given by Equation 2-28, concluding the derivation of the objective function. The resulting iterative scheme is given by three distinct steps:

1. D step: generalized state and input estimation,
2. E step: parameter estimation,
3. M step: noise/error hyperparameter estimation,

where the E and M steps are equivalent to the EM algorithm.

Chapter 3

Model development

This chapter provides an overview of the work carried out to obtain a model of emotional memory that fulfills the project objectives. Firstly, Section 3-1 introduces the dataset and the experimental setting. Additionally, all the design decisions involved in Dynamic Causal Modeling (DCM) are described, such as input signal modeling and choice of equations describing the system. Then, Section 3-2 summarizes all the steps performed in MATLAB to pre-process the data, fit the DCM model, and analyze the results.

3-1 Modeling approach

The goal of the project is to use functional Magnetic Resonance Imaging (fMRI) data from the experiment described below to build a dynamical model of the neural underpinnings of emotional memory. The selected modeling framework is DCM, which models neural dynamics in a network of pre-selected brain regions that respond to external stimuli and collectively give rise to the signal measured with fMRI. Based on the literature, the selected nodes are the Amygdala (Amy), the Hippocampus (Hip), and the Orbitofrontal Cortex (OFC), see Section 1-2 and Section 1-3.

3-1-1 Experimental paradigm

The dataset used in this thesis comes from an experiment conducted by Zhu et al. at Donders Institute for Brain, Cognition, and Behaviour [1]. The purpose of the experiment was to study associative memory of neutral and emotional images. It should be noted that the experiment studied both the encoding and retrieval phases of memory, whereas the focus of this thesis is only on the encoding phase. In the study, 70 healthy adult subjects were asked to memorize 48 so-called "ABC triplets" of images, while having their brain activity measured using fMRI. An "ABC triplet" consists of three images, where:

- "A" corresponds to a neutral cue (one of 48 distinct locations in a cartoon map)

- "B" and "C" correspond to differently valenced images including neutral images (e.g., a building) and negative images (e.g., an angry barking dog).

To study the effects of different emotional values of visual stimuli, the participants were divided into 3 groups, based on the type of images they would see:

- group Neutral-Neutral (NN), where images "B" and "C" were neutral,
- group Emotional-Emotional (EE), where images "B" and "C" were emotional,
- group Neutral-Emotional (NE), where image "B" was neutral and "C" was emotional.

The experimental design is depicted in Figure 3-1 using an example of an "ABC triplet", where the map location with the house with a blue roof ("A") is associated with an image of skyscrapers ("B") and an image of a barking wolf ("C"). Images "B" and "C" are not displayed simultaneously. Instead, the cue is shown with each image separately, indicating that the images should all be associated with each other. During each trial, the entire map was first shown for 0.5s, after which the cue was highlighted on the map and presented for 1s. Then, only the highlighted cue remained side-by-side with the paired image for 2.5s. Finally, there was a break lasting between 0.5-1.5s, followed by the next trial.

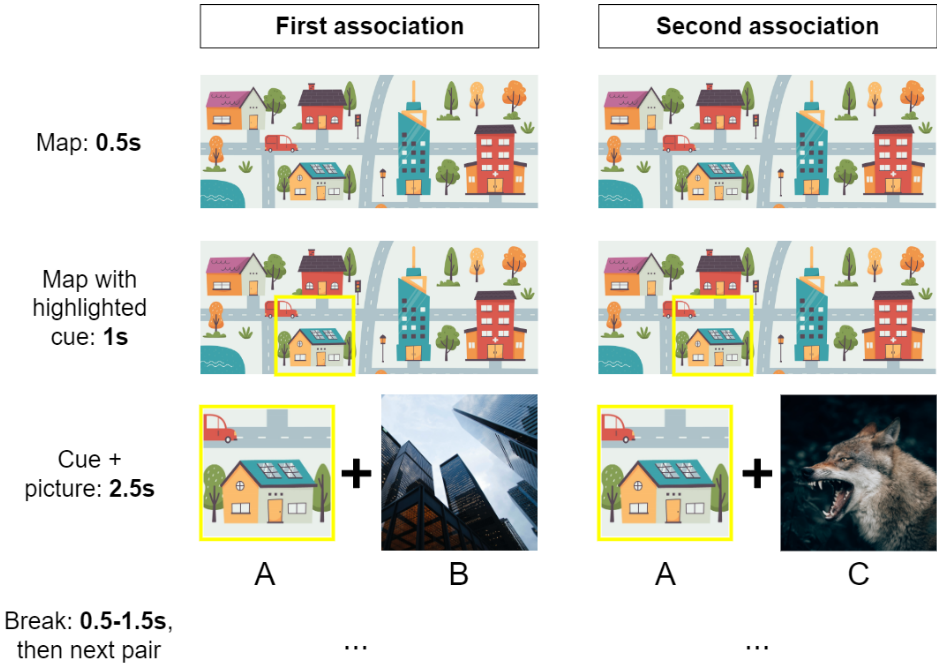


Figure 3-1: Experimental design used to obtain data for this project [1].

The 48 triplets were shown in 4 runs, each run including 12 different triplets. Within each run, all 12 first associations ("A" + "B") were shown first, followed by all 12 second associations ("A" + "C"). After all the triplets had been shown once, the entire process was repeated such that each triplet was shown exactly twice. Once all triplets were displayed twice, a run was completed and the next run with 12 new triplets commenced.

3-1-2 Input signal

In DCM, the input signal used to fit the dynamical model is a step function equal to 1 when an external stimulation is present and equal to 0 otherwise. Since the experiment used in this project includes several different types of images (map → map with highlighted cue → cue and image), a decision has to be made on what should constitute the input signal. Undoubtedly, the most relevant part of each trial is the last one, when the cue is presented next to an image. Hence, the input signal for model fitting is a step function equal to 1 while the cue and image are presented (for 2.5s) and 0 otherwise (the breaks are also about 2.5s long), see Figure 3-1 for details on experiment structure. Moreover, a distinction is made between the first associations (cue + image B) and the second associations (cue + image C), corresponding to two input elements u_1 and u_2, respectively. Figure 3-2 depicts a visualization of input signals in a dummy example with 2 "ABC triplets".

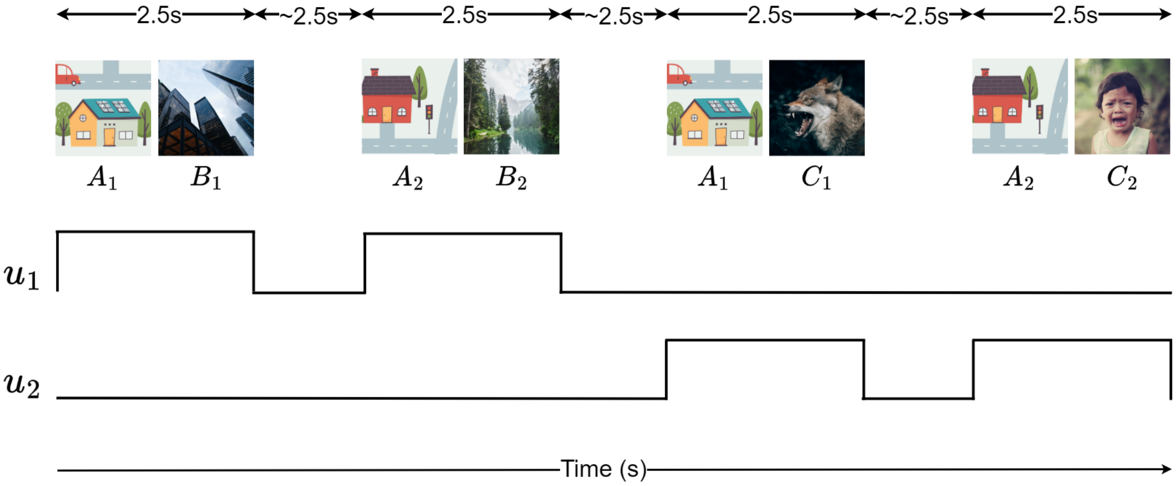


Figure 3-2: Example of input signals u_1 and u_2 used for fitting DCM in a dummy case with only 2 triplets in group NE. The graphic is not to scale.

3-1-3 Model space

DCM requires the user to specify which model connections are present/allowed. If a connection is allowed, the algorithm will update its value during model fitting. However, if a connection is not allowed, its strength will remain at 0. Since the actual model structure (allowed vs. non-allowed connections) is rarely known beforehand, DCM users typically define a model space of plausible model structures that should all be fitted. The model space choice should be rooted in biological assumptions and results from similar past studies. Then, Bayesian Model Selection (BMS) is performed to identify the optimal structure, see Section 2-2-3 for details. Having to fit multiple models just to find the best structure is computationally expensive, which is a common point of criticism of DCM.

The size of the model space in this project is kept to a minimum because the dataset includes a large number of subjects and the computational power is limited. The idea is thus to first focus on the small model space and see if any of the models fits the data sufficiently well. Firstly, all connections in the A matrix are allowed, which is supported by similar studies introduced in

Section 1-3. The related studies modeled (a subset of) the network consisting of the Amy, the Hip, and one or more subregions of the Prefrontal Cortex (PFC). The connections identified in these studies are summarized in Table 3-1, where the various PFC subregions are treated as one node. As can be seen in the table, most studies found bidirectional connections between all nodes, with the exception of [49], where only top-down connections with the PFC were found. Based on these results, it is expected that the nodes in the network modeled in this project should also all be interconnected. Furthermore, the self-connections corresponding to the diagonal matrix entries should also be allowed, as they serve an important biological role in inhibiting neural activity from reaching too high magnitudes [92]. Overall, this means that all the models fitted in this project have A matrices with all connections allowed.

Study in [46]	**Study in [14]**	**Study in [47]**	**Study in [48]**	**Study in [49]**
Amy ↔ Hip	Amy ↔ PFC	Amy ↔ Hip	Amy ↔ Hip	Amy ↔ Hip
		Amy ↔ PFC	Amy ↔ PFC	Amy ← PFC
		Hip ↔ PFC	Hip ↔ PFC	Hip ← PFC

Table 3-1: Between-region connections in the A matrix found in similar studies. The double-sided arrow ↔ means a bidirectional connection was found, and the one-sided arrow ← means only a connection in the specified direction was found. Not all studies modeled all three Amy, Hip, and PFC, hence the empty entries.

Moving on to the B matrices, it is reasonable to assume that both inputs affect the same connections so B_1 and B_2 should have the same structure. This is because the inputs correspond to the same type of external stimulation, i.e., viewing pairs of images. Consequently, all connections in the B matrices should be allowed such that the different influences of inputs (emotional vs. neutral) can be reflected in the strength of the connections. This approach has also been adapted by [14], which is the only other found DCM study of emotional associative memory in a similar network.

Similarly, both columns of matrix C should have the same allowed connections such that the two inputs affect the same nodes. In this case, however, not all connections should be allowed. All of the related studies ([14, 46, 47, 48, 49]) found that the input acts directly on only one of the nodes. This can be seen as the external stimuli entering the brain network through one node, which passes on the influence of the input to the other nodes. As a result, three possible C matrices should be explored in this project, reflecting the external input acting directly on the Amy, the Hip, or the OFC. This concludes the definition of the model space. The allowed connections of the 3 models to be fitted are summarized in matrix form in Equation 3-1 and visually in Figure 3-3.

$$\begin{gathered}\frac{dz}{dt} = \underbrace{\begin{bmatrix}1&1&1\\1&1&1\\1&1&1\end{bmatrix}}_{\mathrm{A}} z + \underbrace{\begin{bmatrix}1&1&1\\1&1&1\\1&1&1\end{bmatrix}}_{\mathrm{B}_1} u_1 z + \underbrace{\begin{bmatrix}1&1&1\\1&1&1\\1&1&1\end{bmatrix}}_{\mathrm{B}_2} u_2 z + \mathrm{C}\begin{bmatrix}u_1\\u_2\end{bmatrix}, \\ \text{where } \mathrm{C} = \begin{bmatrix}1&1\\0&0\\0&0\end{bmatrix} \text{ or } \mathrm{C} = \begin{bmatrix}0&0\\1&1\\0&0\end{bmatrix} \text{ or } \mathrm{C} = \begin{bmatrix}0&0\\0&0\\1&1\end{bmatrix}.\end{gathered} \tag{3-1}$$

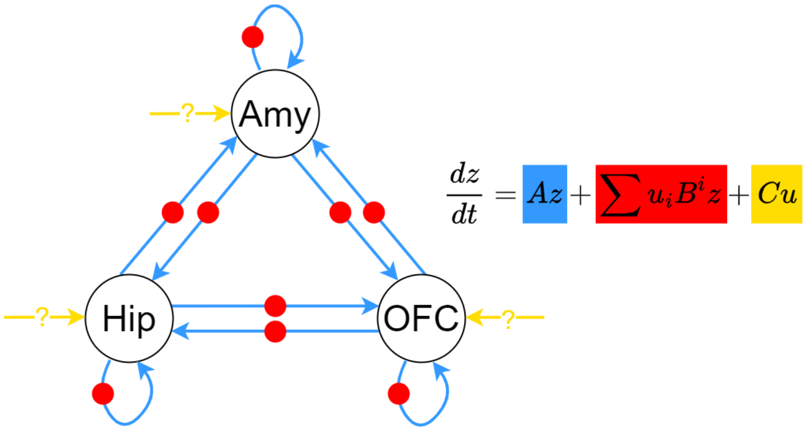
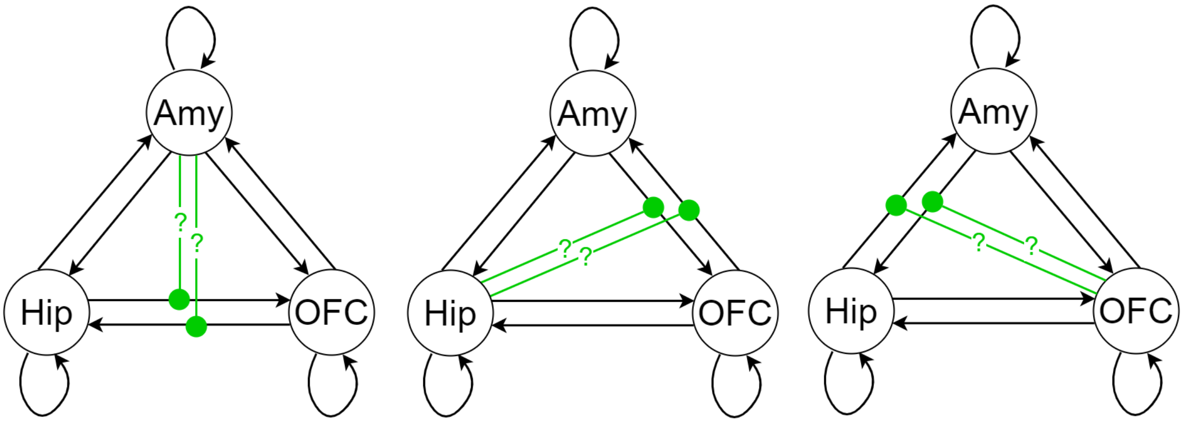


Figure 3-3: Model structures to be fitted. The arrows symbolize connections, whereas the dots correspond to the modulatory influences of the inputs on the connections.

Finally, as elaborated on in Section 2-4, apart from the classical formulation of DCM, other variations that are relevant to this project are Nonlinear DCM, Two-State DCM, and Stochastic DCM. The aim is to fit all 4 types of DCM and identify which one is best suited for the goal of this thesis. Most of these variations do not require making changes to the model space. The exception is Nonlinear DCM, which requires specifying which nodes are assumed to modulate the connections between other nodes. This modulatory influence is reflected by the $\sum z_j D^j z$ term, which in this case contains three matrices D_1, D_2, and D_3 since there are three neuronal states z. In this project, it is assumed that only one node may exert modulatory influence on the other connections, which is consistent with the findings of related studies [48, 49]. Furthermore, it is assumed that the modulatory influence affects both connections between the other nodes, to avoid exploring too many combinatorial possibilities. This means that there are three possible structures of the D matrices, summarized in Equation 3-2 and depicted in Figure 3-4. Since there are also three possibilities for the C matrix, there are $3 \times 3 = 9$ Nonlinear DCM models to be estimated.

$$\begin{aligned}\frac{dz}{dt} &= Az + B_1 u_1 z + B_2 u_2 z + \mathrm{C}\begin{bmatrix} u_1 \\ u_2 \end{bmatrix} + D_1 z_1 z + D_2 z_2 z + D_3 z_3 z,\\ \text{where } \mathrm{D}_1 &= \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix}, D_2 = D_3 = [0],\\ \text{or } \mathrm{D}_2 &= \begin{bmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \\ 1 & 0 & 0 \end{bmatrix}, D_1 = D_3 = [0],\\ \text{or } \mathrm{D}_3 &= \begin{bmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}, D_1 = D_2 = [0].\end{aligned} \tag{3-2}$$



$$\frac{dz}{dt} = Az + \sum u_i B^i z + Cu + \sum z_j D^j z$$

Figure 3-4: Structures of the D matrix to be fitted in Nonlinear DCM. The black arrows symbolize connections in the A matrix, which may be modulated by other nodes through the green connections.

3-2 Model fitting

DCM has been implemented in the Statistical Parametric Mapping (SPM12) toolbox [93] in MATLAB, which is a commonly used tool for analyzing functional imaging data. DCM model fitting requires several steps, including extracting time series from the relevant brain regions, fitting all models from the model space, and comparing the fitted models to find the optimal one. The model fitting procedure applied in this project is based on an example from the SPM12 manual [94]. The goal of this section is to summarize and explain all the modeling steps performed in MATLAB. Furthermore, scripting in SPM12 is not well documented (not even in the manual) so one of the contributions of this thesis is the quick guide to SPM12 DCM scripting in Appendix B.

3-2-1 Data pre-processing

To begin with, the first 10 values from each run are removed because they correspond to the initial 10 seconds, during which the scanner was still being set up. Furthermore, it has been decided to fit the models using data from only the first out of the four runs of each participant. The runs were recorded separately, each involving 12 different ABC triplets displayed over a time span of around 250 seconds. As there is a considerable number of subjects and model structures to fit, and DCM is generally computationally heavy, reducing the size of the dataset is necessary. In fact, even with this smaller dataset, the entire model-fitting process takes around 24 hours on a powerful data-processing server that uses an Intel Xeon Silver 4314 processor.

The next stage in pre-processing data is estimating a General Linear Model (GLM). In the context of neuroscience, GLM is a statistical framework that aims to capture the relationship between the observed signal (fMRI data) and the experimental conditions [94]. This step

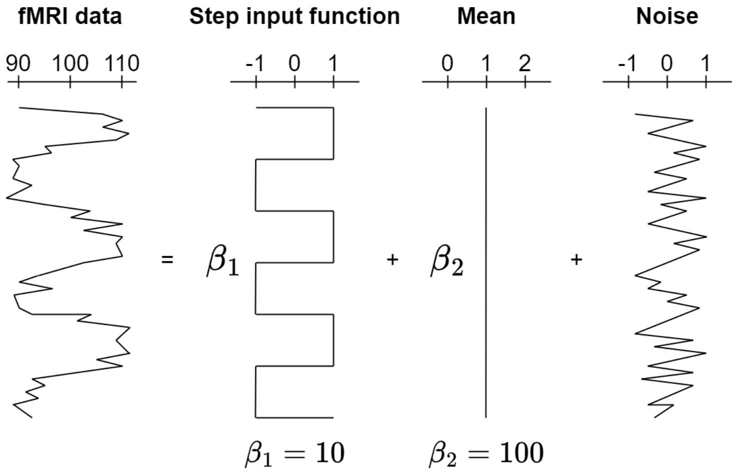


Figure 3-5: Example of a GLM observation model.

serves to denoise the data and identify the neural responses that correspond to experimental stimulation. The model assumes that the observed data is a linear combination of the input signal, a constant mean value, and noise or error. An example of a GLM model is depicted in Figure 3-5. During the estimation process, the observed data is separated into the 3 components such that each component represents the contribution of its associated variable (experimental conditions, mean, and noise). This step requires specifying the experimental conditions (here, u_1 corresponding to the first associations and u_2 corresponding to the second associations), their timings (here, step functions as explained in Section 3-1-2), as well as the sampling frequency of the fMRI machine used to collect data (here, 1 second [1]).

After the GLM has identified the data components that correspond to external stimulation, the next step is to define the brain areas of interest, known as Volumes of Interest (VOIs). In this project, there are three brain regions to be modeled, namely the Amy, the Hip, and the OFC. The goal is to identify the relevant voxels (units on a 3D grid) for each of the brain regions. As depicted in Figure 3-6, each VOI is defined as the voxels that constitute the anatomical region and that show significant responses to both inputs. The anatomical locations of selected brain regions are specified using WFU_PickAtlas [95], which is implemented as a MATLAB toolbox that generates brain masks based on brain atlases. The atlas used in this project is the Automated Anatomical Labelling (AAL) atlas [96]. As Amy and Hip are bilateral (located on both sides of the brain), it is important to decide which side to include in the model. Here, only the left Amy and Hip are selected because they are known to play a more important role in emotional memory and it is common practice to restrict the DCM analysis of emotional memory to the left side only [46, 48, 49].

Furthermore, identifying voxels that respond to external stimulation requires specifying T and F contrasts [94]. These contrasts are mixtures of GLM parameter estimates assessed in relation to some null hypothesis. Firstly, T-contrasts test whether a specific parameter in the GLM has significant importance. Here, two T-contrasts are applied to test whether each individual input (u_1, u_2) has a significant effect on the given brain region. Then, an F-contrast is applied to test whether a set of experimental conditions (u_1 and u_2) has a joint effect on neural activity within a VOI. Applying the T and F contrasts requires specifying

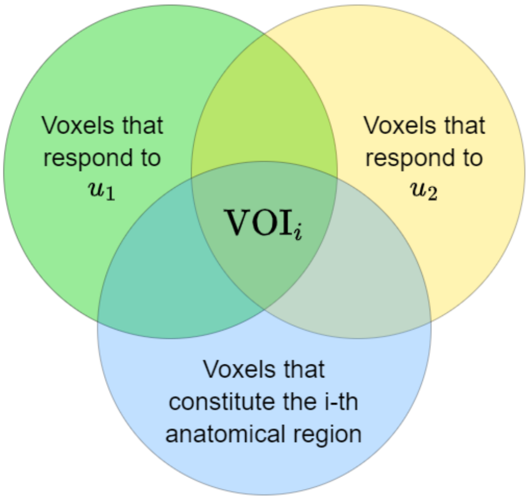

Voxels that respond to u_1
Voxels that respond to u_2
$\mathbf{VOI}_i$
Voxels that constitute the i-th anatomical region

Figure 3-6: Venn diagram visualizing the definition of what voxels constitute VOIs.

their p-values. The p-value approach is a statistical significance test that checks whether there is evidence to reject the null hypothesis [97]. In this project, the p-value is set to 0.05, which is a commonly used p-value in other DCM studies as well as in the example in the SPM12 manual [94].

Finally, by applying the T and F contrasts on each anatomical region, relevant voxels are identified for all three VOIs. The last step is to obtain a mean of the signals from voxels in each VOI, resulting in three 1-dimensional time-series corresponding to the Amy, the Hip, and the OFC. This procedure is carried out for all 70 subjects but for some of the people, no significant voxels are found in at least one of the brain regions. These subjects are excluded from further analysis, leaving 22 people in the NN group, 20 people in the NE group, and 23 people in the EE group. Furthermore, Figure 3-7 depicts a histogram that shows the distribution of the number of relevant voxels found in each VOI for all subjects. The figure shows that for all nodes, around 90% of the values are within the first 3 bins with overall ranges: 0-1500 for Amy, 0-3000 for Hip, and 0-600 for OFC. Furthermore, the averages of the bins constituting the top 90% highest values are 589 for Amy, 1277 for Hip, and 257 for OFC.

3-2-2 Subject-level DCM

The goal of DCM in this project is to obtain one optimal model per group (NN/NE/EE) that explains the data of all subjects within that group sufficiently well. This requires performing a subject-level analysis first, meaning that all models from the model space (see Section 3-1-3) are fitted for every person. The model space consists of 18 models including 3 classical DCM models, 3 Stochastic DCM models, 3 Two-State DCM models, and 9 Nonlinear DCM models.

At this stage, the step-function input signals are convolved with a canonical Hemodynamic Response Function (HRF) (see Figure 2-2). This is a crucial step in modeling how neural activity translates into the observed hemodynamic response in fMRI data. Furthermore, two timing parameters need to be specified, namely the Echo Time (TE) and the slice timings. The former is a property of the fMRI machine used to collect data and in this case, it is equal

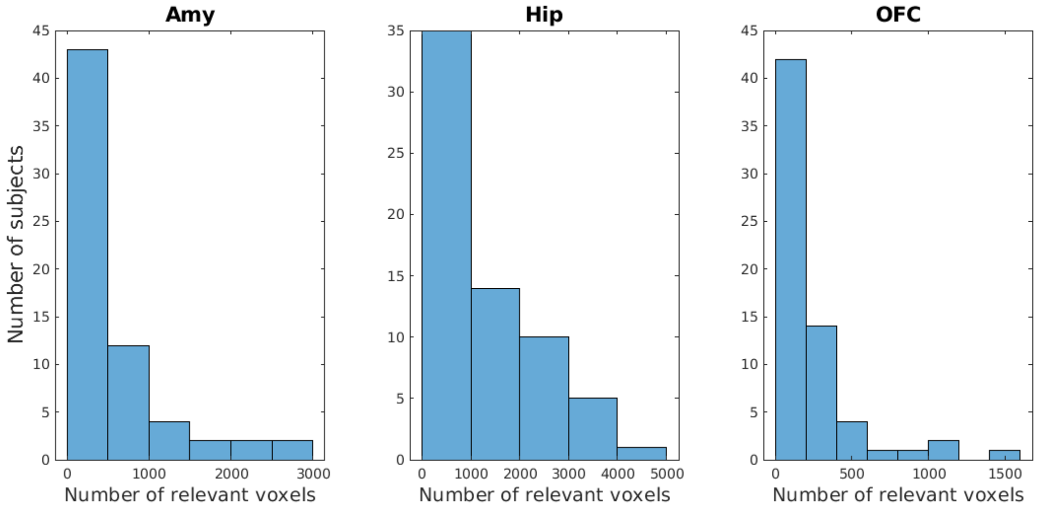


Figure 3-7: Histogram of the distribution of the number of relevant voxels found in each brain region for all subjects.

to 0.04 seconds [1]. Moreover, slice timing refers to the process of correcting for the time differences in the acquisition of different slices while obtaining fMRI data. This is necessary because fMRI machines sequentially capture the activity in 2-dimensional slices to obtain a full 3-dimensional image. During the sampling period (here, the Temporal Resolution (TR) is 1 second), the consecutive slices are obtained at slightly different times and so a slice timing correction has to be applied to reduce the errors caused by the mismatch. In this project, data is shifted to correspond to the middle of each slice (TR/2 = 0.5 seconds), as recommended in the SPM12 manual. As a result, 18 DCM models are estimated for all 65 subjects. This concludes setting up the subject-level DCM analysis.

3-2-3 Group-level DCM

The next step is identifying which one of the 18 types of models fits the data of subjects from each group (NN,NE,EE) best. This analysis is carried out by performing Bayesian Model Selection (BMS), which is a statistical method described in Section 2-2-3. The approach computes the log model evidence $\ln p(y|m)$ of each model structure, which is a metric that assesses the model's ability to explain the data with minimal structural complexity. If the difference in log evidence between one model and all the other models is at least 3, it is considered that the evidence is strong that this model is optimal.

When applying BMS, the user must decide whether to assume that one model structure is optimal for all subjects in a group (Fixed-Effects (FFX)) or if every subject might have a different optimal model (Random-Effects (RFX)). In this project, FFX analysis is applied because it is assumed that there is an optimal model that fits all subjects from a group. This assumption is valid because all participants were exposed to the same experimental conditions. In contrast, if conditions varied across subjects, it would not be reasonable to assume that one model should fit all people within a given group.

After BMS has revealed the optimal model structure for each group, it is time to estimate the parameters of these optimal models. This is done by performing Bayesian Parameter Averaging (BPA), which takes the models with the optimal structure and obtains "average" values of model parameters. To obtain the "average" values, BPA computes the joint posterior distribution by taking the posterior of one subject and treating it as the prior for obtaining the posterior of the next subject. In this project, the average optimal models are obtained separately for each type of DCM (Classical, Nonlinear, Two-State, and Stochastic). As a result, each group (NN,NE,EE) has 4 optimal models, one per DCM type. The purpose of the following chapter is to examine which DCM variation results in the best average model.

Chapter 4

Results and analysis

In order to obtain a nonlinear dynamical model for the Amygdala (Amy)-Hippocampus (Hip)-Orbitofrontal Cortex (OFC) network, several variations of the Dynamic Causal Modeling (DCM) framework are utilized, including the Classical, Nonlinear, Two-State, and Stochastic formulations. This chapter presents all the developed models, compares them, and provides an analysis of the chosen model. First, Section 4-1 discusses the performance of each DCM variation and identifies the most suitable one for the thesis objectives. The chosen model is then thoroughly analyzed in Section 4-2. This section discusses the estimated structure, parameter values, and stability properties across all experimental groups. Analyzing these properties brings insights into how the networks communicate in different groups, which is an indicator of how well memories form depending on the experimental condition.

4-1 Model comparison

In order to identify the best DCM variation, the Amy-Hip-OFC network is modeled using Classical, Two-State, Nonlinear, and Stochastic DCM. The last three methods extend the model and/or the system identification algorithm of the classical approach with different features. Specifically, Two-state DCM models two states per node, as opposed to one in the classical approach, which allows to capture the excitatory and inhibitory neuronal activity. This renders the model more biologically plausible and showcases more complex dynamical phenomena. Furthermore, Nonlinear DCM adds a nonlinear term $\sum x_i D^i x$ to the neuronal state equation, making the model capable of describing how a node can modulate the connectivity between other nodes. Lastly, Stochastic DCM abandons the deterministic assumptions, allowing for more error, neuronal noise, and system misspecification, leading to a more robust model.

For each of the DCM variations, subject-level and group-level analyses are performed to identify the optimal model structure and average parameters that capture the dynamics of all subjects within a group. First, the optimal model structure is identified by performing Bayesian Model Selection (BMS), where several structures are fitted and compared. The BMS procedure chooses the model structure with the highest model evidence, which is a metric that assesses the model's ability to explain data with minimal structural complexity. Then, the parameters of this model structure are obtained during Bayesian Parameter Averaging (BPA), which is an approach for estimating average parameters that explain the data of all subjects within a group. This approach is discussed in detail in Sections 3-2-2 and 3-2-3. As a result, 3 optimal models (one per each group) are found for each of the 4 DCM variations.

Classical, Two-State, and Nonlinear DCM all result in a poor fit. The simulated responses of the average optimal models are depicted in Figure A-1, Figure A-2, and Figure A-3. The responses are plotted against the average measured signal for each group. By looking at the figures, it is clear that most of the models fail to capture any details of the system dynamics. At best, the models roughly follow a rolling average of the data. On top of that, as it turns out, none of these approaches are capable of incorporating the initial conditions of the system. As a result, the simulated responses always start at zero and continue to attempt to follow system dynamics with a large offset. While the initial conditions could in principle be incorporated into the system equations, it is difficult to estimate the initial conditions of all the hidden states in the first place. Estimating the values of these states is one of the goals of the Bayesian inference in DCM. Therefore, it appears that a zero initial condition is an assumption of the DCM framework and the toolbox does not allow the user to generate responses that start elsewhere.

Conversely, the optimal models fitted with Stochastic DCM generate responses that match the measured data considerably better, as seen in Figure A-4 and Figure A-5. Not only does the response follow all the data oscillations but also, the model deals a lot better with non-zero initial conditions. While the responses still start at zero, they immediately jump to the correct level and continue to follow the measured signal without an offset.

To better quantify how the different DCM variations perform, their R-squared value is calculated according to

$$R^2 = 1 - \frac{\sum_i (y_i - f_i)^2}{\sum_i (y_i - \bar{y})^2}, \tag{4-1}$$

where y_i is the measured response at time i, $\bar{y}$ is its mean, and f_i is the predicted response at time i. This metric assesses how much of the data's variance is explained by the model - the higher the value, the better the fit. Furthermore, a negative value indicates that the model does not explain any of the variance in the data and fits the signal worse than a horizontal line. Table 4-1 presents the R-squared values calculated for the optimal models for each DCM type. Most of the values for Classical, Nonlinear, and Two-State DCM are negative or have a very low positive value. On the other hand, the R-squared values are mostly positive with significantly higher magnitudes for Stochastic DCM, with the exception of the OFC in the Neutral-Emotional (NE) and Emotional-Emotional (EE) group. The relative fit among the groups for Stochastic DCM is further examined in the following section but it is clear that this DCM variation gives the best overall fit. Based on this quantitative comparison and qualitative analysis of the model fit figures, the model obtained with Stochastic DCM is considered to be the best model and is further analyzed in the subsequent section.

		R^2_{Amy}	R^2_{Hip}	R^2_{OFC}
Classical DCM	Group NN	-6.7550	-7.8580	-3.1585
	Group NE	-2.9134	-2.0622	-1.5884
	Group EE	-3.3053	-3.2494	-0.0100
Nonlinear DCM	Group NN	-0.0411	-0.8667	-0.2948
	Group NE	-0.0422	-0.0120	-0.0272
	Group EE	0.0075	0.0154	0.0005
Two-State DCM	Group NN	-0.0411	-0.8667	-0.2948
	Group NE	0.0241	0.0065	0.0088
	Group EE	-1.9679	-2.1009	-0.0840
Stochastic DCM	Group NN	0.6493	0.7626	0.3101
	Group NE	0.3889	0.2011	-0.0344
	Group EE	0.4802	0.5599	-0.6269

Table 4-1: The R-squared between the response of the optimal average models and the average measured data, calculated per group, per node, and per DCM variation.

4-2 Analysis of the selected model

The optimal models identified using Stochastic DCM give the best fit and are therefore selected and further analyzed. This section presents various aspects of the selected models, including their structures, estimated parameters, stability properties, and state evolution properties.

4-2-1 Model structure

The network model includes 3 nodes, namely z_{Amy}, z_{Hip}, and z_{OFC} that give rise to neural activities described by

$$\dot{z}(t) = Az(t) + \sum u_i(t)B^i z(t) + Cu(t), \tag{4-2}$$

where the A, B^i, and C matrices explain how the nodes influence each other and how the external stimulation $u(t)$ impacts the network connectivity. The structure of these matrices (which connections and influences are present) is unknown so a set of plausible model structures is defined and compared during Bayesian Model Selection (BMS). In this thesis, the model set includes 3 distinct structures, where the A and B matrix structures are constant but the C matrix structure is allowed to vary, see Section 4-2-1 for a detailed explanation. This is intended to identify how the external stimulation and input from the rest of the brain enter the modeled network. The C matrix has 3 possible structures that model the external inputs entering through the Amy, through the Hip, or through the OFC. In the remainder of the discussion, the different model structures are named based on their corresponding C matrix (Amy/Hip/OFC).

Bayesian Model Selection

Bayesian Model Selection (BMS) calculates the log model evidence for each model, which quantifies how probable it is that a given model generated the measured response and how far the estimated parameters are from the priors. As a result, this metric is meant to prioritize model structures that have a high probability of having generated the data, while having the lowest model complexity. The log evidence values estimated for each model in this project are summarized in Table 4-2.

	Absolute log model evidence		
	Group NN	**Group NE**	**Group EE**
Model 1: Input to Amy	-0.0002	-32.0000	-11.4785
Model 2: Input to Hip	-32.0002	-32.0000	-0.0242
Model 3: Input to OFC	-8.5119	-0.0000	-3.7351

Table 4-2: Absolute log model evidence for models Amy, Hip, and OFC, estimated during BMS.

However, it is easier to analyze these values on a relative scale where the lowest log model evidence (ME) is subtracted from the other log model evidences:

$$\text{ME}_{\text{ relative, model i}} = \text{ME}_{\text{ absolute, model i}} - \min\{\text{ME}_{\text{ absolute}}\}. \tag{4-3}$$

This is done because the absolute log evidence metric assumes negative values so the most plausible model has the least negative log evidence. Furthermore, it is usually considered that there is strong evidence for a model being the optimal one if the difference between its log evidence and the log evidences of all other models is at least 3.

The relative log model evidence values are plotted in a bar graph depicted in Figure 4-1. For each group, there is one model structure with log evidence greater by at least 3 compared to the other models:

- group NN - input to Amy (log evidence greater by 8.5),
- group NE - input to OFC (log evidence greater by 32),
- group EE - input to Hip (log evidence greater by 3.7).

Undoubtedly, the results are the strongest for group NE, where the difference is the largest and there is no relative evidence for the other models. Conversely, the results are the least decisive for group EE, where the difference between the two best models barely exceeds the threshold of 3. Interestingly, in all groups, there is considerable evidence for the model, where input enters through the OFC, even if this model is not always the winner.

Comparison of coefficients of determination

As discussed in Section 2-3-2, it has been shown that the optimal model structure selected with BMS does not necessarily fit the data well in all network nodes. To examine which

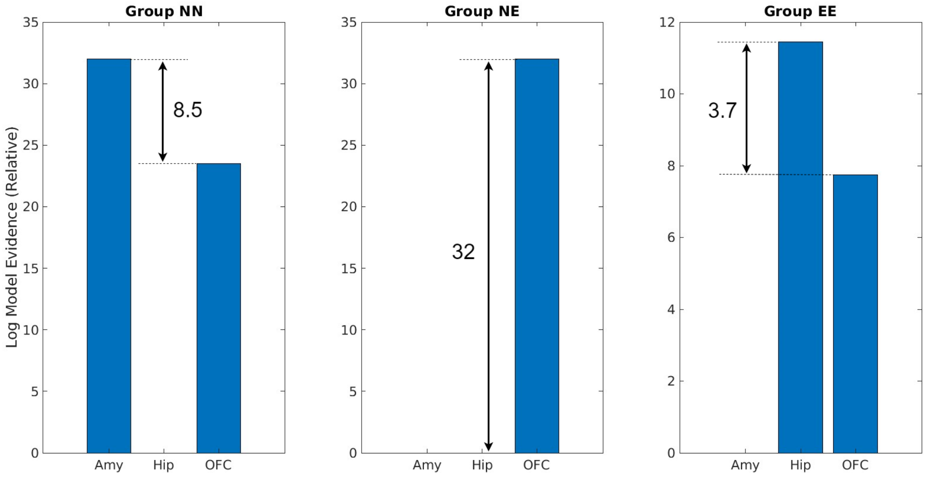


Figure 4-1: Results of BMS for Stochastic DCM. The bars depict the relative log model evidence of different model structures.

model structure has the best fit, it is useful to compare the coefficients of determination, also called R-squared values, of all the models, not only the one that is found to be optimal by BMS. Table 4-3 presents the R-squared values calculated separately for each node (Amy, Hip, OFC) in each of the 3 model structures (input into Amy/ input into Hip/ input into OFC) for every group (NN, NE, EE). The higher the R-squared value, the better the model fits the data. Several interesting insights can be drawn from analyzing the table. Firstly, when considering only the models that were found optimal by BMS, the R-squared values are the highest in group NN. This may suggest that emotional images might lead to more intricate and less predictable patterns in neural activity, making the signal harder to explain with a dynamical model.

Group NN			
	R^2_{Amy}	R^2_{Hip}	R^2_{OFC}
Model 1: Input to Amy (BMS optimal)	0.6493	0.7626	0.3101
Model 2: Input to Hip	0.4631	0.6996	0.0277
Model 3: Input to OFC	0.4632	0.6996	0.0278
Group NE			
	R^2_{Amy}	R^2_{Hip}	R^2_{OFC}
Model 1: Input to Amy	0.2719	0.0304	-0.5625
Model 2: Input to Hip	0.2717	0.0305	-0.5625
Model 3: Input to OFC (BMS optimal)	0.3890	0.2011	-0.0344
Group EE			
	R^2_{Amy}	R^2_{Hip}	R^2_{OFC}
Model 1: Input to Amy	0.4730	0.5534	-0.5530
Model 2: Input to Hip (BMS optimal)	0.4802	0.5599	-0.6269
Model 3: Input to OFC	0.4730	0.5535	-0.5518

Table 4-3: R-squared values in each node (Amy, Hip, OFC) for all average Stochastic DCM models.

Secondly, the R-squared value in the OFC for all the model structures in groups NE and EE is negative, which means that in this node, the model does not explain any of the variance and it fits worse than a horizontal line. This may be because the time series used to estimate the group-optimal average models during BPA contain different numbers of significant voxels, as seen in Table 4-4. Groups NE and EE contain significantly fewer voxels in the OFC, indicating either that this region is not as relevant in the process as previously assumed or that the statistical significance threshold has been set too high and important voxels have been omitted. Showing a lower activity does not necessarily mean that these voxels are less important in the network because they may still exert important influences on the emotional memory circuit, as evidenced by the analysis of coupling strengths in Section 4-2-2. However, it is also likely that the OFC is simply not as crucially involved in the process as the Amy and Hip so it is strongly recommended to reevaluate the node choice for future research in this direction.

	Number of significant voxels in BPA		
	Group NN	**Group NE**	**Group EE**
Amy	177	729	217
Hip	347	212	421
OFC	509	69	7

Table 4-4: Number of significant voxels in the Volumes of Interest (VOIs) used to estimate group-optimal average models during BPA for all groups.

Finally, it can be observed that in groups NN and NE, the R-squared values are the highest for the model that was found to be optimal by BMS. It means that the R-squared and the BMS analyses provide consistent results for these two groups. Conversely, the R-squared analysis for group EE does not reveal a clear winner, as all models have similar values for each node. Similarly, BMS for this group also showed the least confidence that any of the models is significantly better than the others. This again points to the conclusion that group EE experiences the most interference and noise.

Analysis of model generalizability

The R-squared metric explains how close the data is to the fitted regression line (the model). Since the analysis reveals values smaller than one, all models exhibit some level of underfitting. The next important aspect to assess is the level of overfitting/generalizability, i.e., how well the models perform on unseen data. In the case of generative models, this can be assessed by simulating how the model responds to the inputs from a dataset that was not part of the training dataset. In the case of DCM, neuroimaging data is challenging to obtain so the entire dataset is used for training. Instead, cross-validation may be performed (see Section 2-3-2 for details) by splitting the dataset into n sections and using n-1 of those sections for training and the remaining section for validation. The procedure should be repeated n times, each time using a different section for validation.

Before performing a cross-validation study, it is also interesting to check how the average group-optimal models respond to the inputs of individual subjects from the corresponding group. To clarify, the average group-optimal models have so far been compared to the average signal across a given group. It is thus important to evaluate whether the models represent the individual subjects' dynamics well, which also reflects the level of generalizability.

Figure 4-2 depicts an example of how the optimal model of group NN responds to the external stimulation of one of the subjects from this group. Unfortunately, it can be seen that the generated response fits the data very poorly. In fact, all the optimal models perform equally badly when simulating responses to the inputs of all the subjects from the corresponding group. Since the average models fail to generate correct signals to data of individual subjects, they are expected to perform even worse for unseen data. For this reason, a cross-validation study is not been carried out.

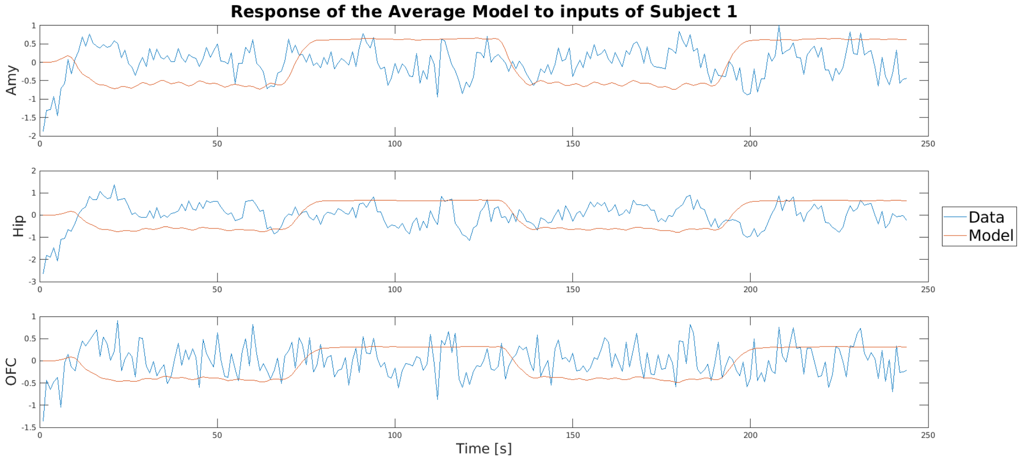


Figure 4-2: Example of how the average optimal model (here, of group NN) responds to the external stimulation of a subject that belongs to the group (here, subject 1 from group NN).

The poor level of generalizability suggests that the DCM framework is not capable of identifying models that generalize well. Such generalizability analysis does not appear to be a common practice in DCM - the literature survey has revealed a lack of studies employing this approach. DCM is mostly used to test hypotheses about average network connectivity, rather than to find one model that can simulate well the neural activity of several people. Especially when dealing with emotional stimuli, people's responses to the same material may differ considerably so there might not exist one truly optimal model.

4-2-2 Estimated parameters

The parameters of the optimal models are obtained using Bayesian Parameter Averaging (BPA), which estimates average parameters based on the data of all subjects within a group. The three groups (NN, NE, EE) have different model structures and may also have very different parameters constituting the A, B, and C matrices. The estimated parameters of all matrices for all groups are shown in the Appendix in Table A-1. DCM analysis involves also estimating the values of the hemodynamic parameters (see Section 2-2-1) but these are of lesser interest. This is because they only serve to fully model the Blood Oxygenation Level-Dependent (BOLD) signal but do not constitute the network model of neuronal activity between the brain regions. In this project, as in most DCM studies, the focus is on analyzing the effective connectivity between the chosen brain regions, expressed by the neuronal equation involving the A, B, and C matrices. To understand the meaning of these estimates, the remainder of this section analyzes their respective values.

Coupling matrix A

Biologically speaking, A expresses the inherent connectivity of the network in the absence of external inputs so it should be roughly similar for all subjects across groups. To analyze how similar the A matrices are in all groups, the possible maximum and minimum values of each matrix entry are summarized in the form of ranges that reflect the possible values across the 3 groups. The average parameter values are obtained according to

$$\operatorname{avg}(a_{i,j}) = \frac{\max a_{i,j} + \min a_{i,j}}{2}, \tag{4-4}$$

and the variance values are defined as

$$\operatorname{var}(a_{i,j}) = \frac{\max a_{i,j} - \min a_{i,j}}{2}. \tag{4-5}$$

The parameter ranges of all A elements are depicted in Figure 4-3, where line thickness symbolizes the relative connection strength, and the colors refer to the sign of the connection (green - positive sign, red - negative sign).

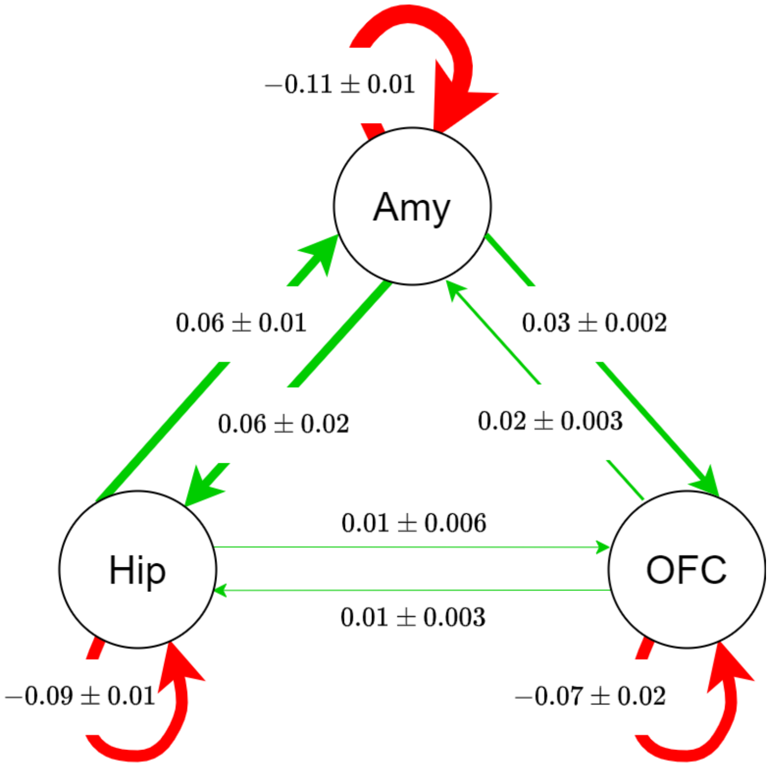


Figure 4-3: Ranges of A matrix values across the groups (NN, NE, EE).

The possible values of each parameter are mostly within $10 - 30\%$ of the averages of their corresponding ranges. While these differences are not negligible and the A matrices are certainly not the same for all groups, the orders of magnitude of all connections are preserved. First of all, it can be observed that the connections between two nodes have similar magnitudes in both directions but there are considerable differences in the connectivity of all pairs of nodes. The strongest between-region connections are observed between the Amy and the Hip, while the weakest connections are found between the Hip and the OFC. Furthermore, the self-connections are the largest in magnitude, which indicates a strong self-inhibition of the nodes, ensuring that the neuronal activity remains bounded.

Coupling matrix B

Perhaps one of the key takeaways from DCM lies in the identification of the B matrix parameters. These indicate how psychological factors affect brain connectivity. This is naturally an important question in this project as the models are developed to identify how different emotional values of images affect neuronal activity. To analyze this, let us consider the parameters of the A' matrix, defined as

$$A' = A + \sum u_i B_i = A + u_1 B_1 + u_2 B_2. \tag{4-6}$$

This augmented A matrix expresses the net product of the inherent network connectivity and the influences the external inputs exert on the connections. As in most DCM studies, the input signal in this thesis has discrete values. The inputs correspond to the onsets of two images constituting associations, where each image is shown at a different time. Consequently, there are 3 possible values of the input vector:

- no image is present $\rightarrow u = [0 \;\; 0]^\top$,
- 1st image is present (image "B" of a triplet) $\rightarrow u = [1 \;\; 0]^\top$,
- 2nd image is present (image "C" of a triplet) $\rightarrow u = [0 \;\; 1]^\top$.

The non-zero input vectors can be plugged into Equation 4-6 to create two distinct adjacency matrices A'. As a result, the model can be viewed as a hybrid system that switches every 2.5s (the length of viewing an image and the length of the break between consecutive images). The parameters of adjacency matrices A, A'_1, and A'_2 are presented in the Appendix in Table A-2. As previously discussed in this section, the A matrix (connectivity when no images are shown) is roughly similar across all experimental groups. Conversely, considerable differences are observed in the A' matrices when experimental stimulation is present. The relative changes between the A' matrices and the A matrix are depicted in Figures 4-4, 4-5, and 4-6, where the line thickness indicates the approximate relative connection strength and the colors signify the sign of the connection (green - positive sign, red - negative sign).

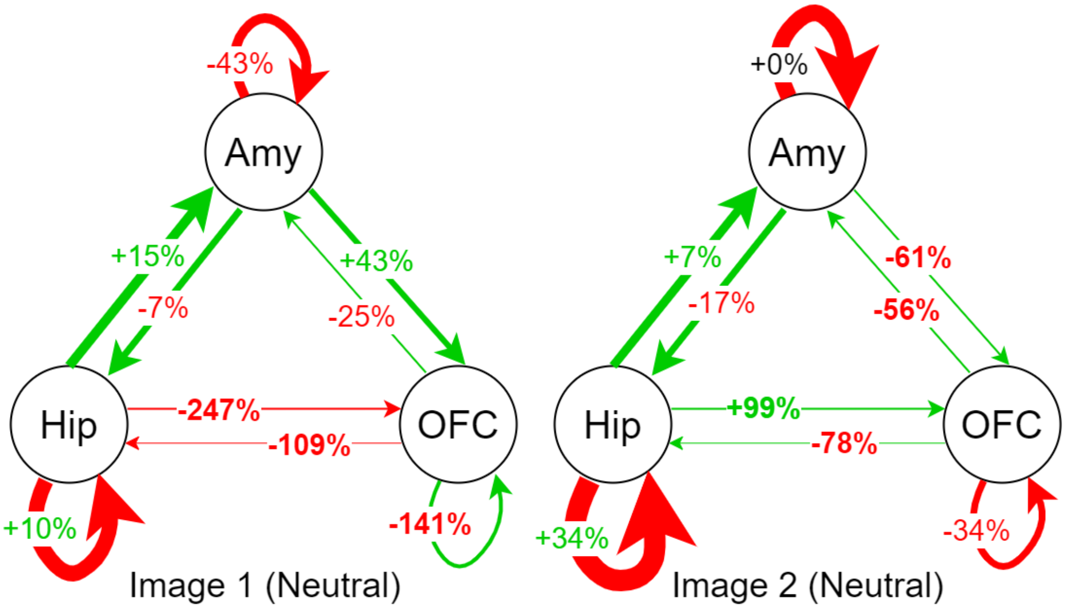


Figure 4-4: Changes in A' caused by the two inputs in the Stochastic DCM in group NN. The percentages represent the relative change in the absolute value of the connection compared to the A matrix.

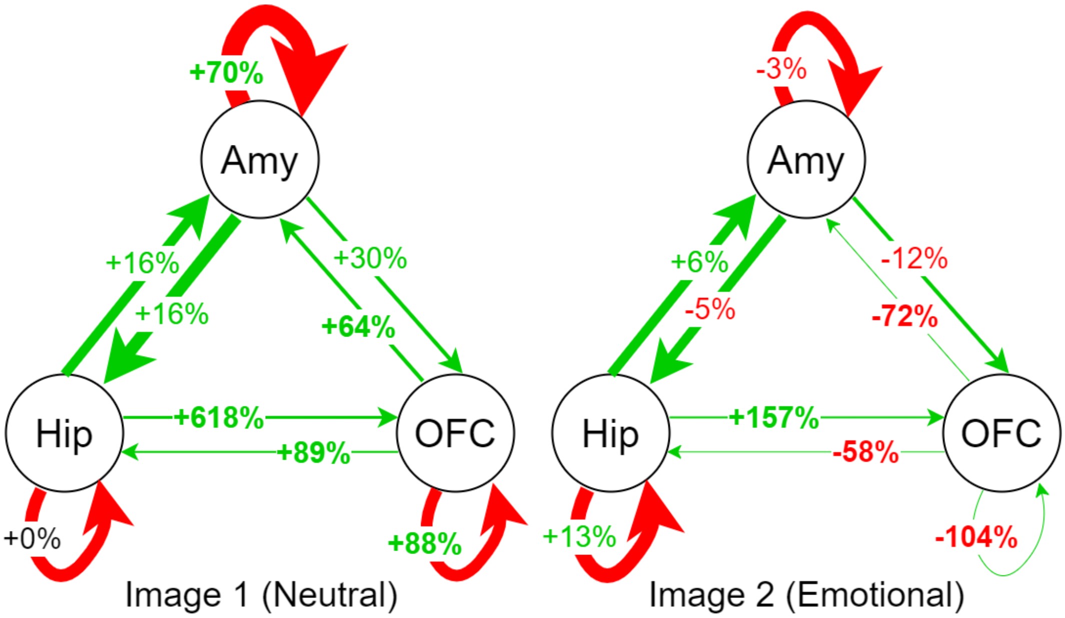


Figure 4-5: Changes in A' caused by the two inputs in the Stochastic DCM in group NE. The percentages represent the relative change in the absolute value of the connection compared to the A matrix.

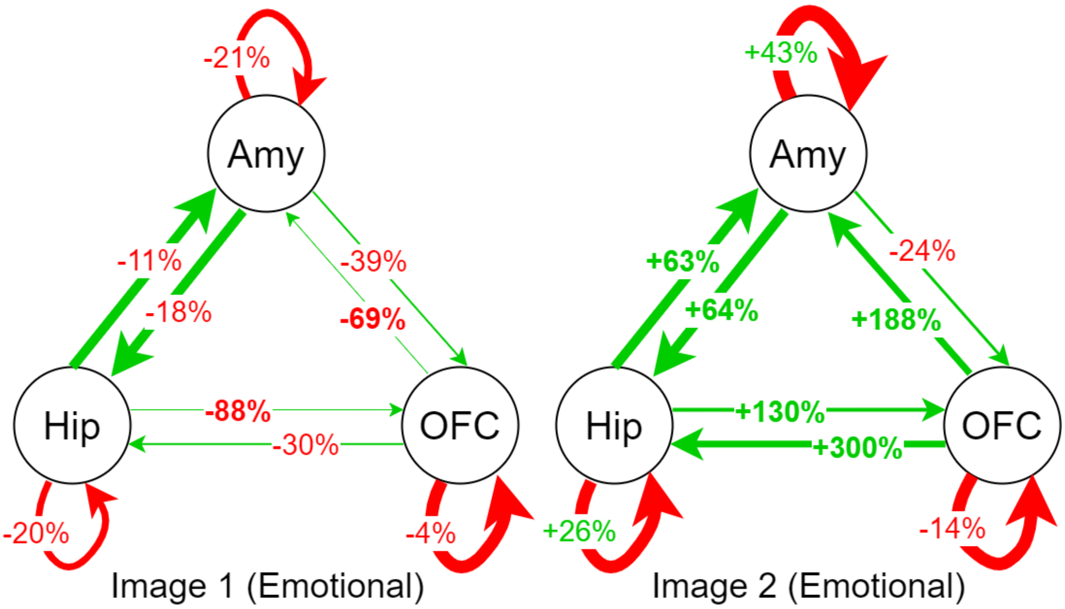


Figure 4-6: Changes in A' caused by the two inputs in the Stochastic DCM in group EE. The percentages represent the relative change in the absolute value of the connection compared to the A matrix.

In most of the figures, it can be observed that, just like in the A matrix, the strongest connections are observed between the Amy and the Hip, while the weakest connections are between the Hip and the OFC. Furthermore, the relative changes between the A' and A matrices are typically the lowest between the Amy and the Hip, varying by around $\pm(5\% - 20\%)$ compared to the A matrix. This suggests that the connectivity between these two nodes is the most robust to external stimulation. On the other hand, the connection strengths between the Hip and the OFC oscillate the most, varying by around $\pm(100\% - 600\%)$ compared to the A matrix. Interestingly, the connectivity between the Hip and the OFC is the most affected by emotions, suggesting a key role of this connection in differently mediating memory encoding depending on the emotional value of this stimulus.

Furthermore, when comparing the figures for all groups, it is evident that the same type of images (neutral or emotional) affect the groups very differently. For example, the first neutral images in group NN cause a decrease in the bidirectional Hip-OFC connectivity and in the connections Amy → Hip and OFC → Amy. However, the first neutral images in group NE increase all of those connections, including a significant increase of around 600% in the connection strength Hip → OFC. This result may seem biologically implausible because, at this stage, the subjects in both groups are only seeing neutral images so their current brain connectivity should be similar. This train of thought is however incorrect because all images are shown twice and the identified models reflect the average connectivity observed during the first and the second time an image is shown. The experimental paradigm is discussed in detail in Section 3-1-1 but in brief, the experiment involved multiple triplets consisting of a cue "A" and two emotionally valenced images "B" and "C". The stimuli were presented such that cue "A" was always accompanied by only one of the images. The cue served to indicate that the two corresponding images "B" and "C" form an association. First, all the "A" and "B" pairs were shown, then all the "A" and "C" pairs, and then, again, all the "A" and "B" pairs, and all the "A" and "C" pairs. This means that when the first image of a triplet is shown for the second time, both images from this triplet have been already shown so an association has been formed, affecting the connectivity.

Finally, it is interesting to analyze which group has the strongest between-region connections on average. This can tell us how strongly the nodes influence each other and how well they exchange information. These average strength values are obtained by averaging over all the off-diagonal entries of the A matrix and the two A' matrices for each group. These mean values are

- 0.0263 for group NN,
- 0.0368 for group NE,
- 0.0327 for group EE.

Evidently, the connections are the strongest in the NE group and the weakest in the NN group. This result is consistent with the findings of [1] (the dataset comes from this study) and with related studies of emotional memory encoding [46, 14] (see Section 1-3), which have also found the strongest between-region connections when emotional stimuli were present.

4-2-3 Stability analysis

It would not be biologically possible for neuronal activity to exponentially diverge to infinity. This means that dynamical computational models of brain activity should be stable. The seminal DCM paper [43] states that the framework ensures the model is "stable" but it is not clear which notion of stability this refers to. The paper explains that the priors on the neuronal parameters have been designed in a way that should ensure the eigenvalues of the A matrix have negative real parts, making the model "stable".

The authors might have been referring to the exponential stability of the unforced system, which is ensured when the A matrix is Hurwitz. This notion of stability is however not very meaningful when modeling neuronal activity in a small brain network. When reducing the immense complexity of the brain to a network with a few nodes, the external input u models not only the external stimulation (here, viewing images) but also the signal from the rest of the brain. While external stimulation can be approximately zero, the brain is always active so the modeled network will always receive some non-zero input. Therefore, the authors of the seminal DCM paper might have been referring to some form of Input-to-State Stability (ISS), which *"roughly states that no matter what is the initial state, if the inputs are uniformly small, then the state must eventually be small"* [98].

The ISS notion of stability may be difficult to prove for real-life systems due to the need to find several bounding functions. There is however a weaker version of ISS, called Integral Input-to-State Stability (IISS), which can be directly proven for bilinear systems. In brief, a bilinear system is said to be IISS if the eigenvalues of the A matrix have negative real parts. This is exactly the requirement that the seminal DCM paper mentions so it is possible that this is the stability notion that the framework has been designed for. The IISS property has been introduced by [98] and is formally defined as follows.

Definition 1 ([98]). *Consider the controlled system $\dot{x} = f(x, u)$, where $f : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}^n$ is continuous, and locally Lipschitz on x for bounded u, and inputs $u(\cdot) : [0, \infty) \to \mathbb{R}^m$ are assumed to be locally essentially bounded.*

The system is Integral Input-to-State Stable if there exist $\alpha, \gamma \in \mathscr{K}_\infty$ and $\beta \in \mathscr{K}\mathscr{L}$ so that the following estimate holds for all initial states $\xi \in \mathbb{R}^n$ and all inputs $u(\cdot)$:

$$\begin{gathered}\alpha(|x(t)|) \leqslant \beta(|\xi|, t) + \int_0^t \gamma(|u(s)|)\mathrm{d}s \\ \textit{for all } t \geqslant 0.\end{gathered} \tag{4-7}$$

If the system satisfies Equation 4-7, then, for any control u such that $\int_0^\infty \gamma(|u(s)|)\mathrm{d}s < \infty$, and for any initial state ξ, it holds for the corresponding trajectory that $x(t) \to 0$ as $t \to \infty$.

Furthermore, bilinear systems of the form $\dot{x} = Ax + \sum u_i B^i x + Cu$ are IISS if A is a Hurwitz matrix, i.e., all eigenvalues have a negative real part [98].

While the seminal DCM paper appears to claim that this stability is always ensured in the estimated models, most of the optimal models identified in this project using Classical, Nonlinear, and Two-State DCM have some eigenvalues with positive real parts. This may be because these DCM variations did not result in models that match the data well and in turn, they do not represent a true neuronal system. Nevertheless, DCM users should be careful in this matter and verify the stability of the identified model to ensure that it is biologically plausible.

In the case of the Stochastic DCM model, the eigenvalues of matrix A corresponding to group NN (e_{NN}), group NE (e_{NE}), and group EE (e_{EE}) all have negative real parts given in Equation 4-8. Therefore, all the identified Stochastic models are IISS, which contributes to their biological plausibility.

$$e_{\mathrm{NN}} = \begin{bmatrix} -0.1593 \\ -0.0696 \\ -0.0252 \end{bmatrix}, \quad e_{\mathrm{NE}} = \begin{bmatrix} -0.1847 \\ -0.0161 \\ -0.0623 \end{bmatrix}, \quad e_{\mathrm{EE}} = \begin{bmatrix} -0.0241 \\ -0.1424 \\ -0.1033 \end{bmatrix}. \tag{4-8}$$

Analysis of network coordination

The stability properties of the identified models are also related to the state coordination. This phenomenon refers to the coordination and alignment of processes within the network to ensure proper functioning and efficient communication [99, 100]. One way of studying the state coordination is by considering the eigenvalues of the graph combinatorial Laplacian L, defined as

$$L = D_{\mathrm{in}} - A, \tag{4-9}$$

where A is the adjacency matrix of the network, and D_{in} is the diagonal node degree matrix with entries equal to the sum of weights of connections going into each node [100]. Matrix L is a symmetric positive semi-definite matrix and all its eigenvalues are real and non-negative.

Group NN

	λ_2 of Amy-Hip	λ_2 of Amy-OFC	λ_2 of Hip-OFC	λ_2 of entire graph
A	0.1046	0.0520	0.0225	0.0536
A'_1	0.1091	0.0582	0.0134	0.0283
A'_2	0.0999	0.0214	0.0195	0.0335

Group NE

	λ_2 of Amy-Hip	λ_2 of Amy-OFC	λ_2 of Hip-OFC	λ_2 of entire graph
A	0.1397	0.0502	0.0118	0.0471
A'_1	0.1642	0.0722	0.0395	0.0872
A'_2	0.1395	0.0325	0.0120	0.0399

Group EE

	λ_2 of Amy-Hip	λ_2 of Amy-OFC	λ_2 of Hip-OFC	λ_2 of entire graph
A	0.0912	0.0437	0.0293	0.0563
A'_1	0.0780	0.0214	0.0120	0.0264
A'_2	0.1494	0.0707	0.0920	0.1079

Table 4-5: Algebraic connectivity λ_2 in graphs and subgraphs for all groups when no input is presented (adjacency matrix A), when input 1, i.e., image "B" in a triplet is presented (adjacency matrix A'_1), and when input 2, i.e., image "C" in a triplet is presented (adjacency matrix A'_2).

Since all rows of L sum up to zero, the lowest eigenvalue is always zero. Moreover, the first non-zero eigenvalue λ_2 is called the "algebraic connectivity" because it reflects the connectivity properties of the graph. A graph with a higher algebraic connectivity tends to have better connected components, making it more robust against disconnections or failures.

The identified models may be viewed as hybrid systems that switch between adjacency matrices defined as $A' = A + u_1 B_1 + u_2 B_2$, leading to 3 different matrices: A (no input is presented), A'_1 (input 1, i.e., image "B" in a triplet, presented), and A'_2 (input 2, i.e., image "C" in a triplet, presented). It has been shown that the states of switching systems converge to similar values (coordinate) with a speed equal to the algebraic connectivity [101]. This means that higher λ_2 values lead to a better state convergence. The algebraic connectivity values have been calculated for the 3 possible adjacency matrices of the full graphs (all nodes included) and subgraphs (considering separately the coupling of Amy-Hip, Amy-OFC, and Hip-OFC) and are presented in Table 4-5. It can be observed that λ_2 of Amy-Hip is the highest compared to other subgraphs in all groups, especially in group NE, indicating the highest state coordination and convergence. On the other hand, λ_2 of Hip-OFC is the lowest compared to other subgraphs in all groups. The connectivity between these nodes has also been found to have the lowest strength and be the most prone to changing upon varying inputs. Overall, these results suggest that Hip-OFC are the least coordinated and robust, compared to other node pairs.

Network's error dynamics: IISS stability and bounds

State coordination is also related to the dynamics of the state error defined by

$$z_e = \begin{bmatrix} z_{Amy} - z_{Hip} \\ z_{Amy} - z_{OFC} \\ z_{Hip} - z_{OFC} \end{bmatrix}. \tag{4-10}$$

A low error between the nodes implies that they are functioning coherently and efficiently exchanging and integrating information. On the other hand, a high error points to lower coordination between the nodes, potentially leading to poorer information flow and possible cognitive impairments. The error dynamics can be derived analogously to the state dynamics given by

$$\dot{z}(\tau) = Az(\tau) + \sum B^i u_i(\tau) z(\tau) + Cu(\tau). \tag{4-11}$$

Using the notation, where a_{ij} is the element in the i-th row and j-th column of matrix A, let us define the state error adjacency matrix

$$A_e = \begin{bmatrix} a_{11} - a_{21} & a_{12} - a_{22} & a_{13} - a_{23} \\ a_{11} - a_{31} & a_{12} - a_{32} & a_{13} - a_{33} \\ a_{21} - a_{31} & a_{22} - a_{32} & a_{23} - a_{33} \end{bmatrix}. \tag{4-12}$$

The same operation may be repeated for the remaining state matrices to arrive at B_{e1}, B_{e2}, and C_e. Now, the state error dynamics can be defined as

$$\dot{z}_e(\tau) = A_e z_e(\tau) + \sum B_e^i u_i(\tau) z_e(\tau) + C_e u(\tau). \tag{4-13}$$

Lemma 1. *The error dynamics corresponding to the network described in Equation 4-13 are a bilinear system and are IISS.*

The eigenvalues of A_e of group NN ($e_{\text{e,NN}}$), group NE ($e_{\text{e,NE}}$), and group EE ($e_{\text{e,EE}}$) all have negative real parts given in Equation 4-14. Therefore, the error dynamics are IISS, meaning that they will eventually converge to zero for sufficiently small inputs.

$$e_{\text{NN}} = \begin{bmatrix} -1.0510 \cdot 10^{-17} \\ -0.0245 + 0.1230i \\ -0.0245 - 0.1230i \end{bmatrix}, \quad e_{\text{NE}} = \begin{bmatrix} -7.8838 \cdot 10^{-17} \\ -0.0483 + 0.0960i \\ -0.0483 - 0.0960i \end{bmatrix}, \quad e_{\text{EE}} = \begin{bmatrix} -1.3878 \cdot 10^{-17} \\ -0.1881 \\ -0.1138 \end{bmatrix}. \tag{4-14}$$

To analyze how large the error can be and thus, how synchronized the nodes are, it is possible to derive bounds for the state error. As this bound depends on the state matrices estimated through DCM, the bound will differ for every group and indicate which group exhibits better information flow related to improved memory encoding.

Proposition 1. *The bounds for the network error dynamics given by Equation 4-13 obey*

$$\|z_e(t)\| \leq \left(-\frac{M^3}{2\omega}\right)^{1/2} \left(\int_0^\infty \left\|\sum B_e^i u_i(\tau) z_e(\tau) + C_e u(\tau)\right\|^2 d\tau\right)^{1/2}, \tag{4-15}$$

where $M \geq 1$ and $\omega > \max\{\text{Re}(\lambda) : \lambda$ eigenvalue of $A_e\}$ are constants that satisfy Lemma 2.

Lemma 2 ([102]). *Let $A \in \mathbb{C}^{n\times n}$ be a matrix. Then for every $\omega > \max\{\mathrm{Re}(\lambda) : \lambda$ eigenvalue of $A\}$, there exists a constant $M \geq 1$ such that*

$$\left\| e^{At} \right\| \leq M e^{\omega t} \quad \forall \ t \geq 0. \tag{4-16}$$

Proof. The proof is based on the results in [102]. The bound derivation begins by pre-multiplying both sides of Equation 4-13 with $e^{-A_e\tau}$, observing that the left-hand side is equivalent to a time derivative of $e^{-A_e\tau}z_e(\tau)$, and integration both sides, resulting in

$$\begin{aligned}
e^{-A_e\tau}\dot{z}_e(\tau) &= e^{-A_e\tau}A_e z_e(\tau) + e^{-A_e\tau}\sum B_e^i u_i(\tau)z_e(\tau) + e^{-A_e\tau}C_e u(\tau)\\
e^{-A_e\tau}\dot{z}_e(\tau) - e^{-A_e\tau}A_e z_e(\tau) &= e^{-A_e\tau}\sum B_e^i u_i(\tau)z_e(\tau) + e^{-A_e\tau}C_e u(\tau)\\
\frac{d}{d\tau}(e^{-A_e\tau}z_e(\tau)) &= e^{-A_e\tau}\sum B_e^i u_i(\tau)z_e(\tau) + e^{-A_e\tau}C_e u(\tau)\\
\int_0^t \frac{d}{d\tau}(e^{-A_e\tau}z_e(\tau))\,d\tau &= \int_0^t \left(e^{-A_e\tau}\left(\sum B_e^i u_i(\tau)z_e(\tau) + C_e u(\tau)\right)\right)d\tau\\
e^{-A_e t}z_e(t) - z_{e0} &= \int_0^t \left(e^{-A_e\tau}\left(\sum B_e^i u_i(\tau)z_e(\tau) + C_e u(\tau)\right)\right)d\tau\\
z_e(t) &= e^{A_e t}z_{e0} + \int_0^t \left(e^{A_e(t-\tau)}\left(\sum B_e^i u_i(\tau)z_e(\tau) + C_e u(\tau)\right)\right)d\tau
\end{aligned} \tag{4-17}$$

This expression may be now turned into an inequality by taking the norms of both sides and using the fact that

$$\|a+b\| \leq \|a\| + \|b\|, \tag{4-18}$$

resulting in the following bound for the state error

$$\begin{aligned}
\|z_e(t)\| &= \left\| e^{A_e t}z_{e0} + \int_0^t \left(e^{A_e(t-\tau)}\left(\sum B_e^i u_i(\tau)z_e(\tau) + C_e u(\tau)\right)\right)d\tau \right\|\\
\|z_e(t)\| &\leq \|e^{A_e t}\|\|z_{e0}\| + \left\| \int_0^t \left(e^{A_e(t-\tau)}\left(\sum B_e^i u_i(\tau)z_e(\tau) + C_e u(\tau)\right)\right)d\tau \right\|\\
\|z_e(t)\| &\leq \|e^{A_e t}\| \left(\|z_{e0}\| + \left\| \int_0^t \left(e^{-A_e\tau}\left(\sum B_e^i u_i(\tau)z_e(\tau) + C_e u(\tau)\right)\right)d\tau \right\| \right).
\end{aligned} \tag{4-19}$$

By applying Lemma 2 on all exponentials, the expression becomes

$$\|z_e(t)\| \leq M e^{\omega t}\left(\|z_{e0}\| + \left\| \int_0^t \left(M e^{-\omega\tau}\left(\sum B_e^i u_i(\tau)z_e(\tau) + C_e u(\tau)\right)\right)d\tau \right\| \right). \tag{4-20}$$

Now, let us use the Cauchy-Schwartz inequality [103] for a product of two arbitrary integrable functions $f(x)$ and $g(x)$, given by

$$\left\| \int f(x)g(x)dx \right\| \leq \left(\int \|f(x)\|^2 dx \right)^{1/2} \left(\int \|g(x)\|^2 dx \right)^{1/2}, \tag{4-21}$$

to separate terms in the integral expression, resulting in

$$\|z_e(t)\| \leq Me^{\omega t}\left(\|z_{e0}\| + \left(\int_0^t \|Me^{-\omega\tau}\|^2\, d\tau\right)^{1/2}\left(\int_0^t \left\|\sum B_e^i u_i(\tau)z_e(\tau) + C_e u(\tau)\right\|^2 d\tau\right)^{1/2}\right). \tag{4-22}$$

Let us now evaluate one of these integrals,

$$\left(\int_0^t \|Me^{-\omega\tau}\|^2\, d\tau\right)^{1/2} = \left(\int_0^t Me^{-2\omega\tau}\, d\tau\right)^{1/2} = \left(\left[-\frac{M}{2\omega}e^{-2\omega\tau}\right]_0^t\right)^{1/2} = \left(\frac{M}{2\omega}\left(1-e^{-2\omega t}\right)\right)^{1/2}. \tag{4-23}$$

Furthermore, let us bring the common denominator of the bound $Me^{\omega t}$ back into the bracket and use the expression derived above, resulting in

$$\|z_e(t)\| \leq Me^{\omega t}\|z_{e0}\| + Me^{\omega t}\left(\frac{M}{2\omega}\left(1-e^{-2\omega t}\right)\right)^{1/2}\left(\int_0^t \left\|\sum B_e^i u_i(\tau)z_e(\tau) + C_e u(\tau)\right\|^2 d\tau\right)^{1/2}. \tag{4-24}$$

Let us further simplify one of the expressions with the exponential functions, to arrive at

$$\begin{aligned} Me^{\omega t}\left(\frac{M}{2\omega}\left(1-e^{-2\omega t}\right)\right)^{1/2} &= \left(Me^{\omega t}\right)^{2\cdot\frac{1}{2}}\left(\frac{M}{2\omega}\left(1-e^{-2\omega t}\right)\right)^{1/2} \\ &= \left(M^2 e^{2\omega t}\frac{M}{2\omega}\left(1-e^{-2\omega t}\right)\right)^{1/2} \\ &= \left(\frac{M^3}{2\omega}\left(e^{2\omega t}-1\right)\right)^{1/2}. \end{aligned} \tag{4-25}$$

Plugging this expression back into the bound, we obtain

$$\|z_e(t)\| \leq Me^{\omega t}\|z_{e0}\| + \left(\frac{M^3}{2\omega}\left(e^{2\omega t}-1\right)\right)^{1/2}\left(\int_0^t \left\|\sum B_e^i u_i(\tau)z_e(\tau) + C_e u(\tau)\right\|^2 d\tau\right)^{1/2}. \tag{4-26}$$

Let us now analyze the two terms constituting this upper bound, starting with $Me^{\omega t}\|z_{e0}\|$. The requirement for ω is that it should be larger than the maximum real part of eigenvalues of A_e. Since A_e is Hurwitz, ω may be chosen to be negative. This means that the first term of the bound exponentially decays to zero as $t \to \infty$. Moving on to the second term, we observe that it is composed of some function of $e^{\omega t}$ and an integral involving the state error matrices, inputs, and state errors. The expression given by Equation 4-25 dominates and defines the rate of increase of the entire term. In the limit, this dominant term reduces to

$$\lim_{t\to\infty}\left(\frac{M^3}{2\omega}\left(e^{2\omega t}-1\right)\right)^{1/2} = \left(-\frac{M^3}{2\omega}\right)^{1/2}. \tag{4-27}$$

Now, putting the two limits together, the limit of the bound as $t \to \infty$ of the state error becomes

$$\|z_e(t)\| \leq \left(-\frac{M^3}{2\omega}\right)^{1/2} \left(\int_0^\infty \left\|\sum B_e^i u_i(\tau) z_e(\tau) + C_e u(\tau)\right\|^2 d\tau\right)^{1/2}. \tag{4-28}$$

$\square$

Before evaluating the limit, the term involving M and ω dominated the dynamics because it involved an exponential function. To compare the state synchronization in different groups, let us then consider this dominant term. While M can be arbitrarily chosen to satisfy Lemma 2, ω should be larger than $\lambda_{\max} = \max\{\mathrm{Re}(\lambda) : \lambda \text{ eigenvalue of } A_e\}$. Looking at the eigenvalues given in Equation 4-14, $\lambda_{\max}$ for groups NN, NE, and EE is

$$\begin{aligned} \lambda_{\max,\mathrm{NN}} &= -1.0510 \cdot 10^{-17}, \\ \lambda_{\max,\mathrm{NE}} &= -7.8838 \cdot 10^{-17}, \\ \lambda_{\max,\mathrm{EE}} &= -1.3878 \cdot 10^{-17}. \end{aligned}$$

It can be observed that this eigenvalue has the largest absolute value for group NE and the smallest absolute value for group NN. The ω value may now be arbitrarily chosen to be slightly larger than these eigenvalues, as long as it remains negative and the inequality in LEmma 2 holds. Since ω is negative, the minuses in Equation 4-28 cancel out, leaving an expression depending on the inverse of the absolute value of ω. As group NE has the largest absolute eigenvalue, this will result in the lowest dominant term in the bound, compared to other groups. As a result, the state error has the strictest bounds for group NE. Therefore, this group is expected to achieve the best synchronization, improving information flow and memory encoding.

Let us now verify this result by looking at the state error evolution for the group-optimal DCMs. These state error trajectories are depicted in Figure 4-7 for group NN, in Figure 4-8 for group NE, and in Figure 4-9 for group EE. In every group, it has been estimated that the input enters the network through a different node (NN: input enters through Amy, NE: input enters through OFC, EE: input enters through Hip). The states of nodes that receive the input (e.g., z_{Amy} in group NN) assume higher magnitudes than the other states. Furthermore, during the experiment, multiple inputs and breaks occur so the adjacency matrix rapidly changes between A, A_1', and A_2'. This is the reason for the zigzag pattern observed in most trajectories. Consistent with the bounds derived earlier, the errors of group NE are the lowest in magnitude. In particular, the Amy and Hip exhibit the highest coordination of all node pairs across all groups. This finding is in line with the analysis of the eigenvalues of the graph Laplacian in Section 4-2-3, where it has been shown that nodes Amy and Hip synchronize the fastest, compared to other node pairs.

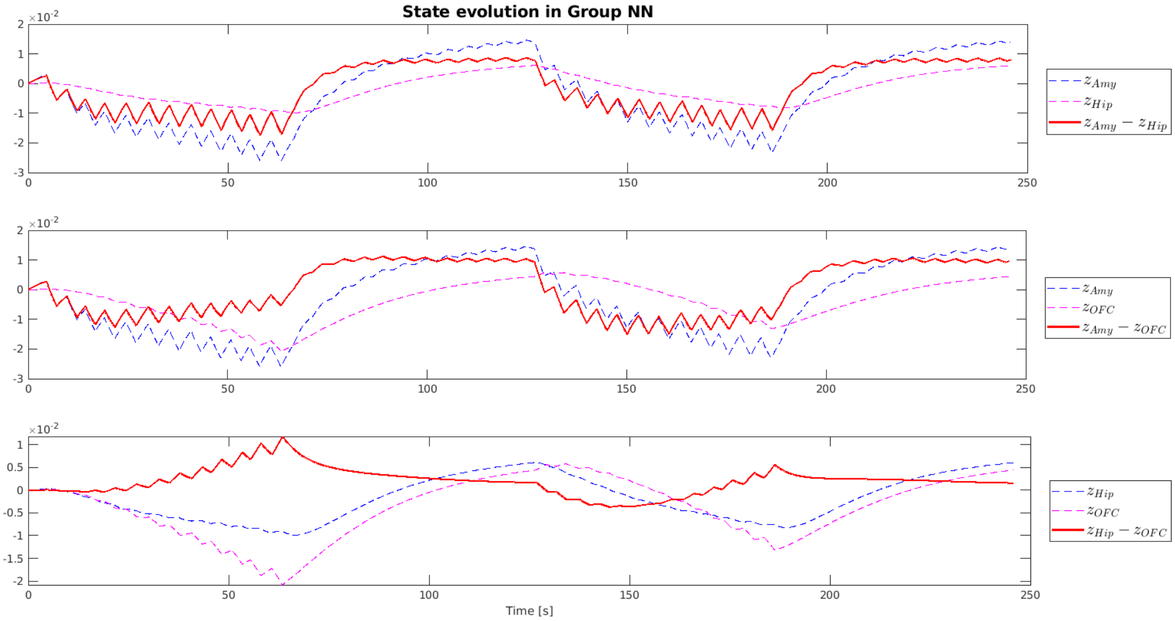


Figure 4-7: State and state-error evolution of the NN group in the Stochastic DCM model. The states are denoted by z_{Amy}, z_{Hip}, and z_{OFC}, whereas the state-errors are the differences between the states.

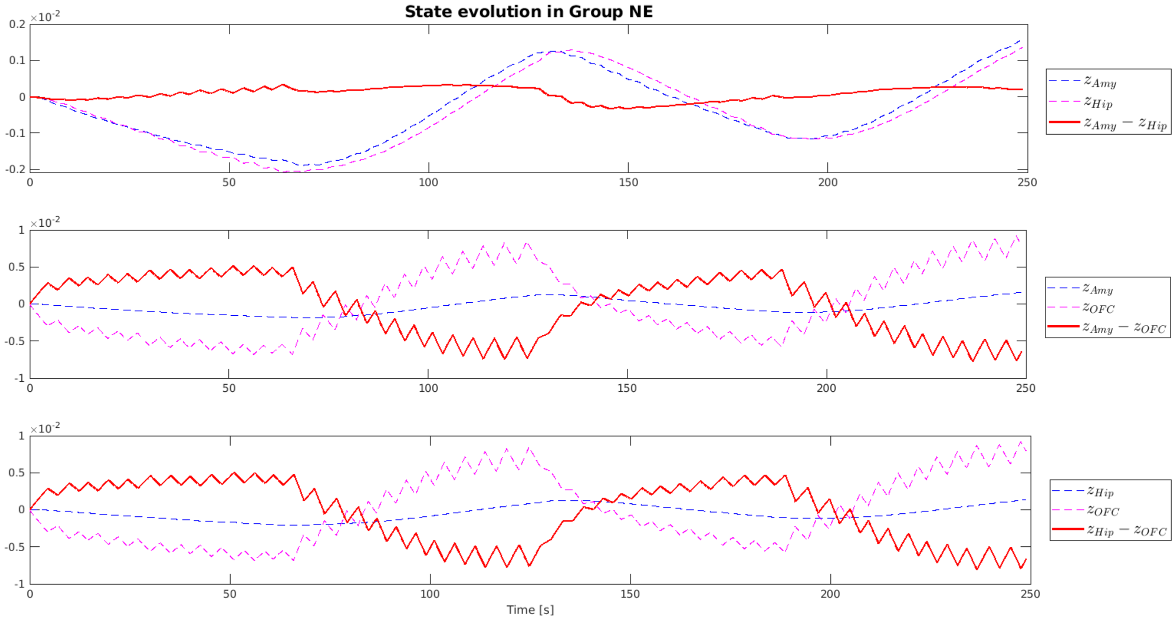


Figure 4-8: State and state-error evolution of the NE group in the Stochastic DCM model. The states are denoted by z_{Amy}, z_{Hip}, and z_{OFC}, whereas the state-errors are the differences between the states.

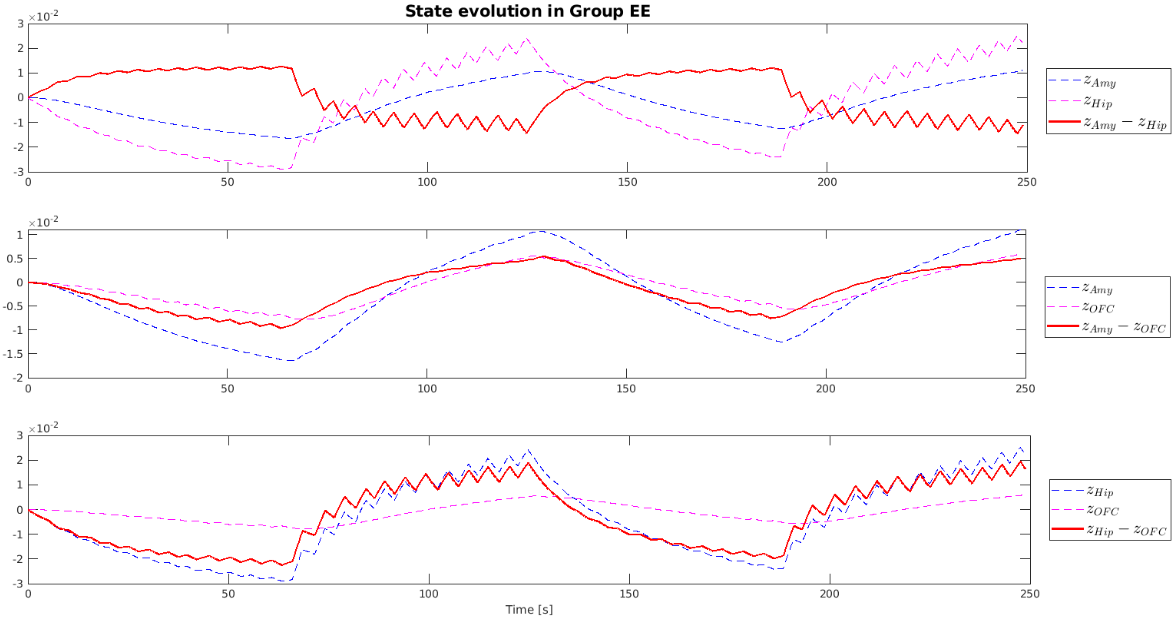


Figure 4-9: State and state-error evolution of the EE group in the Stochastic DCM model. The states are denoted by z_{Amy}, z_{Hip}, and z_{OFC}, whereas the state-errors are the differences between the states.

Chapter 5

Conclusions

The aim of this thesis has been to develop a dynamical model of the neural activity governing the process of encoding emotional associative memories. This cognitive process is responsible for remembering the relationships between stimuli from the environment and assigning an emotional value to them. By understanding this phenomenon better, researchers can improve the treatment of mental disorders associated with emotional dysregulation, as well as enhance bio-inspired technologies such as artificial intelligence. However, much remains unknown about emotional associative memory and studies find discrepant results that are highly sensitive to the experimental paradigm, the length and intensity of the elicited emotions, and the examined brain regions. For this reason, developing a dynamical model of the underlying neural activity is a promising approach that may bright insights that cannot be observed otherwise.

In order to be able to draw conclusions about this cognitive process, the modeling framework must provide biologically interpretable results. In particular, a thorough literature study has revealed that the most crucially involved brain regions in emotional associative memory are the Amygdala (Amy), the Hippocampus (Hip), and the Orbitofrontal Cortex (OFC). When selecting the modeling framework, it was thus important to find an approach for modeling the activity in a network of pre-defined brain regions. The most prominent method that satisfies this requirement is Dynamic Causal Modeling (DCM) [43], where network activity is modeled using a bilinear, biologically-inspired differential equation with parameters estimated through Bayesian inference. Furthermore, DCM models how external manipulation (e.g., displayed images or sounds) affects the brain network and gives rise to the measured signal.

To correctly apply this approach, DCM has been rigorously studied and documented in Chapter 2. This part of the report provides a comprehensive overview of the modeling framework, consolidating information from over 50 scientific articles that discuss DCM's mathematical foundations, the system identification algorithm, critical comments raised against the approach, studies of its statistical validity, and advice on implementation. Apart from the classical version of the framework, the chapter introduces three DCM variations that offer attractive extensions to the modeling framework, namely Nonlinear DCM, Two-state DCM, and Stochastic DCM. As all these methods provide different capabilities, it has been decided to apply them all and identify which one provides the best representation of the problem at hand.

The dataset used to develop the models comes from a functional Magnetic Resonance Imaging (fMRI) study, where participants were asked to memorize pairs of images including two emotionally potent images (group Emotional-Emotional (EE)), two neutral images (group Neutral-Neutral (NN)), or one neutral and one emotional image (group Neutral-Emotional (NE)) [1]. To understand how the differently valenced stimuli are encoded, it has been decided to develop a separate DCM for each group. The modeling approach and all the design decisions have been discussed in detail in Chapter 3. In applying the approach in practice, the Statistical Parametric Mapping (SPM12) MATLAB toolbox [94] has been utilized but developing the script has proved to be very challenging due to a lack of detailed documentation. To bridge this gap, one of the contributions of this thesis is the guide to SPM12 DCM scripting presented in Appendix B, which provides explanations of all the necessary steps.

After successfully developing DCM models of all three groups and using four different DCM variations, the findings have been presented and discussed in Chapter 4. The results have shown that the stochastic version of the modeling framework has given far superior results compared to the other methods. The other types of DCM have all resulted in highly underfitting models that generate signals with a large offset from the correct mean value. Many of these models have also been found to have non-Hurwitz A matrices, meaning that they are unstable in the absence of external input, which is biologically implausible as brain dynamics should retain finite values. Conversely, Stochastic DCM has resulted in models that fit the data considerably better, are globally exponentially stable when unforced, and are Integral Input-to-State Stable (IISS), meaning that the states remain finite for sufficiently small inputs.

Furthermore, the structure, parameters, and stability properties of the optimal Stochastic DCM models have been thoroughly analyzed in Section 4-2. As the level of network state coordination is linked to the efficiency of information flow, this property has been meticulously inspected. By analyzing the graph combinatorial Laplacian matrix L, it has been shown that the Amy-Hip pair in group NE achieves state coordination the fastest, compared to other node pairs in all groups. Moreover, the error dynamics have been analyzed, proving that they are also Integral Input-to-State Stable (IISS) and that analytic bounds can be derived that show differences between the groups. Consistent with the previous analysis, group NE has been found to have the narrowest error bounds of all groups, meaning that the differences between the states are expected to be the lowest, indicating high state coordination. These findings corroborate the results in [1], where group NE has been found to exhibit the best memory of the image associations. To further analyze the meaning of the modeling outcome, Section 5-1 discusses why Stochastic DCM has given the best fit, how the models compare to the behavioral results of the study, and what the models say about emotional associative memory.

To conclude, this thesis has successfully answered all the research questions by providing a detailed overview of the modeling framework, identifying that Stochastic DCM is the most suitable variation of DCM, estimating distinct models for groups NN, NE, and EE, and analyzing what the models reveal about emotional associative memory depending on the emotional value of the stimuli. As a result, this thesis has provided the first DCM model of the encoding of emotional associative memory in the network consisting of the Amy, the Hip, and the OFC. There has only been one other DCM study of the same cognitive process but the Hip has not been included in the modeled network [14]. This brain region is believed to be an indispensable part of the emotional memory circuit [28, 7] and has been shown in this thesis to play a key role in responding to changing stimuli. On top of that, the experiment in [14] included pairs of two emotional or two neutral stimuli, unlike this project, where also two differently valenced stimuli were included (group NE). Additionally, the images in this project have been displayed separately and modeled as distinct inputs, proving that images within a pair affect network connectivity very differently from each other. Overall, by including the hippocampus, NE image pairs, and modeling images within pairs as distinct inputs, this thesis is believed to have provided the most comprehensive representation of the causal connectivity governing the encoding of emotional associative memory to date.

5-1 Discussion of the results

The following discussion aims to present reasons supporting the assertion that the stochastic variation of DCM represents the most appropriate approach for modeling the given experimental paradigm. Additionally, the estimated models are compared to the behavioral results of the experiment, demonstrating a high level of consistency and highlighting the additional insights provided by the computational models.

Why is Stochastic DCM the most suitable approach for this experimental paradigm?

Stochastic DCM models the measured signal as a reaction to both external stimulation (here, displayed images) and neuronal noise, which is an integral part of how the brain functions. As a result, the estimated models have the capability to model random state noise, rendering them more robust to model misspecification. On top of that, Stochastic DCM employs a novel system identification method, which has been proven to provide superior results in model structure and parameter inference, as compared to classical DCM [65, 84]. One of the most crucial differences between the algorithms is the use of generalized coordinates in Stochastic DCM. This type of coordinates tracks several time derivatives of each estimated parameter, as opposed to considering only point estimates. Consequently, much more information about the shape of the function is considered, allowing the model to correctly represent vastly more complex signals.

There are three main reasons why Stochastic DCM is believed to have provided superior results in this project. Firstly, DCM requires specifying the model structure (which connections and influences are present in the network) in advance, and typically, researchers define a space with several different structures and compare their performance through Bayesian Model Selection (BMS). In this project, the model space has included a very limited number of different model structures to reduce the computational complexity. This approach is considered a valid simplification because instead of identifying the presence or absence of connections, their relative importance can be reflected by the magnitudes of connection strengths. However, it may, naturally, lead to model misspecification if some of the aspects of model structure play a key role in the network dynamics. As Stochastic DCM employs a superior identification method and includes state noise, the parameter estimates are more precise and model structure misspecification may be compensated for by modeling the discrepancies as noise.

The second possible reason why the stochastic model represents the neural dynamics significantly better than the other DCM variations is related to the nature of the experimental paradigm. Firstly, the images were displayed very briefly (each image shown for 2.5s) and frequently (new images shown every ∼5s), compared to other DCM studies, where signals typically last for several tens of seconds (e.g., [92]). As explained in Section 2-1-1, fMRI estimates neural activity by measuring the Blood Oxygenation Level-Dependent (BOLD) signals. This working principle is based on the fact that when neurons fire, the supply of oxygen to surrounding neurons increases and peaks after 3-6s and then drops to baseline after 12-30s. As a result, the measurements are taken several seconds after the activity itself. When experimental inputs change as often as in this project, their BOLD signals overlap and it becomes increasingly difficult to differentiate between the impacts of the consecutive inputs. This challenge appears to be handled much better by Stochastic DCM thanks to its more efficient system identification algorithm.

Finally, neural noise is believed to increase when the discrepancy between the environment and a person's internal model (prediction) is high [85, 86, 87, 88]. In this experiment, this is the case because images were shown rapidly and elicited emotional responses, which are inherently highly subjective and nuanced. Consequently, neural noise is believed to be specifically high in this experimental paradigm, necessitating the use of a stochastic modeling framework.

What do the models say about emotional associative memory of differently valenced images?

The models in this project have been developed using data from the study performed by Zhu et al. at the Donders Institute for Brain, Cognition and Behaviour [1], where the authors have analyzed the fMRI data and the behavioral results of the memorization task. The overall outcome of the study has shown that emotional information forms stronger and better-remembered associations with neutral information, while it interferes with the integration with other emotional information. This finding is consistent with both the facilitation theory, i.e., emotionally salient stimuli can strengthen related neutral stimuli, as well as the interference theory, i.e., overlapping representations of related emotional memories with equal priority compete with each other and overall suppress their memory integration.

In particular, statistical tests in [1] have revealed that subjects in group NE have exhibited the best integrated memory (both images in an association remembered correctly). Furthermore, subjects in group EE have shown worse non-integrated memory (one of the images in an association remembered correctly) of the first images and better non-integrated memory of the second images, compared to the other groups. This finding indicates that emotional stimuli interfere with their integrated memory of related events. Moving on to the functional connectivity analysis, the hippocampus has been found to serve a key role, and strong coupling with this region has been shown to be an indicator of the successful encoding of integrated memories. Specifically, the experiment has found the strongest coupling between the hippocampus and other relevant regions (including the amygdala) in the NE group, and the weakest in the NN group, with the EE group in the middle.

	Average Hip coupling strength		
	Group NN	Group NE	Group EE
u_1	0.0239	0.0509	0.0225
u_2	0.0299	0.0379	0.0604
Avg	0.0269	0.0444	0.0414

Table 5-1: Average coupling strength over the Hip-Amy and Hip-OFC connections in both directions. The first row (u1) shows the strength when input 1 is presented, the second row shows the strength when input 2 is presented, and the last row (Avg) is the average of the first two values. The colors indicate the highest magnitude (red), medium magnitude (orange), and the lowest magnitude (yellow) in each row.

The models in this thesis support the findings about hippocampal functional connectivity in [1]. The average coupling strengths over A'_1 (adjacency matrix when input 1 is presented) and A'_2 (adjacency matrix when input 2 is presented) between Hip-Amy and Hip-OFC are presented in Table 5-1. The overall coupling (average over both inputs) is the the highest in group NE, with a slightly lower value in group EE, and a significantly lower value in group NN. This is consistent with the study's finding about the level of functional connectivity of the hippocampus. Moreover, when considering the inputs separately, group EE exhibits the most extreme values - it has the lowest coupling of all groups for input 1 and the highest coupling of all groups for input 2. This is in line with the behavioral results of the study that state that group EE has exhibited the worst item memory for images 1 and the best item memory for images 2.

An additional insight that the computational models offer over the data analysis done in [1] can be derived by looking at the stability and network coordination analysis in Section 4-2-3. The analyzed dynamical properties and plots of state-error evolution show that group NE exhibits the highest state coordination and that its Amy-Hip node pair converges to similar values the fastest. As a result, the network model of this group shows the most efficient information flow, which is consistent with the finding that this group has exhibited the best integrated memory. Furthermore, these results indicate an important role of the Amy-Hip connection in successful memory encoding. Conversely, groups NN and EE have shown a lower state coordination, consistent with their worse integrated memory results.

On top of that, Table A-2 shows that in all groups, the connection strengths between Amy-Hip are also the highest with or without external stimulation, compared to other node pairs. Their magnitudes are also the least dependent on the input since most groups exhibit a change of only up to around 15% compared to when no input is present. Conversely, while all groups exhibit the weakest coupling between Hip-OFC, this connection is the most affected by inputs, changing by up to 600% and even flipping the sign from positive to negative in one group. Overall, these results indicate that the amygdala and the hippocampus have the best and most robust to external manipulation information flow in all groups. Whereas the hippocampus and the OFC, show the weakest but most sensitive to stimulation coupling, indicating that this node pair plays a crucial role in governing reactions to the environment.

5-2 Future research

The DCM models developed in this project have answered the research questions and at the same time, revealed many new directions worth exploring in future work. First of all, the representation of the neural underpinnings in the experiment could be refined by improving the selection of brain regions included in the model. In this thesis, the nodes have been preselected based on related studies and the known theory of emotional learning. To address the issue of potentially excluding important brain regions, an approach called Granger Causality Mapping (GCM) has been proposed as a method for modeling effective connectivity without pre-selecting regions and connections between them [60]. The approach uses vector autoregressive (VAR) modeling of fMRI time series based on Granger causality [104]. This measure of causality quantifies the usefulness of the response of one brain region in predicting the signal of another region. As a result, the method identifies Granger causal maps, which are discrete-time linear models of voxels that are sources or targets of directed influence. Contrary to the models used in DCM, GCM is thus less biologically detailed. However, to improve the work done in this thesis, GCM could be applied first to identify all the relevant brain regions and the connections between them, creating a base of the DCM model structure.

Furthermore, it is strongly recommended to refine the model space and focus on identifying the optimal model structure since in this project, very few structures have been explored. One way of improving the model structure search in DCM is by using the Network Discovery approach [77]. This method relies on efficiently scoring a large number of model structures by approximating the log evidence of the reduced models (some connections disabled) from the log evidence of the full model (all connections allowed). This allows for the exploration of large model spaces without the need to invert every model structure. The Network Discovery approach provides additional functionality over GCM in that it is able to also infer how the input drives the neural activity and how nodes modulate the connectivity (matrix D in Nonlinear DCM). It cannot however identify the relevant brain regions so GCM is still recommended as the first step in the analysis.

Moreover, this project has only evaluated the performance of enhancing classical DCM with either Stochastic or Nonlinear terms but the two concepts can be combined. To gain additional insights into the neural underpinnings of the experiment, it is recommended to fit a model that has both state modulatory influences (matrix D) and neuronal state noise (additive noise ϖ). Based on the related studies summarized in Section 1-3, it appears that the OFC is highly likely to exert gating (nonlinear) influences on the connectivity related to emotional memory but it has not been confirmed. The Nonlinear DCM fitted in this thesis has not given a good fit to data because it does not incorporate the effect of neuronal noise, which has turned out to be crucial in the project.

Lastly, this project has studied brain connectivity in healthy subjects but it would be an interesting direction for future research to focus on subjects with brain damage or neurological disorders. For instance, in a classical conditioning experiment by [25], lesions on the amygdala have been shown to prevent subjects from acquiring an emotional response, while lesions on the hippocampus have been shown to preclude the subjects from remembering what stimuli were presented in the experiment. Furthermore, several neurological disorders such as Post-Traumatic Stress Disorder (PTSD), depression, or Obsessive-Compulsive Disorder (OCD) are associated with emotional dysregulation that affects both the anatomical and functional connectivity in the brain. Developing a model of effective connectivity in subjects with such disorders has the potential to improve the diagnosis and treatment through behavioral therapy and brain stimulation.

Appendix A

Additional results and figures

Matrix A

	A→A	H→A	O→A	A→H	H→H	O→H	A→O	H→O	O→O
NN	-0.1033	0.0574	0.0190	0.0472	-0.1093	0.0077	0.0271	0.0051	-0.0416
NE	-0.1248	0.0649	0.0191	0.0808	-0.0939	0.0092	0.0321	0.0095	-0.0444
EE	-0.1015	0.0552	0.0229	0.0471	-0.0782	0.0243	0.0220	0.0163	-0.0902

Matrix B1

	A→A	H→A	O→A	A→H	H→H	O→H	A→O	H→O	O→O
NN	0.0495	0.0080	-0.0060	-0.0036	-0.0097	-0.0155	0.0122	-0.0204	0.1030
NE	-0.0757	0.0107	0.0126	0.0138	0.0003	0.0076	0.0094	0.0201	-0.0407
EE	0.0206	-0.0056	-0.0121	-0.0077	0.0154	-0.0044	-0.0101	-0.0130	0.0035

Matrix B2

	A→A	H→A	O→A	A→H	H→H	O→H	A→O	H→O	O→O
NN	0.0002	0.0037	-0.0133	-0.0084	-0.0336	-0.0112	-0.0172	0.0082	0.0246
NE	0.0033	0.0037	-0.0140	-0.0039	-0.0120	-0.0049	-0.0036	0.0051	0.0481
EE	-0.0411	0.0310	0.0332	0.0271	-0.0201	0.0436	-0.0062	0.0191	0.0130

Matrix C

	u_1 →A	u_2 →A	u_1 →H	u_2 →H	u_1 →O	u_2 →O
NN	-0.0042	0.0008	0	0	0	0
NE	0	0	0	0	0.0008	0.0020
EE	0	0	-0.0006	0.0049	0	0

Table A-1: Estimated parameters in the optimal Stochastic DCM models for all groups. The columns are labeled to indicate which directed connection the corresponding values belong to. "A" stands for the Amygdala, "H" for the Hippocampus, and "O" for the Orbitofrontal Cortex. All values are rounded to 4 decimal places.

Adjacency matrices in Group NN

	A→A	H→A	O→A	A→H	H→H	O→H	A→O	H→O	O→O
$\mathbf{A}$	-0.1033	0.0574	0.0190	0.0472	-0.1093	0.0077	0.0271	0.0051	-0.0416
$\mathbf{A}'_1$	-0.0662	0.0465	0.0404	0.0625	-0.1084	-0.0121	0.0177	-0.0012	0.0297
$\mathbf{A}'_2$	-0.1155	0.0417	0.0110	0.0582	-0.1323	0.0164	0.0104	0.0031	-0.0487

Adjacency matrices in Group NE

	A→A	H→A	O→A	A→H	H→H	O→H	A→O	H→O	O→O
$\mathbf{A}$	-0.1248	0.0649	0.0191	0.0808	-0.0939	0.0092	0.0321	0.0095	-0.0444
$\mathbf{A}'_1$	-0.1826	0.0922	0.0401	0.0720	-0.0907	0.0234	0.0321	0.0161	-0.0869
$\mathbf{A}'_2$	-0.1036	0.0745	0.0270	0.0650	-0.1030	0.0084	0.0055	0.0036	0.0019

Adjacency matrices in Group EE

	A→A	H→A	O→A	A→H	H→H	O→H	A→O	H→O	O→O
$\mathbf{A}$	-0.1015	0.0552	0.0229	0.0471	-0.0782	0.0243	0.0220	0.0163	-0.0902
$\mathbf{A}'_1$	-0.0758	0.0346	0.0159	0.0433	-0.0617	0.0018	0.0055	0.0102	-0.0908
$\mathbf{A}'_2$	-0.1376	0.0694	0.0198	0.0799	-0.0972	0.0339	0.0509	0.0582	-0.0812

Table A-2: Estimated parameters of A and A' in the optimal Stochastic DCM models for all groups. A'_1 is the coupling matrix when input 1 is presented, and A'_2 when input 2 is presented. The columns are labeled to indicate which directed connection the corresponding values belong to. "A" stands for the Amygdala, "H" for the Hippocampus, and "O" for the Orbitofrontal Cortex. All values are rounded to 4 decimal places.

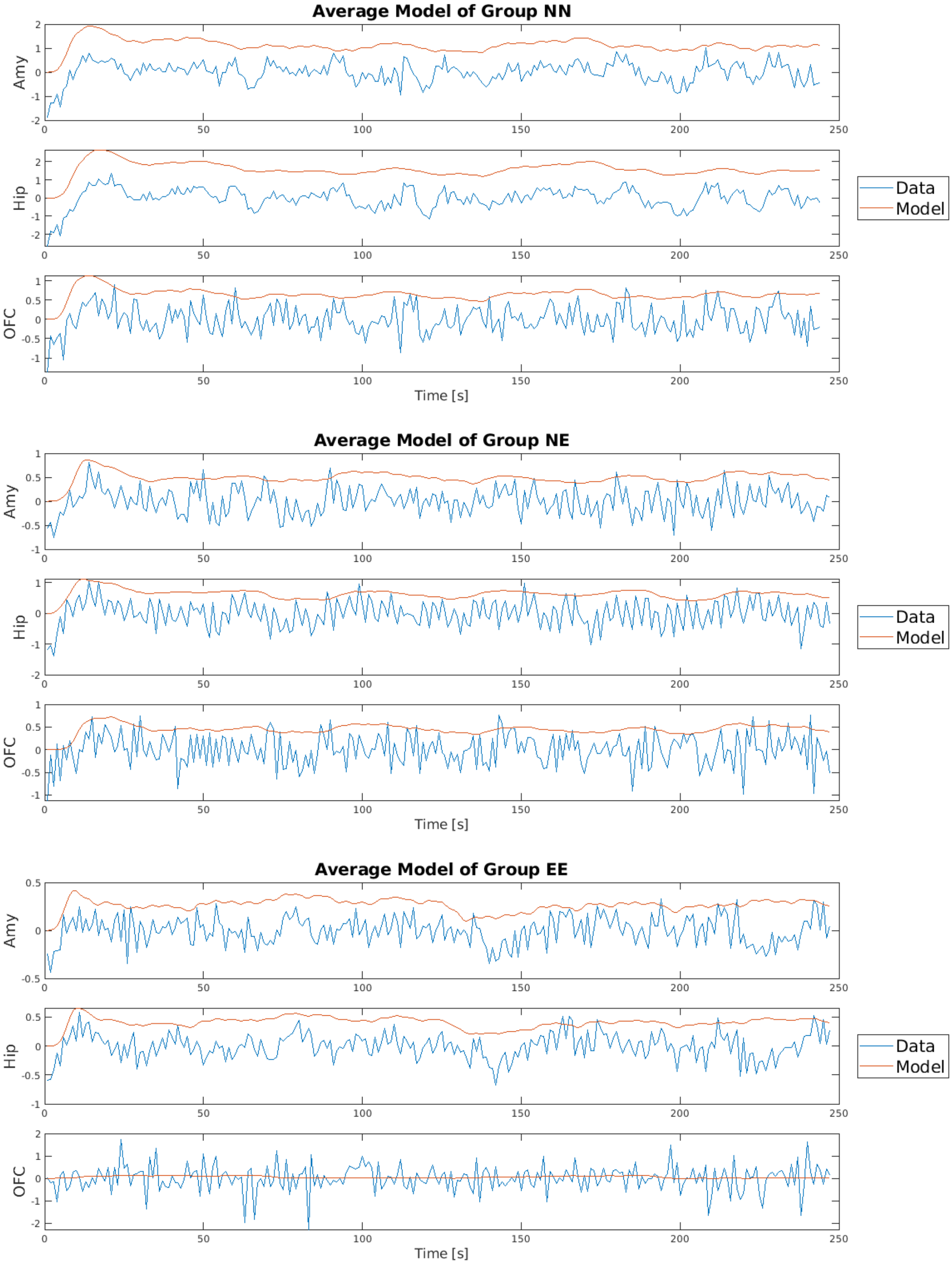


Figure A-1: Average optimal models vs. data (average signal for each group) for Classical DCM.

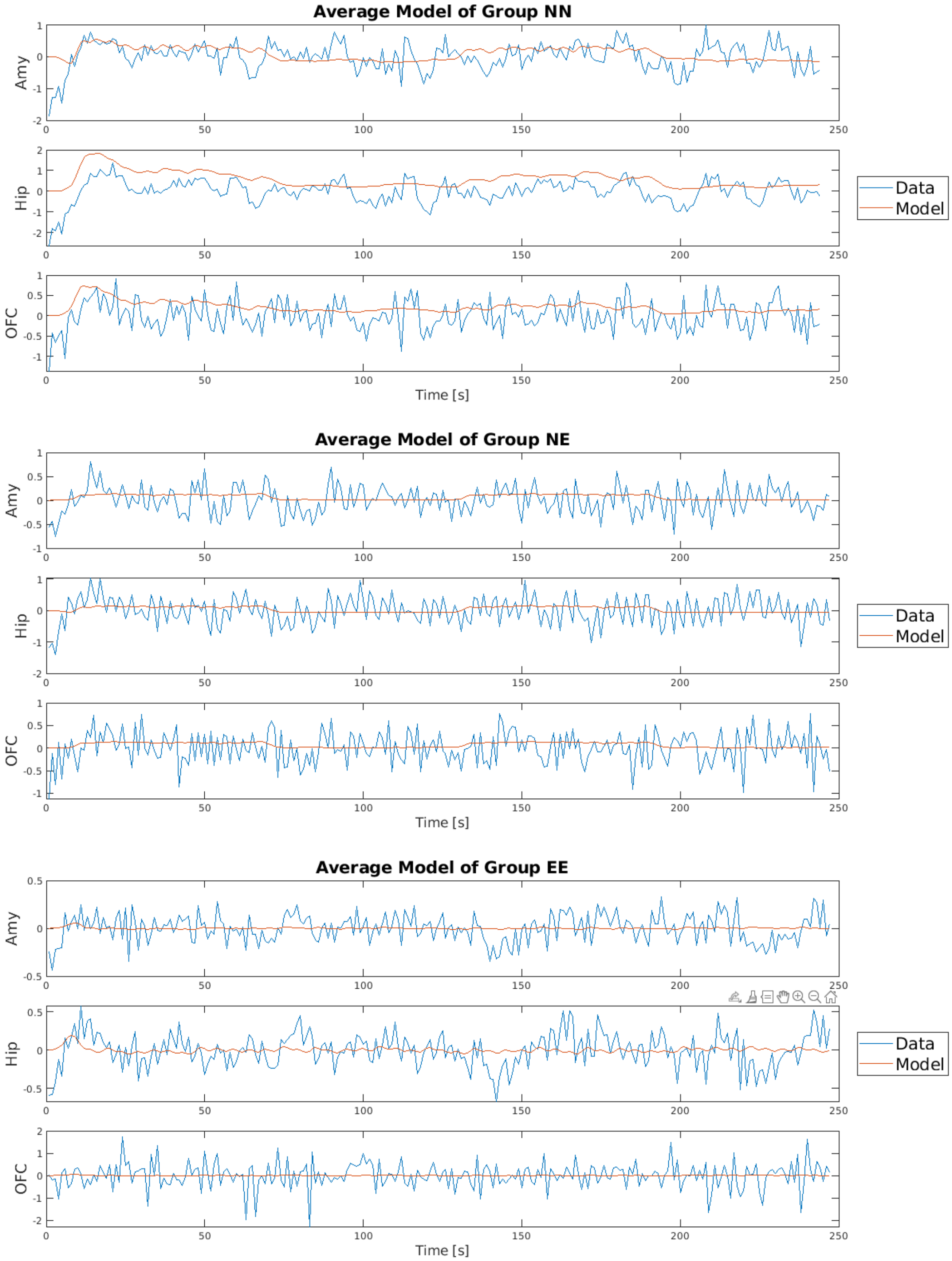


Figure A-2: Average optimal models vs. data (average signal for each group) for Nonlinear DCM.

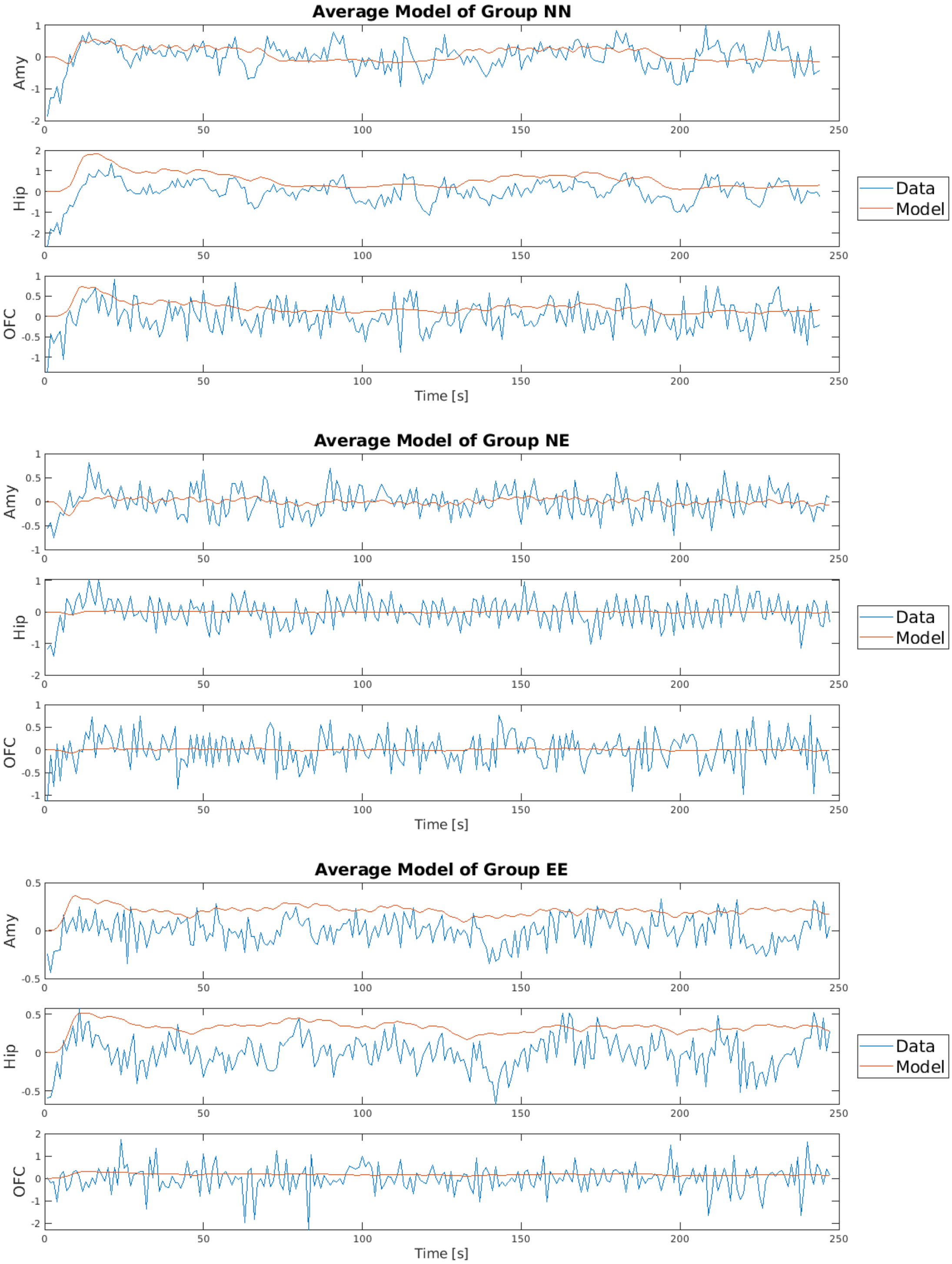


Figure A-3: Average optimal models vs. data (average signal for each group) for Two-State DCM.

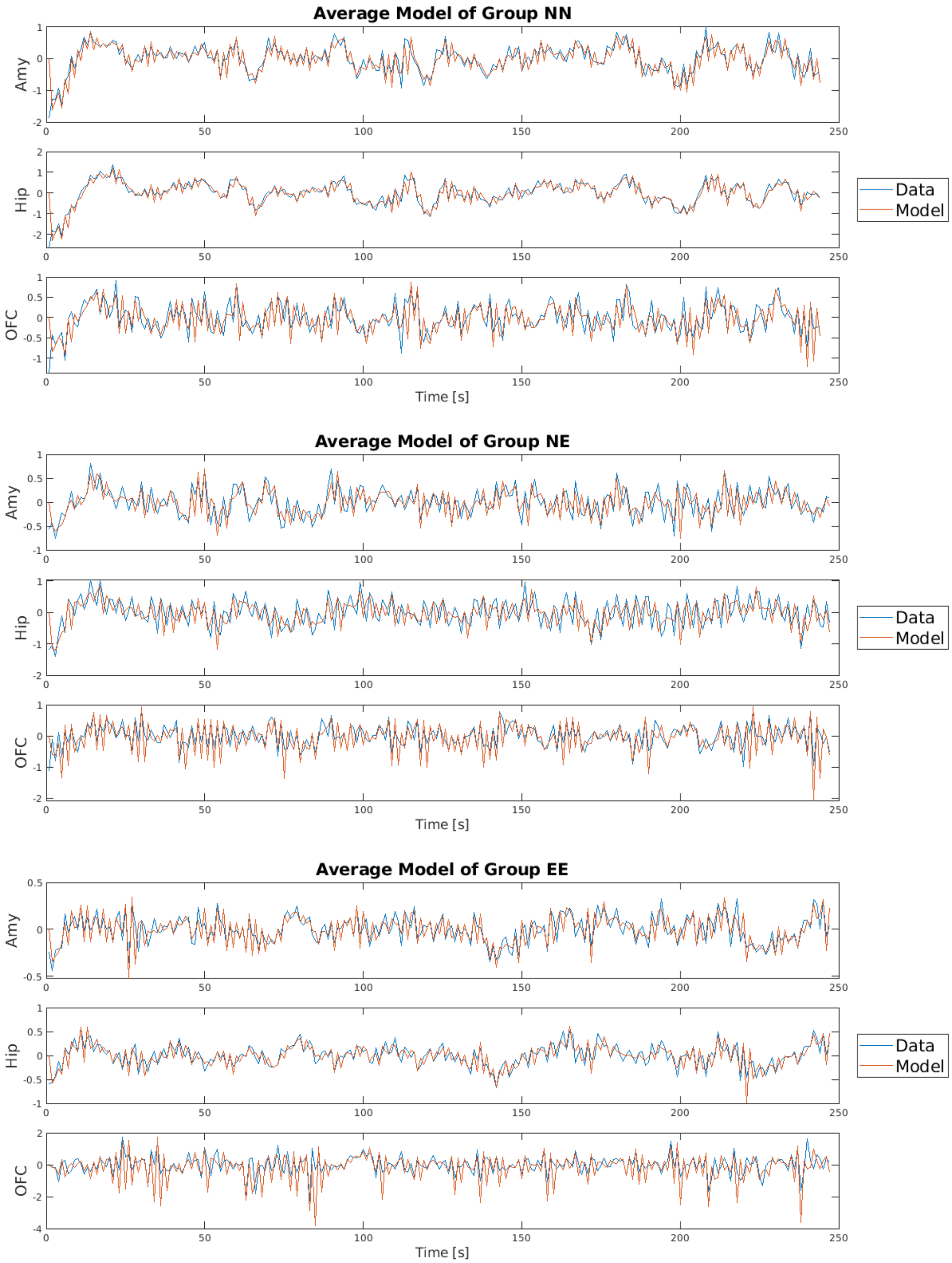


Figure A-4: Average optimal models vs. data (average signal for each group) for Stochastic DCM.

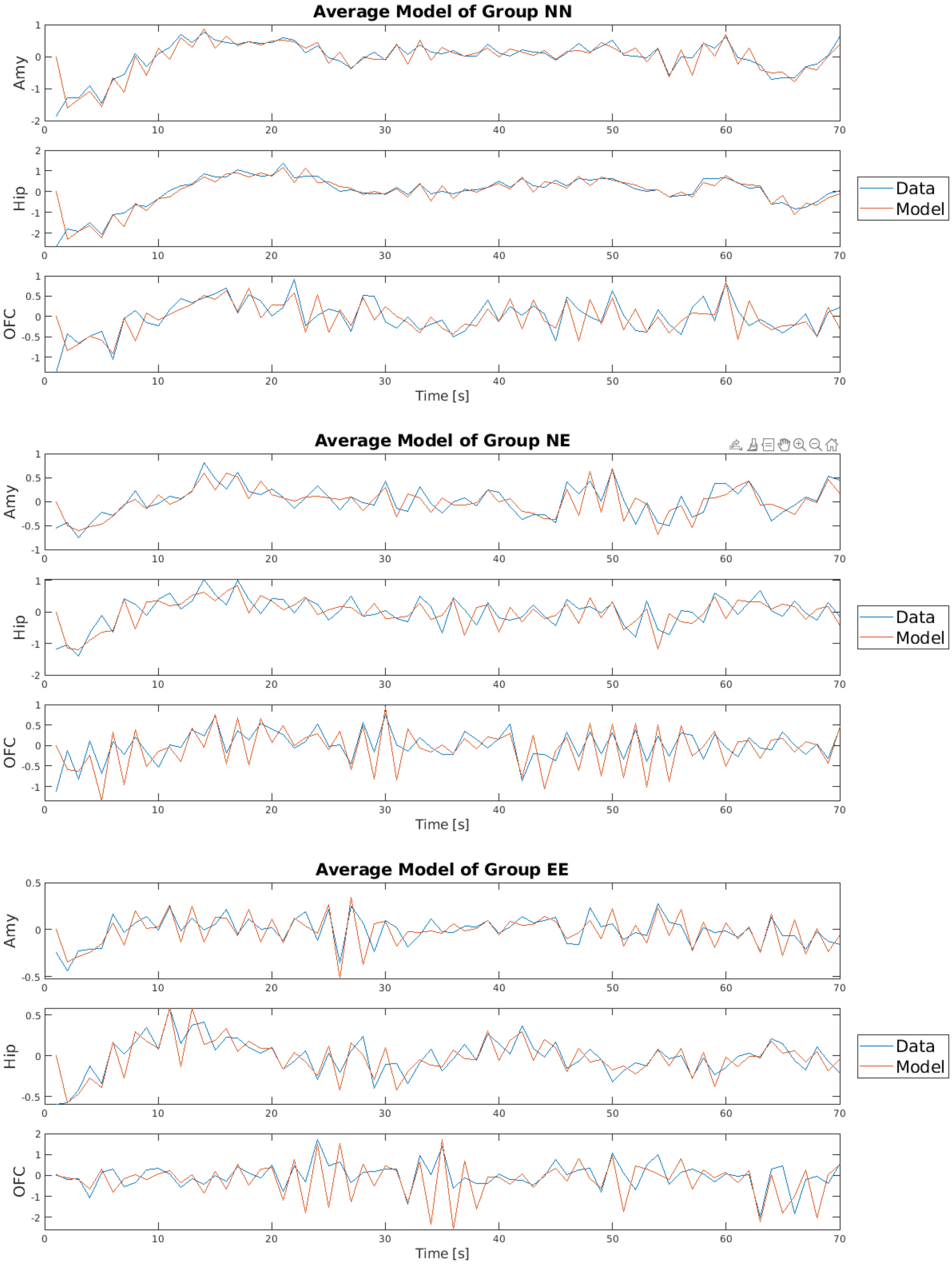


Figure A-5: Zoomed-in average optimal models vs. data (average signal for each group) for Stochastic DCM.

Appendix B

A quick guide to SPM12 DCM scripting

Statistical Parametric Mapping (SPM12) [93] is a toolbox implemented in MATLAB meant for analyzing functional imaging data. Crucially, the toolbox includes various functions necessary to estimate and compare Dynamic Causal Modeling (DCM) systems.

The creators of the toolbox have released a user manual [94] but the content focused on DCM is limited. In particular, the only chapter about DCM for functional Magnetic Resonance Imaging (fMRI) data explains solely how to use the Graphical User Interface (GUI) and there is no mention of how to access the same functionality through a MATLAB script. Such manual selection of modeling parameters would take a considerable amount of time, be highly prone to errors, and require the user to perform the same routine steps for every subject and every time the user wishes to try some new settings.

There is in fact a built-in functionality for SPM12 scripting but the manual does not explain how to use it for DCM. The only online tutorial on this matter that has been found is [74] but even this one does not give a comprehensive overview of scripting in the toolbox. However, combining the advice from the online tutorial, the sole DCM example script included in the SPM12 toolbox, and a fair amount of trial and error, a full SPM12 DCM script has been developed for the use in this project. The goal of this appendix is to briefly explain how to develop a script to access the functionality necessary to fit DCM models on a subject and group level.

B-1 Main code

The main code sets up the environment, directories, and experiment parameters, and calls all the necessary functions.

```
% Start the script by adding to the path the folder where you have saved
   SPM12 and initializing the toolbox.
```

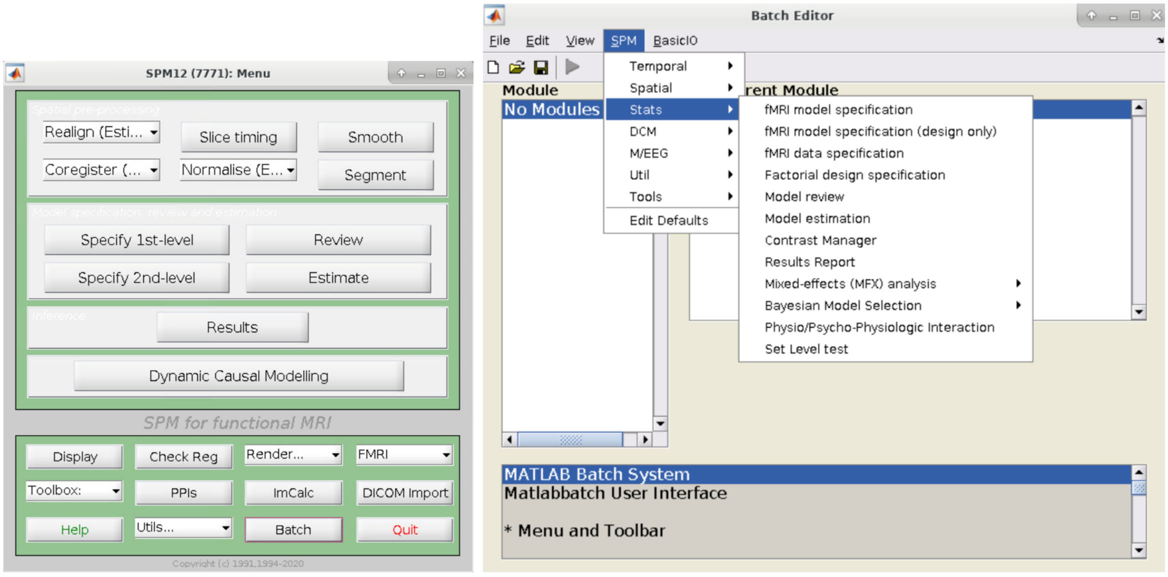
```
addpath /home/cogaff/werdzi/Downloads/spm12
spm('Defaults','fMRI');
spm_jobman('initcfg');

% Then set the directory with your fMRI data and the directory where you
    would like to save the results of model estimation (conventionally,
    this folder is called GLM).
dir_data = '/project/3013067.02/msc_thesis/Encoding/Data/';
glm_name = '/project/3013067.02/msc_thesis/Encoding/GLMs';

% Now, import your table (or other type of file) that contains the
    subject IDs and information on which group they belong to.
sheet = readtable('/project/3013067.02/Data_analysis/Behavioral_results/
    EMC_memory_performance_final.xlsx','Sheet',2);

% As your analysis will probably show, some subjects do not exhibit high
    activation due to external stimulation. As a result, it is possible
    that no significant voxels will be found for some of the subjects and
    the script would break with an error message. To avoid terminating the
     run, the "try and catch" functionality is useful. To record the
    success and failure rate, create a structure array.
subjectID = string(table2array(sheet(1:70,1))));
group = string(table2array(sheet(1:70,2)));
groupID = sheet{1:70,3};
status = repmat("Not attempted",70,1); % This will change to "Success" or
     the error message after fitting has been attempted.
info_indv = table(subjectID,group,groupID,status);

% Now it is time to loop over all subjects and extract their relevant
    time series (GLM) and fit DCM models on a subject level. Functions GLM
    (id,glm_name,dir_data) and DCM(id,glm_name,dir_data) are explained
    later in the tutorial.
for i = 1:length(info_indv{:,1})
    id = convertStringsToChars(info_indv{i,1});
    try
        % Fit DCM for subject i
        GLM(id,glm_name,dir_data);
        DCM(id,glm_name,dir_data);
        % Print a success message
        fprintf('Subject %s analysis completed.\n', id);
        info_indv{i,4} = "Success";
    catch exception
        % Print an error message
        fprintf('Error processing subject %s: %s\n', id, exception.
            message);
        info_indv{i,4} = string(exception.message);
        % Continue to the next iteration
        continue;
    end
end

% Save the structure array.
```

```
save(strcat('/project/3013067.02/msc_thesis/Encoding/GLMs_',glm_name,'/
    info_indv.mat'),'info_indv');

% Now that the subject-level DCMs have been fitted, it is time to find
    the optimal group-level models. In this project, the suitable group-
    level analysis is Bayesian Model Selection (BMS) and Bayesian
    Parameter Averaging (BPA). Refer to Section 2-2-3 of this report to
    select the suitable procedure for your case. The BMSBPA(info_indv,
    glm_name) function is explained later in the tutorial.
BMSBPA(info_indv,glm_name);
```

B-2 Time-series extraction

The script for extracting time series through a General Linear Model (GLM) and Volume of Interest (VOI) analysis can be generated using the "Batch" menu from the GUI. The "Batch" menu allows the user to select the desired functionalities in the GUI and create a script that may be used in a loop for all subjects or altered to try new settings later.

To open the SPM12 GUI, type "spm fmri" in the command window and the menu will pop up. Press on the "Batch" button, then select "SPM" from the toolbar. The two menus are depicted in Figure B-1.



MATLAB Batch System
Matlabbatch User Interface

* Menu and Toolbar

Figure B-1: The SPM12 GUI for fMRI data: main menu (on the left) and the batch editor (on the right).

There are several options in the batch editor that are necessary to extract time series for DCM, namely "Stats → fMRI model specification", "Stats → Contrast Manager", and "Util → Volumes of Interest". After all the desired settings have been selected, click on "File → Save Batch and Script". This will save the script in your current working directory. You can now open it and copy-paste its content to create your custom script.

Function GLM presented below has been constructed in exactly this manner. To further explain the various involved steps, the script includes detailed comments. It should be noted that there are multiple design decisions to make here so make sure to change all the parameters in the script to match your modeling approach.

```matlab
function GLM(id,glm_name,dir_data)

% Add the directory containing the data and the GLM folder to path.
addpath /project/3013067.02/msc_thesis/Encoding

% Every time a GLM is estimated for a subject, a file called SPM.mat is created and there is no way to change the file name for every subject. Instead, create a separate subfolder for each subject in your GLM folder.
dir_glm = strcat('/project/3013067.02/msc_thesis/Encoding/GLMs_',glm_name,'/GLM',id);

% Clear matlabbatch to avoid using settings from previous runs
clear matlabbatch

% The code below was generated by the Batch Editor using the Stats -> fMRI model specification option. This portion of the code specifies the data directories, timing parameters, and input onsets, and then it estimates the GLM.
matlabbatch{1}.spm.stats.fmri_spec.dir = {dir_glm};
matlabbatch{1}.spm.stats.fmri_spec.timing.units = 'scans';
matlabbatch{1}.spm.stats.fmri_spec.timing.RT = 1;
matlabbatch{1}.spm.stats.fmri_spec.sess(1).scans = {strcat(dir_data,id,'_scans',mode,'.nii')};
matlabbatch{1}.spm.stats.fmri_spec.sess(1).cond(1).name = 'Image 1';
matlabbatch{1}.spm.stats.fmri_spec.sess(1).cond(1).onset = load(strcat(dir_data,id,'_onsets',mode,'_image1.txt'));
matlabbatch{1}.spm.stats.fmri_spec.sess(1).cond(1).duration = 2.5;
matlabbatch{1}.spm.stats.fmri_spec.sess(1).cond(2).name = 'Image 2';
matlabbatch{1}.spm.stats.fmri_spec.sess(1).cond(2).onset = load(strcat(dir_data,id,'_onsets',mode,'_image2.txt'));
matlabbatch{1}.spm.stats.fmri_spec.sess(1).cond(2).duration = 2.5;
matlabbatch{2}.spm.stats.fmri_est.spmmat(1) = cfg_dep('fMRI model specification: SPM.mat File', substruct('.','val', '{}',{1}, '.','val', '{}',{1}, '.','val', '{}',{1}), substruct('.','spmmat'));

% The code below was generated by the Batch Editor using the Stats -> Contrast Manager option. This portion of the code specifies the T and F contrasts for time series extraction.
matlabbatch{3}.spm.stats.con.spmmat(1) = cfg_dep('Model estimation: SPM.mat File', substruct('.','val', '{}',{2}, '.','val', '{}',{1}, '.','val', '{}',{1}), substruct('.','spmmat'));
matlabbatch{3}.spm.stats.con.consess{1}.fcon.name = 'Effects of interest';
matlabbatch{3}.spm.stats.con.consess{1}.fcon.weights = [1 0; 0 1];
matlabbatch{3}.spm.stats.con.consess{1}.fcon.sessrep = 'none';
matlabbatch{3}.spm.stats.con.consess{2}.tcon.name = 'Image 1';
matlabbatch{3}.spm.stats.con.consess{2}.tcon.weights = [1 0];
```

```
matlabbatch{3}.spm.stats.con.consess{2}.tcon.sessrep = 'none';
matlabbatch{3}.spm.stats.con.consess{3}.tcon.name = 'Image 2';
matlabbatch{3}.spm.stats.con.consess{3}.tcon.weights = [0 1];
matlabbatch{3}.spm.stats.con.consess{3}.tcon.sessrep = 'none';
matlabbatch{3}.spm.stats.con.delete = 0;
spm_jobman('run',matlabbatch);

% The code below was generated by the Batch Editor using the Util ->
    Volumes of Interest option. This portion of the code specifies the
    VOIs including their anatomical location and activation assessed using
     T and F contrasts.
clear matlabbatch
VOI_names = {'Amy', 'Hip', 'OFC'};
VOI_mask_dir = '/project/3013067.02/msc_thesis/Encoding/VOImasks/';
VOI_masks = {'Amy_L.nii', 'Hip_L.nii', 'OFC_L.nii'};

matlabbatch = cell(length(VOI_names),1);
for i = 1:length(VOI_names)
    matlabbatch{i}.spm.util.voi.spmmat = {strcat(dir_glm,'/SPM.mat')};
    matlabbatch{i}.spm.util.voi.adjust = 1;
    matlabbatch{i}.spm.util.voi.session = 1;
    matlabbatch{i}.spm.util.voi.name = VOI_names{i};

    % Anatomical mask
    matlabbatch{i}.spm.util.voi.roi{1}.mask.image = {strcat(VOI_mask_dir,
        VOI_masks{i})};
    matlabbatch{i}.spm.util.voi.roi{1}.mask.threshold = 0.5;

    % Level of activation
    matlabbatch{i}.spm.util.voi.roi{2}.spm.spmmat = {''};
    matlabbatch{i}.spm.util.voi.roi{2}.spm.contrast = 2;
    matlabbatch{i}.spm.util.voi.roi{2}.spm.threshdesc = {'none'};
    matlabbatch{i}.spm.util.voi.roi{2}.spm.thresh = 0.05;
    matlabbatch{i}.spm.util.voi.roi{2}.spm.extent = 0;
    matlabbatch{i}.spm.util.voi.roi{2}.spm.mask.contrast = 3;
    matlabbatch{i}.spm.util.voi.roi{2}.spm.mask.thresh = 0.05;
    matlabbatch{i}.spm.util.voi.roi{2}.spm.mask.mtype = 0;

    % Expression to combine anatomical mask and input contrasts
    matlabbatch{i}.spm.util.voi.expression = 'i1 & i2';
end
spm_jobman('run',matlabbatch);

% Files 'VOI_name_1.mat' have been created for each of your VOIs.

end
```

B-3 Subject-level DCM

Now that time series have been extracted from the relevant voxels, a subject-level DCM may be fitted. The code below has been adapted from an example script from the SPM12 tutorial.

There is no way to generate the full code for DCM specification and estimation using the Batch Editor.

This function creates a DCM structure array that contains all the information about the signal, inputs, timing, and other parameters. This struct is then sent to the ModelSpace, which assigns the connections and saves separate files for every model structure.

```
function DCM(id,glm_name,dir_data)

% Add the directory containing the data and the GLM folder to path.
addpath /project/3013067.02/msc_thesis/Encoding

% Specify the GLM subfolder for the current subject.
dir_glm = strcat('/project/3013067.02/msc_thesis/Encoding/GLMs_',glm_name,'/GLM',id);

% Clear the DCM struct to avoid overwriting a previous struct.
clear DCM

% Load the files created by function GLM
load(fullfile(dir_glm,'SPM.mat'));
load(fullfile(dir_glm,'VOI_Amy_1.mat'),'xY');
DCM.xY(1) = xY;
load(fullfile(dir_glm,'VOI_Hip_1.mat'),'xY');
DCM.xY(2) = xY;
load(fullfile(dir_glm,'VOI_OFC_1.mat'),'xY');
DCM.xY(3) = xY;

% Specify the size of the model
DCM.n = length(DCM.xY);       % number of regions
DCM.v = length(DCM.xY(1).u); % number of time points

% Specify the time series
DCM.Y.dt  = SPM.xY.RT;
DCM.Y.X0  = DCM.xY(1).X0;
for i = 1:DCM.n
    DCM.Y.y(:,i)  = DCM.xY(i).u;
    DCM.Y.name{i} = DCM.xY(i).name;
end
DCM.Y.Q    = spm_Ce(ones(1,DCM.n)*DCM.v);

% Specify the experimental inputs
DCM.U.dt   =  SPM.Sess.U(1).dt;
DCM.U.name = [SPM.Sess.U.name];
DCM.U.u    = [SPM.Sess.U(1).u(1:end,1) ...
              SPM.Sess.U(2).u(1:end,1)];

% Specify the DCM timing
DCM.delays = [0.5 0.5 0.5];
DCM.TE     = 0.04;

```

```
% Create the DCM model space including all the model structures you wish to fit and compare later. The function ModelSpace is explained later in the tutorial.
ModelSpace(DCM,dir_glm);

% Load the file with filenames of all the DCMs in the model space, created by ModelSpace.
DCMfilenames = load(strcat(dir_glm,'/DCMfilenames.mat')).DCMfilenames;

% Estimate all the DCMs from the model space.
clear matlabbatch
matlabbatch{1}.spm.dcm.fmri.estimate.dcmmat = DCMfilenames;
spm_jobman('run',matlabbatch);

end
```

The function ModelSpace is a custom function that serves as an example of how a large space can be efficiently created using for loops. The actual connections and other settings depend on your modeling approach.

```
function ModelSpace(DCM,dir_glm)

% Constant values - all the models in this model space share these.
DCM.options.nograph    = 1;
DCM.options.centre     = 1;

% In this project, several DCM types are used: classical bilinear, nonlinear, stochastic, and two-state. For explanations of the possible connections, refer to section 3-1-3.

% A & B: all connections always allowed

% C: one row is always all 1s and the other rows are all 0s

% D: only one node can exert modulatory influences on both connections
% between the other nodes:

%      Amy  Hip  OFC        Amy  Hip  OFC        Amy  Hip  OFC        Amy  Hip  OFC
% Amy   0    *    *     Amy  0    1    0     Amy  0    0    1     Amy  0    0    0
% Hip   *    0    *     Hip  1    0    0     Hip  0    0    0     Hip  0    0    1
% OFC   *    *    0     OFC  0    0    0     OFC  1    0    0     OFC  0    1    0

% Possible D off-diagonal connections
D_opts = [1  1  0  0  0  0;
          0  0  1  1  0  0;
          0  0  0  0  1  1];

id = 0;
% Loop over the 3 possible C matrices
for i = 1:3
    C = zeros(3,2);
    C(i,:) = 1;
    DCM.a = ones(3);
    DCM.b = ones(3);
    DCM.c = C;
    DCM.d = zeros(3);

    % Bilinear
    DCM.options.nonlinear  = 0;
    DCM.options.two_state  = 0;
    DCM.options.stochastic = 0;
    id = id + 1;
    filename = strcat(dir_glm,'/DCM',num2str(id),'.mat');
    save(filename,'DCM')

    % Stochastic
    DCM.options.nonlinear  = 0;
    DCM.options.two_state  = 0;
    DCM.options.stochastic = 1;
    id = id + 1;
    filename = strcat(dir_glm,'/DCM',num2str(id),'.mat');
    save(filename,'DCM')

    % Two State
    DCM.options.nonlinear  = 0;
    DCM.options.two_state  = 1;
    DCM.options.stochastic = 0;
    id = id + 1;
    filename = strcat(dir_glm,'/DCM',num2str(id),'.mat');
    save(filename,'DCM')

    % Nonlinear
    DCM.options.nonlinear  = 1;
    DCM.options.two_state  = 0;
    DCM.options.stochastic = 0;
    % Loop over the 3 possible D matrices
    for k = 1:size(D_opts,1)
        D = zeros(3,3,3);
        D([6 8 12 16 20 22]) = D_opts(k,:);
        DCM.d = D;
        id = id + 1;
        filename = strcat(dir_glm,'/DCM',num2str(id),'.mat');
        save(filename,'DCM')
    end
end

% Create and save a cell array with filenames of all DCMs
DCMfilenames = cell(id,1);
for i = 1:id
    DCMfilenames{i} = strcat(dir_glm,'/DCM',num2str(i),'.mat');
end
save(strcat(dir_glm,'/DCMfilenames.mat'),'DCMfilenames');

end
```

B-4 Group-level DCM

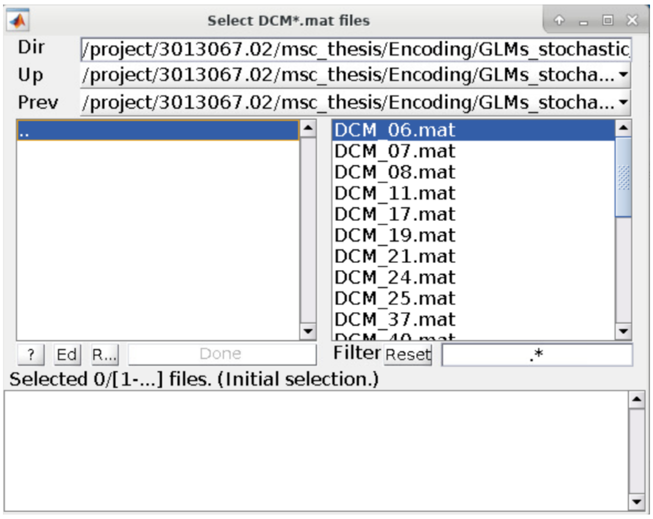
Now, it is time to find the optimal model structure per group using Bayesian Model Selection (BMS) and then, find average parameters for the optimal structures using Bayesian Parameter Averaging (BPA). The code for BMS has been obtained using Batch Editor → SPM → DCM → Bayesian Model Selection → Model Inference.

Furthermore, it appears that the Batch Editor does not have an option to generate a script for BPA so this has to be done manually using the GUI. To save time, the optimal models of all participants from a group can be saved in one file. This way, it will be easy to select all the files from a folder and quickly perform BPA for all groups. The code below prepares such folders for each group.

```
function BMSBPA(info_indv,glm_name)

% Make a table to store optimal models per group.
group = ["NN";"NE";"EE"];
bestModel = zeros(3,1);
info_group = table(group,bestModel);

% Load the file with filenames of all the DCMs in the model space, created by ModelSpace.
DCMfilenames = load(strcat(dir_glm,'/DCMfilenames.mat')).DCMfilenames;

% Make folders to store all models per group for later BPA using the GUI.
mkdir(strcat('/project/3013067.02/msc_thesis/Encoding/GLMs_',glm_name,'/BPA_NN'));
mkdir(strcat('/project/3013067.02/msc_thesis/Encoding/GLMs_',glm_name,'/BPA_NE'));
mkdir(strcat('/project/3013067.02/msc_thesis/Encoding/GLMs_',glm_name,'/BPA_EE'));

% Consider only the subject for whom relevant voxels have been found and thus, GLM has been successful.
info_indv_success = info_indv(strcmp(info_indv.status,"Success."), :);

% Loop over the groups to compare the model structure through BMS
for groupID = 1:3
    ids_success_group = table2array(info_indv_success(info_indv_success.groupID == groupID, 1));

    % Find the optimal model structure for the current group.
    clear matlabbatch
    matlabbatch{1}.spm.dcm.bms.inference.dir = {strcat('/project/3013067.02/msc_thesis/Encoding/GLMs_',glm_name)};
    for i = 1:length(ids_success_group)
        id = convertStringsToChars(ids_success_group(i));
        matlabbatch{1}.spm.dcm.bms.inference.sess_dcm{i}.dcmmat = {DCMfilenames};
    end
    matlabbatch{1}.spm.dcm.bms.inference.model_sp = {''};
    matlabbatch{1}.spm.dcm.bms.inference.load_f = {''};
    matlabbatch{1}.spm.dcm.bms.inference.method = 'FFX';
    matlabbatch{1}.spm.dcm.bms.inference.family_level.family_file = {''};
    matlabbatch{1}.spm.dcm.bms.inference.bma.bma_no = 0;
    matlabbatch{1}.spm.dcm.bms.inference.verify_id = 1;
    spm_jobman('run',matlabbatch);

    % Get the id of the optimal model for this group and save it in the info_group struct.
    BMS = load(strcat('/project/3013067.02/msc_thesis/Encoding/GLMs_', glm_name,'/BMS.mat'));
    [~,DCM_bestID] = max(BMS.BMS.DCM.ffx.model.post);
    info_group{groupID,2} = DCM_bestID;
    % The BMS file is called BMS.mat so to avoid overwriting it for the next group, copy the file with a unique name and delete the BMS.mat file.
    save(strcat('/project/3013067.02/msc_thesis/Encoding/GLMs_',glm_name,'/BMS_',group(groupID),'.mat'),'BMS');
    delete(strcat('/project/3013067.02/msc_thesis/Encoding/GLMs_', glm_name,'/BMS.mat'));

    % Copy DCMs with the optimal model structure from the current group
    % into one folder. This will make file selection for BPA easier.
    for i = 1:length(ids_success_group)
        id = convertStringsToChars(ids_success_group(i));
        copyfile(strcat('/project/3013067.02/msc_thesis/Encoding/GLMs_', glm_name,'/GLM',id,'/DCM',num2str(DCM_bestID),'.mat'), ...
            strcat('/project/3013067.02/msc_thesis/Encoding/GLMs_', glm_name,'/BPA_',group(groupID),'/DCM_',id,'.mat'));
    end

end

% Save the info_group structure to access the BMS results easily.
save(strcat('/project/3013067.02/msc_thesis/Encoding/GLMs_',glm_name,'/info_group.mat'),'info_group');

end
```

After the code has run and saved the files in the BPA_group folders, go to the GUI main menu, click on "Dynamic Causal Modeling", and select from the drop-down menu "Action: average" → BPA. A new menu will open, as depicted in Figure B-2. Navigate to the BPA folder for a given group and select all the files that the BMSBPA function has copied there. The editor will fit an average model for this group and save it under the specified name.

This concludes the DCM estimation. Now, you may review your results and analyze the fitted models. Simply load the DCM files and inspect their contents. The estimated A, B, C, and D matrices can be found in the DCM.Ep field. Furthermore, to visualize the fit of the model, you may plot the measured signal DCM.Y.y against the simulated signal DCM.y.



Select DCM*.mat files

Dir /project/3013067.02/msc_thesis/Encoding/GLMs_stochastic

Up /project/3013067.02/msc_thesis/Encoding/GLMs_stocha...

Prev /project/3013067.02/msc_thesis/Encoding/GLMs_stocha...

..

DCM_06.mat
DCM_07.mat
DCM_08.mat
DCM_11.mat
DCM_17.mat
DCM_19.mat
DCM_21.mat
DCM_24.mat
DCM_25.mat
DCM_37.mat
DCM_40.mat

Filter .*

Selected 0/[1-...] files. (Initial selection.)

Figure B-2: The SPM12 GUI for fMRI data: BPA menu.

Bibliography

[1] Y. Zhu, W. Liu, N. Kohn, and G. Fernández, "Emotional information facilitates or disrupts memory integration through distinct hippocampal processes of reactivation and connectivity," Apr. 2023, pages: 2023.04.25.538111 Section: New Results. [Online]. Available: https://www.biorxiv.org/content/10.1101/2023.04.25.538111v1

[2] C. B. MorÉn, Jan, "Emotional Learning: A Computational Model of the Amygdala," *Cybernetics and Systems*, vol. 32, no. 6, pp. 611–636, Sep. 2001, publisher: Taylor & Francis _eprint: https://www.tandfonline.com/doi/pdf/10.1080/01969720118947. [Online]. Available: https://www.tandfonline.com/doi/abs/10.1080/01969720118947

[3] H. Eichenbaum, "Hippocampus: Cognitive Processes and Neural Representations that Underlie Declarative Memory," *Neuron*, vol. 44, no. 1, pp. 109–120, Sep. 2004. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S089662730400529X

[4] E. m, J. L. Plass, E. O. Hayward, and B. D. Homer, "Emotional design in multimedia learning," *Journal of Educational Psychology*, vol. 104, pp. 485–498, 2012, place: US Publisher: American Psychological Association.

[5] S. D'Mello, B. Lehman, R. Pekrun, and A. Graesser, "Confusion can be beneficial for learning," *Learning and Instruction*, vol. in press, p. X, Jan. 2013.

[6] S. Vogel and L. Schwabe, "Learning and memory under stress: implications for the classroom," *npj Science of Learning*, vol. 1, no. 1, p. 16011, Jun. 2016. [Online]. Available: https://www.nature.com/articles/npjscilearn201611

[7] G. Richter-Levin and I. Akirav, "Amygdala-hippocampus dynamic interaction in relation to memory," *Molecular Neurobiology*, vol. 22, no. 1, pp. 11–20, Aug. 2000. [Online]. Available: https://doi.org/10.1385/MN:22:1-3:011

[8] G. Okada, Y. Okamoto, Y. Kunisato, S. Aoyama, Y. Nishiyama, S. Yoshimura, K. Onoda, S. Toki, H. Yamashita, and S. Yamawaki, "The Effect of Negative and Positive Emotionality on Associative Memory: An fMRI Study," *PLOS ONE*, vol. 6, no. 9, p. e24862, Sep. 2011, publisher: Public Library of Science. [Online]. Available: https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0024862

[9] C. R. Madan, E. Fujiwara, J. B. Caplan, and T. Sommer, "Emotional arousal impairs association-memory: Roles of amygdala and hippocampus," *NeuroImage*, vol. 156, pp. 14–28, Aug. 2017. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1053811917303841

[10] F. N. Ahmad and W. E. Hockley, "The role of familiarity in associative recognition of unitized compound word pairs," *Quarterly Journal of Experimental Psychology*, vol. 67, no. 12, pp. 2301–2324, Dec. 2014, publisher: SAGE Publications. [Online]. Available: https://doi.org/10.1080/17470218.2014.923007

[11] J. A. Bisby, A. J. Horner, L. D. Hørlyck, and N. Burgess, "Opposing effects of negative emotion on amygdalar and hippocampal memory for items and associations," *Social Cognitive and Affective Neuroscience*, vol. 11, no. 6, pp. 981–990, Jun. 2016. [Online]. Available: https://doi.org/10.1093/scan/nsw028

[12] L. Cahill, M. Uncapher, L. Kilpatrick, M. T. Alkire, and J. Turner, "Sex-Related Hemispheric Lateralization of Amygdala Function in Emotionally Influenced Memory: An fMRI Investigation," *Learning & Memory*, vol. 11, no. 3, pp. 261–266, May 2004, company: Cold Spring Harbor Laboratory Press Distributor: Cold Spring Harbor Laboratory Press Institution: Cold Spring Harbor Laboratory Press Label: Cold Spring Harbor Laboratory Press Publisher: Cold Spring Harbor Lab. [Online]. Available: http://learnmem.cshlp.org/content/11/3/261

[13] T. Canli, J. E. Desmond, Z. Zhao, and J. D. E. Gabrieli, "Sex differences in the neural basis of emotional memories," *Proceedings of the National Academy of Sciences*, vol. 99, no. 16, pp. 10 789–10 794, Aug. 2002, publisher: Proceedings of the National Academy of Sciences. [Online]. Available: https://www.pnas.org/doi/abs/10.1073/pnas.162356599

[14] B. Ćurčić Blake, M. Swart, and A. Aleman, "Bidirectional Information Flow in Frontoamygdalar Circuits in Humans: A Dynamic Causal Modeling Study of Emotional Associative Learning," *Cerebral Cortex*, vol. 22, no. 2, pp. 436–445, Feb. 2012. [Online]. Available: https://doi.org/10.1093/cercor/bhr124

[15] M. G. Craske, M. Treanor, C. Conway, T. Zbozinek, and B. Vervliet, "Maximizing Exposure Therapy: An Inhibitory Learning Approach," *Behaviour research and therapy*, vol. 58, pp. 10–23, Jul. 2014. [Online]. Available: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4114726/

[16] R. K. McHugh, B. A. Hearon, and M. W. Otto, "Cognitive-Behavioral Therapy for Substance Use Disorders," *The Psychiatric clinics of North America*, vol. 33, no. 3, pp. 511–525, Sep. 2010. [Online]. Available: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2897895/

[17] B. Kopell, B. Greenberg, and A. Rezai, "Deep Brain Stimulation for Psychiatric Disorders," *Journal of clinical neurophysiology : official publication of the American Electroencephalographic Society*, vol. 21, pp. 51–67, Jan. 2004.

[18] M. S. George, H. A. Sackeim, A. J. Rush, L. B. Marangell, Z. Nahas, M. M. Husain, S. Lisanby, T. Burt, J. Goldman, and J. C. Ballenger, "Vagus nerve stimulation: a new tool for brain research and therapy,"

Biological Psychiatry, vol. 47, no. 4, pp. 287–295, Feb. 2000. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S000632239900308X

[19] A. Post and M. E. Keck, "Transcranial magnetic stimulation as a therapeutic tool in psychiatry: what do we know about the neurobiological mechanisms?" *Journal of Psychiatric Research*, vol. 35, no. 4, pp. 193–215, Jul. 2001. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0022395601000231

[20] T. Tsoneva, G. Garcia-Molina, and A. Nijholt, "Emotional Brain-Computer Interfaces," *International Journal of Autonomous and Adaptive Communications Systems*, vol. 6, pp. 9–25, Jan. 2013.

[21] Z. He, Z. Li, F. Yang, L. Wang, J. Li, C. Zhou, and J. Pan, "Advances in Multimodal Emotion Recognition Based on Brain–Computer Interfaces," *Brain Sciences*, vol. 10, no. 10, p. 687, Oct. 2020, number: 10 Publisher: Multidisciplinary Digital Publishing Institute. [Online]. Available: https://www.mdpi.com/2076-3425/10/10/687

[22] A. Al-Nafjan, M. Hosny, Y. Al-Ohali, and A. Al-Wabil, "Review and Classification of Emotion Recognition Based on EEG Brain-Computer Interface System Research: A Systematic Review," *Applied Sciences*, vol. 7, no. 12, p. 1239, Dec. 2017, number: 12 Publisher: Multidisciplinary Digital Publishing Institute. [Online]. Available: https://www.mdpi.com/2076-3417/7/12/1239

[23] L.-F. Rodríguez and F. Ramos, "Computational models of emotions for autonomous agents: major challenges," *Artificial Intelligence Review*, vol. 43, no. 3, pp. 437–465, Mar. 2015. [Online]. Available: https://doi.org/10.1007/s10462-012-9380-9

[24] S. Ojha, J. Vitale, and M.-A. Williams, "Computational Emotion Models: A Thematic Review," *International Journal of Social Robotics*, vol. 13, no. 6, pp. 1253–1279, Sep. 2021. [Online]. Available: https://doi.org/10.1007/s12369-020-00713-1

[25] A. Bechara, D. Tranel, H. Damasio, R. Adolphs, C. Rockland, and A. R. Damasio, "Double Dissociation of Conditioning and Declarative Knowledge Relative to the Amygdala and Hippocampus in Humans," *Science*, vol. 269, no. 5227, pp. 1115–1118, Aug. 1995, publisher: American Association for the Advancement of Science. [Online]. Available: https://www.science.org/doi/abs/10.1126/science.7652558

[26] F. D. Raslau, I. T. Mark, A. P. Klein, J. L. Ulmer, V. Mathews, and L. P. Mark, "Memory Part 2: The Role of the Medial Temporal Lobe," *American Journal of Neuroradiology*, vol. 36, no. 5, pp. 846–849, May 2015, publisher: American Journal of Neuroradiology Section: Functional Vignette. [Online]. Available: http://www.ajnr.org/content/36/5/846

[27] M. P. Richardson, B. A. Strange, and R. J. Dolan, "Encoding of emotional memories depends on amygdala and hippocampus and their interactions," *Nature Neuroscience*, vol. 7, no. 3, pp. 278–285, Mar. 2004.

[28] E. A. Phelps, "Human emotion and memory: interactions of the amygdala and hippocampal complex," *Current Opinion in Neurobiology*, vol. 14, no. 2, pp. 198–202, Apr. 2004. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0959438804000479

[29] P. Vuilleumier, “How brains beware: neural mechanisms of emotional attention,” *Trends in Cognitive Sciences*, vol. 9, no. 12, pp. 585–594, Dec. 2005. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1364661305003025

[30] M. Davis and P. J. Whalen, “The amygdala: vigilance and emotion,” *Molecular Psychiatry*, vol. 6, no. 1, pp. 13–34, Jan. 2001, number: 1 Publisher: Nature Publishing Group. [Online]. Available: https://www.nature.com/articles/4000812

[31] A. Öhman, A. Flykt, and F. Esteves, “Emotion drives attention: Detecting the snake in the grass,” *Journal of Experimental Psychology: General*, vol. 130, pp. 466–478, 2001, place: US Publisher: American Psychological Association.

[32] E. Fox, R. Russo, R. Bowles, and K. Dutton, “Do threatening stimuli draw or hold visual attention in subclinical anxiety?” *Journal of Experimental Psychology: General*, vol. 130, pp. 681–700, 2001, place: US Publisher: American Psychological Association.

[33] J. L. McGaugh, “Memory–a Century of Consolidation,” *Science*, vol. 287, no. 5451, pp. 248–251, Jan. 2000, publisher: American Association for the Advancement of Science. [Online]. Available: https://www.science.org/doi/full/10.1126/science.287.5451.248

[34] L. Cahill and M. T. Alkire, “Epinephrine enhancement of human memory consolidation: Interaction with arousal at encoding,” *Neurobiology of Learning and Memory*, vol. 79, no. 2, pp. 194–198, Mar. 2003. [Online]. Available: https://linkinghub.elsevier.com/retrieve/pii/S1074742702000369

[35] L. Cahill, L. Gorski, and K. Le, “Enhanced Human Memory Consolidation With Post-Learning Stress: Interaction With the Degree of Arousal at Encoding,” *Learning & Memory*, vol. 10, no. 4, pp. 270–274, Jul. 2003. [Online]. Available: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC202317/

[36] L. Cahill, R. J. Haier, J. Fallon, M. T. Alkire, C. Tang, D. Keator, J. Wu, and J. L. McGaugh, “Amygdala activity at encoding correlated with long-term, free recall of emotional information.” *Proceedings of the National Academy of Sciences*, vol. 93, no. 15, pp. 8016–8021, Jul. 1996, publisher: Proceedings of the National Academy of Sciences. [Online]. Available: https://www.pnas.org/doi/abs/10.1073/pnas.93.15.8016

[37] D. J.-F. de Quervain, B. Roozendaal, and J. L. McGaugh, “Stress and glucocorticoids impair retrieval of long-term spatial memory,” *Nature*, vol. 394, no. 6695, pp. 787–790, Aug. 1998, number: 6695 Publisher: Nature Publishing Group. [Online]. Available: https://www.nature.com/articles/29542

[38] E. Phelps, K. O’Connor, C. Gatenby, J. Gore, C. Grillon, and M. Davis, “Activation of the left amygdala to a cognitive representation of fear,” *Nature neuroscience*, vol. 4, pp. 437–41, May 2001.

[39] E. K. Miller and J. D. Cohen, “An Integrative Theory of Prefrontal Cortex Function,” *Annual Review of Neuroscience*, vol. 24, no. 1, pp. 167–202, 2001, _eprint: https://doi.org/10.1146/annurev.neuro.24.1.167. [Online]. Available: https://doi.org/10.1146/annurev.neuro.24.1.167

[40] SoP, “Prefrontal Cortex,” Jan. 2017. [Online]. Available: https://www.thescienceofpsychotherapy.com/prefrontal-cortex/

[41] J. R. Stroop, “Studies of interference in serial verbal reactions,” *Journal of Experimental Psychology*, vol. 18, pp. 643–662, 1935, place: US Publisher: Psychological Review Company.

[42] E. T. Rolls, “A theory of emotion and consciousness, and its application to understanding the neural basis of emotion,” in *The cognitive neurosciences*. Cambridge, MA, US: The MIT Press, 1995, pp. 1091–1106.

[43] K. J. Friston, L. Harrison, and W. Penny, “Dynamic causal modelling,” *NeuroImage*, vol. 19, no. 4, pp. 1273–1302, Aug. 2003. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1053811903002027

[44] D. Wang and S. Liang, “Dynamic Causal Modeling on the Identification of Interacting Networks in the Brain: A Systematic Review,” *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 29, pp. 2299–2311, 2021, conference Name: IEEE Transactions on Neural Systems and Rehabilitation Engineering.

[45] R. Underwood, E. Tolmeijer, J. Wibroe, E. Peters, and L. Mason, “Networks underpinning emotion: A systematic review and synthesis of functional and effective connectivity,” *NeuroImage*, vol. 243, p. 118486, Nov. 2021. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S105381192100759X

[46] M. Fastenrath, D. Coynel, K. Spalek, A. Milnik, L. Gschwind, B. Roozendaal, A. Papassotiropoulos, and D. J. F. d. Quervain, “Dynamic Modulation of Amygdala–Hippocampal Connectivity by Emotional Arousal,” *Journal of Neuroscience*, vol. 34, no. 42, pp. 13 935–13 947, Oct. 2014, publisher: Society for Neuroscience Section: Articles. [Online]. Available: https://www.jneurosci.org/content/34/42/13935

[47] P. Gagnepain, J. Hulbert, and M. C. Anderson, “Parallel Regulation of Memory and Emotion Supports the Suppression of Intrusive Memories,” *Journal of Neuroscience*, vol. 37, no. 27, pp. 6423–6441, Jul. 2017, publisher: Society for Neuroscience Section: Research Articles. [Online]. Available: https://www.jneurosci.org/content/37/27/6423

[48] N. E. Nawa and H. Ando, “Effective connectivity within the ventromedial prefrontal cortex-hippocampus-amygdala network during the elaboration of emotional autobiographical memories,” *NeuroImage*, vol. 189, pp. 316–328, Apr. 2019. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1053811919300424

[49] A. P. R. Smith, K. E. Stephan, M. D. Rugg, and R. J. Dolan, “Task and Content Modulate Amygdala-Hippocampal Connectivity in Emotional Retrieval,” *Neuron*, vol. 49, no. 4, pp. 631–638, Feb. 2006. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0896627306000080

[50] A. L. Hodgkin and A. F. Huxley, “A quantitative description of membrane current and its application to conduction and excitation in nerve,” *The Journal of Physiology*, vol. 117, no. 4, pp. 500–544, Aug. 1952. [Online]. Available: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1392413/

[51] M. Breakspear, "Dynamic models of large-scale brain activity," *Nature Neuroscience*, vol. 20, no. 3, pp. 340–352, Mar. 2017, number: 3 Publisher: Nature Publishing Group. [Online]. Available: https://www.nature.com/articles/nn.4497

[52] F. Cao and C. Perfetti, "Neural Signatures of the Reading-Writing Connection: Greater Involvement of Writing in Chinese Reading than English Reading," *PLOS ONE*, vol. 11, p. e0168414, Dec. 2016.

[53] J. Siero, A. Bhogal, and J. Jansma, "Blood Oxygenation Level–dependent/Functional Magnetic Resonance Imaging : Underpinnings, Practice, and Perspectives," *PET Clinics*, vol. 8, pp. 329–344, Jul. 2013.

[54] M. Ramezanian-Panahi, G. Abrevaya, J.-C. Gagnon-Audet, V. Voleti, I. Rish, and G. Dumas, "Generative Models of Brain Dynamics," *Frontiers in Artificial Intelligence*, vol. 5, 2022. [Online]. Available: https://www.frontiersin.org/articles/10.3389/frai.2022.807406

[55] E. Bullmore and O. Sporns, "Complex brain networks: Graph theoretical analysis of structural and functional systems," *Nature reviews. Neuroscience*, vol. 10, pp. 186–98, Mar. 2009.

[56] M. Rubinov and O. Sporns, "Complex network measures of brain connectivity: Uses and interpretations," *NeuroImage*, vol. 52, no. 3, pp. 1059–1069, Sep. 2010. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S105381190901074X

[57] K. Friston, R. Moran, and A. K. Seth, "Analysing connectivity with Granger causality and dynamic causal modelling," *Current Opinion in Neurobiology*, vol. 23, no. 2, pp. 172–178, Apr. 2013. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0959438812001845

[58] Y. Kuramoto, "Chemical Turbulence," in *Chemical Oscillations, Waves, and Turbulence*, ser. Springer Series in Synergetics, Y. Kuramoto, Ed. Berlin, Heidelberg: Springer, 1984, pp. 111–140. [Online]. Available: https://doi.org/10.1007/978-3-642-69689-3_7

[59] B. van der Pol, "LXXXVIII. On "relaxation-oscillations"," *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science*, vol. 2, no. 11, pp. 978–992, Nov. 1926, publisher: Taylor & Francis _eprint: https://doi.org/10.1080/14786442608564127. [Online]. Available: https://doi.org/10.1080/14786442608564127

[60] A. Roebroeck, E. Formisano, and R. Goebel, "Mapping Directed Influence over the Brain Using Granger Causality and fMRI," *NeuroImage*, vol. 25, pp. 230–42, Apr. 2005.

[61] J. Daunizeau, O. David, and K. E. Stephan, "Dynamic causal modelling: A critical review of the biophysical and statistical foundations," *NeuroImage*, vol. 58, no. 2, pp. 312–322, Sep. 2011. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1053811909012488

[62] G. Lohmann, K. Erfurth, K. Müller, and R. Turner, "Critical comments on dynamic causal modelling," *NeuroImage*, vol. 59, no. 3, pp. 2322–2329, Feb. 2012. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1053811911010718

[63] K. Friston, J. Daunizeau, and K. E. Stephan, "Model selection and gobbledygook: response to Lohmann et al," *NeuroImage*, vol. 75, pp. 275–278, Jul. 2013.

[64] G. Lohmann, K. Müller, and R. Turner, "Response to commentaries on our paper: Critical comments on dynamic causal modelling," *NeuroImage*, vol. 75, pp. 279–281, Jul. 2013. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1053811912007707

[65] B. Li, J. Daunizeau, K. E. Stephan, W. Penny, D. Hu, and K. Friston, "Generalised filtering and stochastic DCM for fMRI," *NeuroImage*, vol. 58, no. 2, pp. 442–457, Sep. 2011. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1053811911001406

[66] A. C. Marreiros, S. J. Kiebel, and K. J. Friston, "Dynamic causal modelling for fMRI: A two-state model," *NeuroImage*, vol. 39, no. 1, pp. 269–278, Jan. 2008. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1053811907007070

[67] K. E. Stephan, L. Kasper, L. M. Harrison, J. Daunizeau, H. E. M. den Ouden, M. Breakspear, and K. J. Friston, "Nonlinear dynamic causal models for fMRI," *NeuroImage*, vol. 42, no. 2, pp. 649–662, Aug. 2008. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1053811908005983

[68] R. Buxton, E. Wong, and L. Frank, "Dynamics of blood flow and oxygenation changes during brain activation: The balloon model," *Magnetic resonance in medicine : official journal of the Society of Magnetic Resonance in Medicine / Society of Magnetic Resonance in Medicine*, vol. 39, pp. 855–64, Jun. 1998.

[69] J. B. Mandeville, J. J. Marota, C. Ayata, G. Zaharchuk, M. A. Moskowitz, B. R. Rosen, and R. M. Weisskoff, "Evidence of a cerebrovascular postarteriole windkessel with delayed compliance," *Journal of Cerebral Blood Flow and Metabolism: Official Journal of the International Society of Cerebral Blood Flow and Metabolism*, vol. 19, no. 6, pp. 679–689, Jun. 1999.

[70] K. J. Friston, A. Mechelli, R. Turner, and C. J. Price, "Nonlinear Responses in fMRI: The Balloon Model, Volterra Kernels, and Other Hemodynamics," *NeuroImage*, vol. 12, no. 4, pp. 466–477, Oct. 2000. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S105381190090630X

[71] R. B. Buxton and L. R. Frank, "A model for the coupling between cerebral blood flow and oxygen metabolism during neural stimulation," *Journal of Cerebral Blood Flow and Metabolism: Official Journal of the International Society of Cerebral Blood Flow and Metabolism*, vol. 17, no. 1, pp. 64–72, Jan. 1997.

[72] R. L. Grubb, M. E. Raichle, J. O. Eichling, and M. M. Ter-Pogossian, "The effects of changes in PaCO2 on cerebral blood volume, blood flow, and vascular mean transit time," *Stroke*, vol. 5, no. 5, pp. 630–639, 1974.

[73] A. Anil Meera and M. Wisse, "Dynamic Expectation Maximization Algorithm for Estimation of Linear Systems with Colored Noise," *Entropy*, vol. 23, no. 10, p. 1306, Oct. 2021. [Online]. Available: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC8534782/

[74] "SPM Tutorial #6: Scripting — Andy's Brain Book 1.0 documentation." [Online]. Available: https://andysbrainbook.readthedocs.io/en/latest/SPM/SPM_Short_Course/SPM_06_Scripting.html

[75] K. Stephan, W. Penny, R. Moran, H. den Ouden, J. Daunizeau, and K. Friston, "Ten simple rules for dynamic causal modeling," *Neuroimage*, vol. 49, no. 4, pp. 3099–3109, Feb. 2010. [Online]. Available: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2825373/

[76] F. Kj, K. J, B. B, and R. A, "A DCM for resting state fMRI," *NeuroImage*, vol. 94, no. 100, Jul. 2014, publisher: Neuroimage. [Online]. Available: https://pubmed.ncbi.nlm.nih.gov/24345387/

[77] K. J. Friston, B. Li, J. Daunizeau, and K. E. Stephan, "Network discovery with DCM," *NeuroImage*, vol. 56, no. 3, pp. 1202–1221, Jun. 2011. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S105381191001623X

[78] R. Patro, "Cross Validation: K Fold vs Monte Carlo," Feb. 2021. [Online]. Available: https://towardsdatascience.com/cross-validation-k-fold-vs-monte-carlo-e54df2fc179b

[79] W. D. Penny, K. E. Stephan, A. Mechelli, and K. J. Friston, "Modelling functional integration: a comparison of structural equation and dynamic causal models," *NeuroImage*, vol. 23, pp. S264–S274, Jan. 2004. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1053811904003891

[80] O. David, I. Guillemain, S. Saillet, S. Reyt, C. Deransart, C. Segebarth, and A. Depaulis, "Identifying Neural Drivers with Functional MRI: An Electrophysiological Validation," *PLOS Biology*, vol. 6, no. 12, p. e315, Dec. 2008, publisher: Public Library of Science. [Online]. Available: https://journals.plos.org/plosbiology/article?id=10.1371/journal.pbio.0060315

[81] M. L. Seghier and K. J. Friston, "Network discovery with large DCMs," *NeuroImage*, vol. 68, pp. 181–191, Mar. 2013. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1053811912011780

[82] S. Frässle, E. I. Lomakina, A. Razi, K. J. Friston, J. M. Buhmann, and K. E. Stephan, "Regression DCM for fMRI," *NeuroImage*, vol. 155, pp. 406–421, Jul. 2017. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S105381191730201X

[83] K. J. Friston, K. H. Preller, C. Mathys, H. Cagnan, J. Heinzle, A. Razi, and P. Zeidman, "Dynamic causal modelling revisited," *NeuroImage*, vol. 199, pp. 730–744, Oct. 2019. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1053811917301568

[84] J. Daunizeau, K. E. Stephan, and K. J. Friston, "Stochastic dynamic causal modelling of fMRI data: Should we care about neural noise?" *NeuroImage*, vol. 62, no. 1, pp. 464–481, Aug. 2012. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1053811912004697

[85] A. Destexhe and M. Rudolph-Lilith, *Neuronal Noise*. Springer Science & Business Media, Jan. 2012, google-Books-ID: ZZJcWvg2HUgC.

[86] D. Guo, M. Perc, T. Liu, and D. Yao, “Functional importance of noise in neuronal information processing,” *Europhysics Letters*, vol. 124, no. 5, p. 50001, Dec. 2018, publisher: EDP Sciences, IOP Publishing and Società Italiana di Fisica. [Online]. Available: https://dx.doi.org/10.1209/0295-5075/124/50001

[87] R. B. Stein, E. R. Gossen, and K. E. Jones, “Neuronal variability: noise or part of the signal?” *Nature Reviews Neuroscience*, vol. 6, no. 5, pp. 389–397, May 2005, number: 5 Publisher: Nature Publishing Group. [Online]. Available: https://www.nature.com/articles/nrn1668

[88] A. A. Faisal, L. P. J. Selen, and D. M. Wolpert, “Noise in the nervous system,” *Nature Reviews Neuroscience*, vol. 9, no. 4, pp. 292–303, Apr. 2008, number: 4 Publisher: Nature Publishing Group. [Online]. Available: https://www.nature.com/articles/nrn2258

[89] J. Daunizeau, K. J. Friston, and S. J. Kiebel, “Variational Bayesian identification and prediction of stochastic nonlinear dynamic causal models,” *Physica D: Nonlinear Phenomena*, vol. 238, no. 21, pp. 2089–2118, Nov. 2009. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0167278909002425

[90] K. J. Friston, N. Trujillo-Barreto, and J. Daunizeau, “DEM: A variational treatment of dynamic systems,” *NeuroImage*, vol. 41, no. 3, pp. 849–885, Jul. 2008. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1053811908001894

[91] K. Friston, K. Stephan, B. Li, and J. Daunizeau, “Generalised Filtering,” *Mathematical Problems in Engineering*, vol. 2010, p. e621670, Jun. 2010, publisher: Hindawi. [Online]. Available: https://www.hindawi.com/journals/mpe/2010/621670/

[92] P. Zeidman, A. Jafarian, M. L. Seghier, V. Litvak, H. Cagnan, C. J. Price, and K. J. Friston, “A guide to group effective connectivity analysis, part 2: Second level analysis with PEB,” *NeuroImage*, vol. 200, pp. 12–25, Oct. 2019. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1053811919305233

[93] “SPM12 Software - Statistical Parametric Mapping.” [Online]. Available: https://www.fil.ion.ucl.ac.uk/spm/software/spm12/

[94] K. Friston, “SPM12 Manual,” 2021. [Online]. Available: https://www.fil.ion.ucl.ac.uk/spm/doc/

[95] “NITRC: WFU_pickatlas: Tool/Resource Info.” [Online]. Available: https://www.nitrc.org/projects/wfu_pickatlas/

[96] E. T. Rolls, C.-C. Huang, C.-P. Lin, J. Feng, and M. Joliot, “Automated anatomical labelling atlas 3,” *NeuroImage*, vol. 206, p. 116189, Feb. 2020. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1053811919307803

[97] T. Dahiru, “P – VALUE, A TRUE TEST OF STATISTICAL SIGNIFICANCE? A CAUTIONARY NOTE,” *Annals of Ibadan Postgraduate Medicine*, vol. 6, no. 1, pp. 21–26, Jun. 2008. [Online]. Available: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4111019/

[98] E. Sontag, “Comments on Integral Variant of ISS,” *Systems & Control Letters*, vol. 34, pp. 93–100, May 1998.

[99] C. J. Stam and J. C. Reijneveld, “Graph theoretical analysis of complex networks in the brain,” *Nonlinear Biomedical Physics*, vol. 1, no. 1, p. 3, Jul. 2007. [Online]. Available: https://doi.org/10.1186/1753-4631-1-3

[100] S. Boccaletti, V. Latora, Y. Moreno, M. Chavez, and D. U. Hwang, “Complex networks: Structure and dynamics,” *Physics Reports*, vol. 424, no. 4, pp. 175–308, Feb. 2006. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S037015730500462X

[101] D. Olfati-Saber and R. Murray, “Agreement Problems in Networks with Directed Graphs and Switching Topology,” *Proceedings of the IEEE Conference on Decision and Control*, Jan. 2004.

[102] M. v. d. Ven, “Input-to-State Stability for bilinear systems,” Sep. 2020, publisher: University of Twente. [Online]. Available: https://essay.utwente.nl/83420/

[103] J. M. Steele, *The Cauchy-Schwarz Master Class: An Introduction to the Art of Mathematical Inequalities*. Cambridge University Press, Apr. 2004, google-Books-ID: 7GDyRMrlgDsC.

[104] C. W. J. Granger, “Testing for causality: A personal viewpoint,” *Journal of Economic Dynamics and Control*, vol. 2, pp. 329–352, Jan. 1980. [Online]. Available: https://www.sciencedirect.com/science/article/pii/016518898090069X

Glossary

List of Acronyms

DCSC	Delft Center for Systems and Control
DCM	Dynamic Causal Modeling
fMRI	functional Magnetic Resonance Imaging
sMRI	structural Magnetic Resonance Imaging
CT	Computed Tomography
PET	Positron Emission Tomography
EEG	Electroencephalography
MEG	Magnetoencephalography
BOLD	Blood Oxygenation Level-Dependent
HRF	Hemodynamic Response Function
EM	Expectation Maximization
PFC	Prefrontal Cortex
PTSD	Post-Traumatic Stress Disorder
OCD	Obsessive-Compulsive Disorder
TMS	Transcranial Magnetic Stimulation
DBS	Deep Brain Stimulation
SUD	Substance Use Disorder
VNS	Vagus Nerve Stimulation
AI	Artificial Intelligence
GCM	Granger Causality Mapping
OFC	Orbitofrontal Cortex
PHC	Parahippocampal Cortex
MidFG	Middle Frontal Gyrus

MTL	Medial Temporal Lobe
vmPFC	Ventromedial Prefrontal Cortex
IFG	Inferior Frontal Gyrus
MFG	Medial Frontal Gyrus
Amy	Amygdala
Hip	Hippocampus
DLPFC	Dorsolateral Prefrontal Cortex
SPM	Statistical Parametric Mapping
BMS	Bayesian Model Selection
RFX	Random-Effects
FFX	Fixed-Effects
BPA	Bayesian Parameter Averaging
ANOVA	Analysis of Variance
BMA	Bayesian Model Averaging
SEM	Structural Equation Modeling
BCIs	Brain-Computer Interfaces
DEM	Dynamic Estimation Maximization
NN	Neutral-Neutral
NE	Neutral-Emotional
EE	Emotional-Emotional
SPM12	Statistical Parametric Mapping
GLM	General Linear Model
VOIs	Volumes of Interest
VOI	Volume of Interest
AAL	Automated Anatomical Labelling
TE	Echo Time
TR	Temporal Resolution
ISS	Input-to-State Stability
IISS	Integral Input-to-State Stability
FEP	Free Energy Principle
VFE	Variational Free Energy
GF	Generalised Filtering
GUI	Graphical User Interface
ME	model evidence

List of Symbols

α	Parameters of exact neuronal dynamics
α_i	Grubb's exponent
β	BOLD response confounding effect coefficient
η_a	Prior expectation of elements of matrix A
η_b	Prior expectation of elements of matrix B
η_σ	Expectation of σ
$\eta_{\theta\|y}$	Conditional expectation of θ
η_θ	Prior expectation of θ
γ_i	Rate of flow-dependent elimination
κ_i	Rate of signal decay
κ_i^k	k-th order Volterra kernel
λ	BOLD function nonlinearity
λ_i	Modeling error hyperparameters
$\mathfrak{J}$	Coupling matrix in two-state DCM
ν_a	Prior variance of elements of matrix A
ν_b	Prior variance of elements of matrix B
ν_σ	Variance of σ
ϕ_N	Cumulative normal distribution
ϕ_χ	Cumulative $\chi^2_{l(l-1)}$ distribution
ρ_i	Resting oxygen extraction fraction
σ	Intrinsic decay parameter
τ_i	Hemodynamic transit time
$\tau_z(\sigma)$	Transformed expectation of σ
θ	Parameters of the full model
θ^c	Parameters of the neuronal dynamics model
θ^h	Parameters of the hemodynamic model
ε	BOLD response error
A	Coupling matrix of neuronal dynamics
A'	Augmented A matrix of the neuronal state equation
a_{ij}	Element ij of the coupling matrix A
B^j	Coupling matrix of neuronal dynamics
b_{ij}^k	Element ij of the coupling matrix B^k
C	Coupling matrix of neuronal dynamics
$C_{\theta\|y}$	Conditional covariance of θ
C_θ	Prior covariance of θ
e_{EE}	Eigenvalues of the optimal model for group EE
$e_{\max}$	Largest eigenvalue of the coupling matrix
e_{NE}	Eigenvalues of the optimal model for group NE

e_{NN}	Eigenvalues of the optimal model for group NN
F	Nonlinear function of underlying neuronal dynamics or Free energy optimized in the estimation algorithm
f_i	Inflow
$h(u, \theta)$	Estimated BOLD response
$p(\sigma)$	Probability of σ being negative
$p(\tau_z)$	Transformed probability of the intrinsic decay parameter being negative
$p(e_{\max})$	Probability of the largest eigenvalue being positive
Q_i	Contribution of error covariance components in the i-th region
q_i	Normalized deoxyhemoglobin content
R^2	Model fit metric
s_i	Vasodilatory signal
u	Extrinsic input vector
v_i	Normalized blood volume
X	BOLD response confounding effect signal
x	States of the full model
z	Neuronal state vector