Mechanistic Models of Reward Based Learning and Decision Making for Clinically Motivated Problems

Dissertation
Zur Erlangung des akademischen Grades
Doktor-Ingenieur (Dr.-Ing.)

vorgelegt der Fakultät für Informatik und Automatisierung
Technischen Universität Ilmenau

von: M.Sc. Meltem Sevgi
geboren am: 23. 12. 1985 in Istanbul, Türkei

Gutachter:

1. Prof. Dr.-Ing. habil. Jens Haueisen
2. Dr. Rosalyn Moran
3. Prof. Dr. Tobias H. Donner

Tag der Einreichung: 11. 10. 2016

Tag der wissenschaftlichen Aussprache: 19. 04. 2017

urn:nbn:de:gbv:ilm1-2017000570

Abstract

Mechanistic models of learning and decision making can help test specific hypotheses about observed behavior and brain function. This thesis presents a framework for integrating computational models of adaptive intelligence systems such as Reinforcement Learning and Bayesian Learning algorithms to address clinically motivated problems. In order to provide a comprehensive evaluation, Functional Magnetic Resonance Imaging (fMRI) data were analyzed with behavioral and connectivity models.

In this work, some of the most widely used reinforcement learning algorithms in neuroimaging and psychological studies were evaluated with simulations to understand the behavior of agents under different model parameters and strategies. The models were then tested on a large empirical dataset, and the prediction errors that were derived from the winning model informed the general linear model for fMRI data analysis. Reinforcement learning models were able to capture differences in the function of dopaminergic brain regions and associated behavior in individuals with different genotype. It was further proposed that integrating learning algorithms in effective connectivity models can provide a complementary framework for studying altered brain network dynamics. This was achieved by constructing bilinear and nonlinear dynamic causal models of brain regions involved in reward and prediction error processing.

Finally, hierarchical Bayesian models were implemented to model an agent's learning behavior in a complex, volatile environment. A parallel learning system approach was developed for learning and combining multiple cues by adopting the Hierarchical Gaussian Filter and the precision weighted response model pair. Simulations of parameter recovery suggest that this approach can be used for learning and combining different sources of information. Further, the proposed model was tested on a real dataset and compared against alternative models including an optimal Bayesian learner. This method allowed us to identify individual subprocesses involved in learning from social cues that differ according to the level of autistic traits. We further propose that the experimental and modeling approach presented here can contribute to mechanistic formulations of many psychiatric disorders.

Zusammenfassung

Mechanistische Modelle für Lernen und zur Entscheidungsfindung können helfen, spezifische Hypothesen über beobachtetes Verhalten und dessen Etablierung in Gehirn zu testen. Die hier vorliegende Arbeit bietet einen Ansatz, um computer-gestützte Modelle zum Verstärkungslernen (Reinforcement Learning) und Bayes'sche Lernalgorithmen zu integrieren, und um klinisch motivierte Probleme zu adressieren. Die so entstandenen Modelle wurden anhand von funktionelle Magnetresonanztomographie (fMRT) gemeinsam mit Verhaltens- und Konnektivitätsmodellen evaluiert.

In dieser Arbeit werden vor allem Algorithmen zum Verstärkungslernen betrachtet, typischerweise die im bildgebenden und in psychologischen Studien zur Anwendung kommen. Eine Bewertung der Algorithmen erfolgte mithilfe von Simulationen, um somit das Verhalten von virtuellen Agenten bei unterschiedlichen Modellparametern und Strategien zu verstehen. Später wurden die generierten Modelle an einem empirischen Datensatz getestet, wobei das beste Modell zur Auswertung der fMRI Daten an ein lineares Modell übergeben wurde. Solche Modelle konnten Unterschiede in der Funktion von dopaminergen Gehirnregionen und dem damit assozierten Verhalten zwischen Individuen mit unterschiedlicher genetische Disposition zeigen. Weiterhin wurde untersucht, ob die Einbeziehung von Lernalgorithmen in effektive Konnektivitätsmodelle als komplementäre Grundlage für die weitere Erforschung von veränderten Netzwerkdynamiken im menschlichen Gehirn dienen könnte. Dazu wurden bilineare und nicht-lineare dynamisch kausale Modelle verschiedener Hirnregionen, welche in Belohnungslernen und Vorhersagefehlerprozessen beteiligt sind, erstellt.

In einer Erweiterung, wurden hierarchische Bayes'sche Modelle betrachtet, welche das Lernverhalten eines virtuellen Agenten in einer komplexen und unbeständigen Umgebung modellieren. Ein paralleler Lernansatz wurde zum Lernen und Kombinieren multipler Hinweisreize entwickelt, indem hierarchische Gauss'schen Filter mit präzisionsgewichteten Antwortmodellen gepaart wurden. Simulationen von Parameterschätzungen deuten darauf hin, dass dieser Ansatz zum Lernen und Kombinieren verschiedener Informationsquellen genutzt werden kann. Das vorgeschlagene Modell wurde auf Grundlage empirische Daten überprüft und mit alternativen Modellen verglichen, wie beispielsweise mit dem optimalen Bayes'schen Agenten. Letztlich hat uns diese Methode ermöglicht, individuelle Subprozesse zu identifizieren, die am Lernen von sozialen Hinweisreizen beteiligt sind und die mit unterschiedlichen Ausprägungen vom autistischen Züge variieren. Darüber hinaus postulieren wir, dass der hier vorgestellte experimentelle und modellierende Ansatz zu einer mechanistischen Beschreibung von unterschiedlichen psychiatrischen Störungen beitragen kann.

Acknowledgement

I would like to express my sincere gratitude to people who made this work possible.

I thank Marc Tittgemeyer for believing in me from my first day in Max Planck Institute, always keeping an open mind and an open door to discuss my ideas and to support me to achieve them. I thank Thomas Knösche and Jens Haueisen for helping me formalize my research questions, and their time that they spend for the development of this thesis.

I cannot thank enough Leonhard Schilbach for being such a great mentor and friend, always motivating me when things go downhill. His unique skills made us, a psychiatrist and an engineer, to be able to speak the same language. Special thanks to Jens Brüning for sharing his knowledge of genetics and his energy and passion for science, and his leadership. I thank Markus Ullsperger, it has been a privilege visiting his group in Magdeburg and working together.

I am very grateful to all our collaborators at Biomedical Engineering department in ETH Zürich for the efficient meetings, especially to Klaas E. Stephan for his inputs in fMRI and DCM analysis and his contagious enthusiasm for neuroimaging, and to Andreea O. Diaconescu for her help in adapting Bayesian modeling to our study.

I would also like to thank my colleagues in MPI: to Corina and to the IT department for their software support, to Alexandra for being a great office-mate and also for proofreading this thesis, to Lionel for sharing his knowledge on computational modeling, to Geraldine for supplying Swiss wine in the most critical times, to our Andreas who will always live in our memories, and to other members of TNC group for all the fun and the memories we had together.

Finally, I thank my boyfriend Gavin for his patience and support during my PhD years. I also thank my lovely, caring family in Istanbul. This work would not be possible without their endless support. I thank my father for everything he did for me. I dedicate this thesis to his memory.

Contents

Nomenclature **xiii**

1 Introduction **1**
- 1.1 Motivation 1
- 1.2 Structure of the Thesis 2

2 Background **5**
- 2.1 Reinforcement Learning 5
  - 2.1.1 Markov Decision Processes 5
  - 2.1.2 Temporal Difference Learning 7
  - 2.1.3 Q-Learning 8
  - 2.1.4 Softmax Action Selection 8
  - 2.1.5 Reinforcement Learning In the Brain 9
    - 2.1.5.1 Neural Correlates of Reward and Punishment 9
    - 2.1.5.2 Prediction Error Processing in the Brain 10
    - 2.1.5.3 Genetic Influences on Reinforcement Learning 14
- 2.2 Bayesian Learning and Decision Making 15
  - 2.2.1 Cue Combination 15
  - 2.2.2 Bayesian Decision Theory 17
  - 2.2.3 Hierarchical Gaussian Filter (HGF) 18
  - 2.2.4 Bayesian Brain 20
- 2.3 Functional Magnetic Resonance Imaging 22
  - 2.3.1 BOLD signal 22
  - 2.3.2 fMRI Data Analysis 23
  - 2.3.3 Brain Connectivity 25
  - 2.3.4 Dynamic Causal Modeling 26
  - 2.3.5 Inference on Model Space 29

3 Functional MR Imaging of Reinforcement Learning **31**
- 3.1 Introduction 31
- 3.2 Methods 33
  - 3.2.1 Reinforcement Learning Models 33
  - 3.2.2 Simulations 33
  - 3.2.3 Participants 34
  - 3.2.4 DNA isolation and SNP genotyping 34
  - 3.2.5 Reinforcement Learning Task 35
  - 3.2.6 Statistical Analysis 35
  - 3.2.7 Parameter Estimation and Model Comparison 35
  - 3.2.8 fMRI Acquisition 37
  - 3.2.9 fMRI Data Analysis 37

3.2.10 Volume of Interest Analysis 38
3.3 Results 39
3.3.1 Simulations 39
3.3.2 Behavioral Results 39
3.3.3 Reinforcement Learning Model Comparison 40
3.3.4 Whole Brain Activations 40
3.3.5 Volume of Interest Analysis 40
3.3.6 Prediction Error Processing and Avoidance Learning 41
3.4 Discussion 41

4 Dynamic Causal Modeling of Reinforcement Learning Circuitry in the Brain 51
4.1 Introduction 51
4.2 Methods 52
4.2.1 Model Construction 52
4.2.2 Time Series Extraction 52
4.2.3 Model Space 53
4.2.4 Bayesian Model Selection and Bayesian Model Averaging 54
4.3 Results 55
4.3.1 Model Selection 55
4.3.2 Effective Connectivity Strengths in an RL Network 56
4.3.3 Genotype, Connectivity and Behavior 56
4.4 Discussion 57

5 Hierarchical Bayesian Modeling to Study Differences in Learning 62
5.1 Introduction 62
5.2 Methods 64
5.2.1 Parallel Learning Systems 64
5.2.2 Precision Weighted Response Model 66
5.2.3 Participants 67
5.2.4 The Experiment 67
5.2.5 Simulations 69
5.2.6 Model Inversion 70
5.2.7 Model Space and Model Selection 70
5.2.8 Multivariate Regression 71
5.2.9 Other Behavioral Measures 72
5.3 Results 72
5.3.1 Simulations 72
5.3.2 Empirical Dataset 73
5.3.3 Construct Validity 74
5.4 Discussion 77

6 General Discussion 83

7 Conclusion 89
7.1 Summary and Outlook 89
7.2 Contributions 91

List of Algorithms 93

List of Figures **95**

List of Tables **97**

Bibliography **99**

Erklärung **114**

Nomenclature

List of Abbreviations

ACC	Anterior cingulate cortex
BMC	Bayesian model comparison
BOLD	Blood oxygenation level dependent
DA	Dopamine
DBS	Deep brain stimulation
DCM	Dynamic causal model
FFX	Fixed effect analysis
fMRI	Functional magnetic resonance imaging
GLM	General linear model
HGF	Hierarchical Gaussian filter
HRF	Hemodynamic response function
ICA	Independent component analysis
MEG	Magnetoencephalography
mPFC	Medial prefrontal cortex
NAcc	Nucleus accumbens
OFC	Orbitofrontal cortex
OTO	Observing the observer
PCA	Principle component analysis
PE	Prediction error
PET	Positron emission tomography
RFX	Random effect analysis
RL	Reinforcement learning
SN	Substantia nigra

TD	Temporal difference
vStr	Ventral striatum
VTA	Ventral tegmental area

List of Symbols

α	Learning rate
β	Decision temperature
δ	Prediction error
κ	Coupling of 3rd level to 2nd level
$\mu_{1,card}$	Posterior expectation of card accuracy
$\mu_{1,gaze}$	Posterior expectation of gaze accuracy
ω	variance parameter for 2nd level
π	Precision (inverse variance)
ζ	Weight on the precision of gaze accuracy
$b^{(t)}$	Belief at trial t
$m^{(p)}$	Perceptual model
$m^{(r)}$	Response model
Q	Action value
B0	Static magnetic field
T1	Spin-lattice relaxation time
T2	Spin-spin relaxation time

1 Introduction

1.1 Motivation

Models are mathematical representations of processes. In this thesis, the focus will be on one type of model: **Learning models**. They explain which actions are taken and which strategies are followed by an agent in certain situations. In the case of learning with feedback, these steps can be explained with well known algorithms such as reinforcement learning and Bayesian learning. It is of great scientific interest to discover the neuronal architecture involved in learning and decision making. Therefore, another type of model that will be discussed and used in conjunction with learning models in this thesis are **connectivity models** that explain the influence of neuronal activity in one brain region on another region in terms of connectivity strengths. These two kinds of models are state-of-the-art methods in understanding altered mechanisms in many disorders. For that reason, it is important to assess them to provide perspectives in various clinical problems.

Dopamine is a neurotransmitter that modulates reward signalling in the brain, influencing our behaviour in response to rewarding stimuli. It is involved in learning from feedback (Schultz et al., 1997). Reinforcement learning algorithms have been adopted to explain dopaminergic function. Therefore, they are of increasing interest in understanding many disorders including addiction, Parkinson's disease, and obesity, where the brain's reward system is impaired. Genetic influences underlying obesity have not been investigated from a reinforcement learning perspective. In this thesis, I will present how these models can be integrated to bring insights to these under-explored questions at the intersection of neurology and endocrinology. For example, specific hypothesis regarding the effects of certain genes on learning and connectivity parameters will be tested. Simulations will compare an agent's learning behavior under different strategies by adopting the most commonly used reinforcement learning models with varying parameters such as learning rate and inverse temperature parameter, and different initial action values. More importantly, they will be tested on a large, real dataset. The prediction errors that are derived from reinforcement learning models can be included in the general linear model for fMRI data analysis. This method will be presented to test the power of learning models in capturing differences in dopaminergic brain function and associated behavior.

Another important foundation of learning is the synaptic plasticity achieved through neurotransmission. Information flow during learning in the brain can be tracked with connectivity methods. Although this is a challenging concept due to technical constraints of imaging modalities, functional connectivity methods can provide statistical dependencies among interacting brain regions. **Dynamic Causal Modeling** (DCM) describes the interactions in a set of brain regions in terms of their effective connections. These type of models are named **effective connectivity models** due to their power in representing the network in terms of the directional influences. During associative learning, changes in synaptic plasticity can be captured with bilinear and nonlinear dynamic causal models between auditory and visual areas (den Ouden et al., 2009), and between cortical and motor areas (den Ouden et al., 2010), respectively. This work will present a framework to combine prediction errors derived from a reinforcement learning model with bilinear and nonlinear DCMs and to map learning behavior onto brain connectivity.

Finally, the features of the environment can influence an agent's learning and decision making processes. When the environment is uncertain, learning and perceptual estimation becomes suboptimal (Landy et al., 2007), i.e it becomes harder for the agent to learn the actions that lead to maximum reward. Cue combination studies have tried to address methods of combining different sources of information that guide an agent's actions. While linear cue combination proposes a weighting of cues based on their reliabilities, Bayesian formulations have also incorporated prior knowledge (Fig. 1.1). Some state-space models, such as the "observing the observer" (Daunizeau et al., 2010b) approach use generative models to explain a perceptual inference process. One example to this is a generative model **Hierarchical Gaussian Filter** (HGF) that describes the relationship between an agent's beliefs and its environment (Mathys et al., 2011). Unlike optimal Bayesian learner models, the HGF can model individual learning trajectories and derive subject specific learning and decision making parameters (Iglesias et al., 2013; Mathys et al., 2014; Diaconescu et al., 2014). In clinical applications, it is important to investigate pathological perceptual mechanisms such as in schizophrenia, autism, and depression. Theories of predictive coding suggest a failure in adapting prediction errors causing alterations in learning and decision making (Friston, 2016). In this thesis, hierarchical Bayesian modeling will be implemented in learning from multiple cues to provide a method for identifying individual differences which can be influenced by personal traits.

In summary, I will show that computational models of learning can answer many clinical questions such as altered behavioral and brain responses in obesity and autism, as they can provide a means of understanding these processes. Neuroscientists and physicians can do quantitative hypothesis testing by integrating learning models into their research. In this thesis, some of the most promising modeling approaches will be evaluated with simulations and model inversion diagnostics, tested on large datasets, compared with possible alternative models, and combined with fMRI data and effective connectivity models to present a comprehensive framework for many neuroimaging applications.

1.2 Structure of the Thesis

This thesis is organized as follows: First, **Chapter 2** introduces the reader to theoretical background necessary for the concepts covered in this thesis. The first section describes the reinforcement learning (RL), its markovian properties, stochastic action selection, as well as the neural representation of the elements of the RL such as processing of positive and negative feedback, action values, and prediction errors. The second section includes the rationale for the need of Bayesian approaches to model learning and decision-making, the mathematical backgrounds of Bayesian reasoning, an overview of its applications to human learning, and explaining the HGF that will be implemented later in the thesis. The subsection about cue

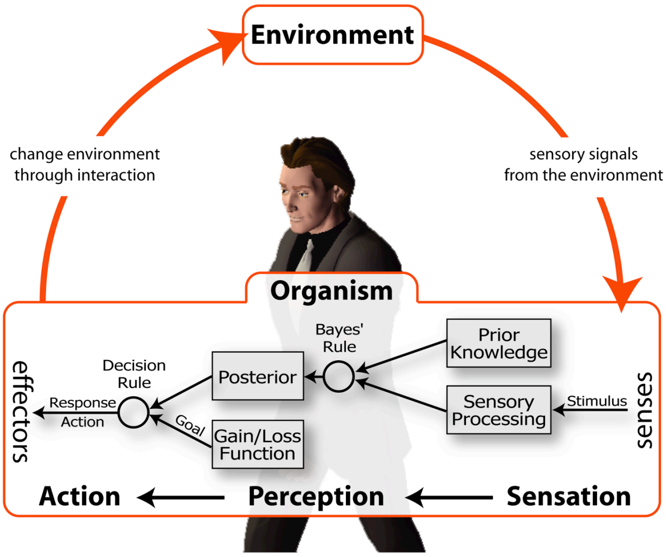


Figure 1.1: Interaction of the organism with the environment through its senses. According to Bayesian theories of learning, the organism combines the sensory information with the prior knowledge to form a posterior belief which then leads to a decision (from Ernst and Bülthoff (2004)).

integration gives an introduction to some of the state-of-the-art Bayesian methods for combining different sources of information in the environment. The third section of this chapter describes principles of functional magnetic resonance imaging (fMRI) signal and analysis, functional, and effective connectivity methods. The process for evaluating the probability distribution of a hypothesis is known as **statistical inference**. The background chapter will also introduce Bayesian inference since it is used for statistical inference of Bayesian models that are presented throughout this thesis.

Chapter 3 evaluates and compares two reinforcement learning algorithms with simulations and an empirical dataset. Learning behavior of an agent with different parameters, exploration-exploitation, convergence of action values to the real values, and the effect of having two different learning rates are shown with simulations. Then the models are compared on a real dataset. Finally, functional MRI data is analyzed with prediction errors derived from the winning model. This chapter further demonstrates that genetic variance can affect dopamine dependent midbrain responses and learning from negative outcomes in humans.

Chapter 4 focuses on modeling of effective connectivity among the brain regions that are involved in the implementation of reinforcement learning such as processing of reward and prediction error. A model space with bilinear and nonlinear dynamic causal models (DCMs) allowed to test interaction dynamics and gating mechanisms within the reward circuitry during a learning task. Similar to Chapter 3, genotype information of the participants was included in the post-hoc analysis to understand how alterations in dopamine genetics system influences connectivity strengths. Chapter 3 and 4 present a novel application of genetics, connectivity and algorithmic models to study obesity.

Chapter 5 is dedicated to the evaluation of Bayesian learning algorithms in combining multiple cues. The hierarchical Gaussian filter and the precision weighted response model are adopted to simulate an agent who learns from multiple sources of information in a volatile environment. The model pair is also tested on a real dataset and compared with alternative models. Finally, this approach is validated for predicting autistic traits related differences of social cognition.

Chapter 6 provides general discussions with the technical and biological limitations of the studies presented here, and the questions that were addressed in this thesis.

Finally, **Chapter 7** provides conclusions from this thesis with a summary of the studies, as well as the future directions that are important for translating present studies into clinical research and practice.

2 Background

2.1 Reinforcement Learning

Reinforcement Learning (RL) is an area in machine learning, which addresses the problem of an agent's interaction with its environment to learn to take actions that maximize the future reward. Depending on the problem, an agent can be human, robot, an autonomous helicopter or even a factory. Although the term was used previously in learning systems, Richard S. Sutton and Andrew G. Barto created the area of RL in 1979 while developing adaptive intelligent systems that could change their behavior according to the environment. As stated by Sutton (Sutton, 1992), all RL can be seen as “reverse engineering of certain psychological processes”. RL research includes many applications in control theory, artificial intelligence, dynamic programming, and neuroscience. One of the most intuitive and earliest examples is a multi-armed bandit task: A scenario of multiple slot machines (or one-armed bandits) where one arm returns a higher reward than the other arms and the agent needs to learn this while facing the problem of exploitation (stick with the arm with a high payoff) versus exploration (try the other arms despite little information).

RL algorithms are formalised as Markov Decision Processes (MDPs) in which the agent does not know the reward function and the transition probabilities explicitly. Therefore, this section will start with an overview of MDPs and then it will examine a special type of RL algorithm called temporal difference learning which is a general case of Q-Learning. RL models will be implemented later in Chapter 3 on a human probabilistic learning problem. The link between the mechanistic formulation of RL and the biology will be explained in the following subsections.

2.1.1 Markov Decision Processes

A Markov Decision Process (MPD) consists of these elements:

A set of finite number of states, or state space (S), a set of actions (A), state transition function $T(s, a, s')$, and the immediate reward that agent receives when arriving at new state s', $r(s, a, s')$.

As the name suggests, it has a Markovian property: At any time point t of the decision process, the next state s_{t+1} depends only on the current state s_t: $P(s_{t+1}|s_t) = P(s_{t+1}|s_t, s_{t-1}, ..., s_1)$. When the actions are non-deterministic the state transition function defines the probability of an agent's arrival in the new state s' after taking the action a when in state s, $P(s'|s, a)$. Now, the goal of the agent is maximizing the future reward which translates the problem to so called policy finding, $\pi(s)$ when in state s. A policy is a sequence of actions that fully defines the agent's behavior. An optimal policy is the policy that maximizes the reward in the equation. Most MDPs use a discounting parameter γ to discount future rewards. In this case, the cumulative discounted reward that the agent will receive is

$$V^{\pi}(s_t) = r_t + \gamma r_{t+1} + \gamma^2 r_{t+2} + ... = \sum_{i=0}^{\infty} \gamma^i r_{t+i}, \tag{2.1}$$

where $0 \leq \gamma < 1$. This approach also fits to animal or human behavior where the agent typically prefers immediate rewards over future ones so that any reward at time step n will be discounted with a rate of γ^n. When $\gamma = 1$, the process becomes a non-discounted MDP. As γ approaches 0, the agent discounts future rewards more. For example, think of a robot agent in a grid world example as in Fig. 2.1. The robot is initially in the state in the left bottom corner and the goal is to reach the docking station in the top right corner. The arrows show one optimal policy for this agent denoted by π^*

$$\pi^* = \arg\max_{\pi} V^{\pi}(s). \tag{2.2}$$

If we assign the discount factor of 0.8 and a numerical value to charging say 100, then according to the Eq. 2.1 the state-value function under the optimal policy will be $V^*(s) = 0 + (0.8)0 + (0.8)^2 100 = 64$.

In general, the optimal policy can be defined in terms of the actions that maximizes rewards plus the state-value function

$$\pi^* = \arg\max_{a} [r(s,a) + \gamma V^*(T(s,a,s'))]. \tag{2.3}$$

If the state value function is known, then the optimal policy can be found, e.g. backward recursion, brute force. To see the recursive property of V^*, Eq. 2.1 can be rewritten

$$V^*(s_t) = r_t + \gamma[r_{t+1} + \gamma r_{t+2} + ...] = r_t + \gamma V^*(s_{t+1}) \tag{2.4}$$

This is known as Bellman equation (Bellman, 1957). Note that the state-value function can be calculated when the transition function $T(s,a,s')$ is known to the agent. Therefore, Bellman equation provides a solution when the agent has perfect knowledge of the environment. This optimization forms the basis of Dynamic Programming (Bertsekas et al., 1995) such that one writes the value function recursively to get rid of future terms that are not avail-

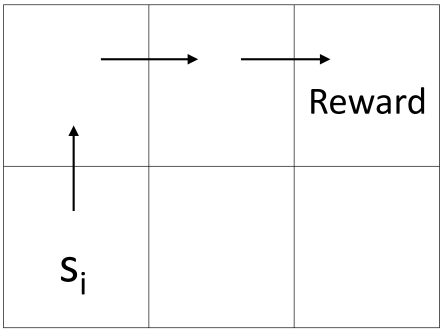


Figure 2.1: An example grid world for an agent in the state s_i. Arrows show one optimal policy to obtain the reward.

able at the time. However, in real life scenarios the agent lacks knowledge of either reward function or transition function, or both. Therefore, it is not possible to define an optimal policy directly by maximizing those two functions. The next section will explain temporal difference learning as a solution to this.

2.1.2 Temporal Difference Learning

Although Temporal Difference (TD)-learning is a general prediction approach, this section will explain it in the context of RL. Introduced by Sutton (1988), TD learning is a general method to predict future value of a state by updating the estimated value of the current state. The optimal policy is found by approximating the state value function. In each consecutive state the agent calculates the error between the prediction $V(s_t)$ and the value of the actual observed state $r_{t+1} + V(s_{t+1})$. This error is used to update the estimated value of current state $V(s_t)$ such that

$$V(s_t) \leftarrow V(s_t) + \alpha[r_{t+1} + V(s_{t+1}) - V(s_t)], \tag{2.5}$$

where α is the learning rate which decides how much weight will be on the new piece of information in this new state s_{t+1}. In many RL problems, one can save the values of V as a lookup table or replace the table with a function approximation (Shi, 2011). The error term $r_{t+1} + V(s_{t+1}) - V(s_t)$ is also known as the temporal difference error. Estimations in the later stages of the decision process will be closer to real values so that error is reduced during learning. This is also known as the simplest TD method or TD(0) because it uses only the prediction of the next state value to update the current state value.

Algorithm 1 TD(0)

```
Initialize state value $V(s) \leftarrow 0$
repeat
    Take action a
    Receive reward r
    Observe new state $s'$
    Update the state value estimate
    $V(s) \leftarrow V(s) + \alpha[r + \gamma V(s') - V(s)]$
    $s \leftarrow s'$
until s terminates
```

For convergence of $TD(\lambda)$ for any value of λ, the reader is referred to the literature (Dayan, 1992).

So far for simplicity we assumed deterministic relationships. In nondeterministic environments, i.e. where the state action transition function and the reward functions are probabilistic, the state value function $V^{\pi}(s_t)$ becomes the expected value of future rewards:

$$V^{\pi}(s_t) = E[\sum_{i=0}^{\infty} \gamma^i r_{t+i}] = \gamma \sum_{s'} P(s_{t+1} = s' | s_t = s, a_t = a) V^{\pi}(s') \tag{2.6}$$

2.1.3 Q-Learning

Q-learning is a special case of TD-learning where the agent selects the action a_t that maximizes $V(s_{t+1})$ while in state s_t. This action specific description of the expected total reward is defined with the action-value function

$$Q^{\pi}(s_t, a_t) = r(s_t, a_t) + \gamma V^{*}(s_{t+1}). \tag{2.7}$$

We can see the relationship between state-value function and the action-value function by comparing equation 2.1 to the equation above. To find the optimal policy, the quantity $Q(s_t, a_t)$ needs to be maximized. Now the agent does not need the knowledge of the reward function and the values of the future states, but it only needs to choose the specific action that maximizes the reward in the consecutive state s_{t+1}.

Algorithm 2 Q-Learning

```
Initialize value of each state $Q(s,a) \leftarrow 0$
repeat
    Take action a
    Receive reward r
    Observe new state $s'$
    Update the estimated action value of the state
    $Q(s,a) \leftarrow Q(s,a) + \alpha[r + \gamma \underset{a'}{max} Q(s',a') - Q(s,a)]$
    $s \leftarrow s'$
until s terminates
```

For brevity, we will not go through convergence properties. For a proof of convergence of Q-Learning, please see Melo (2001).

2.1.4 Softmax Action Selection

At first glance one would expect the agent to choose the action with the largest rewarding value. However, there is a drawback in this approach: Since the agent is learning while executing an action, always choosing the same action which yielded the largest reward in the early stages of learning will cause other actions to remain unexplored. In RL, this is known as the *exploration-exploitation dilemma.* A general solution to this dilemma is to introduce softmax function which converts action values $Q(s, a)$ to probabilistic values of executing that action $P(s, a)$

$$P(s,a) = \frac{e^{Q(s,a)/\beta}}{\sum_{i=1}^{n} e^{Q(s,a_i)/\beta}} \tag{2.8}$$

where β is temperature parameter. Big values of β will assign each action with similar probabilities, whereas small values will result in the highest probability for the action with the highest reward, hence the agent will follow a more deterministic action selection.
This equation stems from Boltzmann distribution in statistical mechanics. It has many application areas from explaining foraging behavior in bees, (Niv et al., 2002) to image classification problems in the form of an activation function in the last layer of an artificial neural network. This is also relevant in neuronal models where a neuron fires when the threshold is exceeded.

Softmax function is the basis of the response models that will be implemented in this thesis. We will go through a classical application of this function in chapter 3 where the subjects assign probability values for the action values of each stimulus pair. Later on, in chapter 5 we will see a modified version of softmax function where the volatility of the environment will influence this transformation, which we will call precision weighting response model.

2.1.5 Reinforcement Learning In the Brain

This section will introduce the brain regions and the biological processes which are key in learning from reward and punishment. This has been a central question for many researchers. Computational models are at the heart of understanding these processes. Advances in neuroimaging methods in the last decade have made it possible to combine these computational models with imaging techniques and analyze the imaging data with model derived parameters to explain underlying cognitive processes. This section will give some insights about what has been accomplished so far by introducing some important studies in the field.

2.1.5.1 Neural Correlates of Reward and Punishment

Rewards are positive reinforcers for an animal to learn about its environment and take appropriate actions in order to survive. Similarly, punishments decrease the probability of a behavior. Reinforcers can be primary such as food and water or secondary such as money. In behavioral psychology, there are two classes of conditioning: In Pavlovian or classical conditioning, reinforcers follow the conditioned stimulus so that once the associations are formed between the two, an unconditioned response follows the conditioned stimulus. While in Pavlovian conditioning, the conditioned stimuli are independent from the animal's actions (Schultz and Dickinson, 2000), in instrumental conditioning, animal's actions determine the type of reinforcement it will receive.

The neurotransmitter dopamine is involved in processing rewarding stimuli. Dopaminergic neurons are mainly found in the substantia nigra (SN) and ventral tegmental area (VTA). Projections of these neurons into different brain structures form dopaminergic pathways: The DA neurons in VTA project to the prefrontal cortex and nucleus accumbens (NAcc), which is called the mesolimbic pathway. Projections from SN to the caudate, and dorsal putamen form the nigrostriatal pathway. Connections from VTA to frontal cortex are the mesocortical pathway. These connections of midbrain DA cells integrate information in different domains (Haber, 2014). The reward circuit in the brain is embedded within the cortico-basal ganglia system: There is a strong dopaminergic input from midbrain to ventral striatum (vStr). vStr projects back to midbrain and to the ventral pallidum and also projects to the orbitofrontal cortex (OFC) and anterior cingulate cortex (ACC). Midbrain sends inputs to prefrontal cortex through thalamus (Haber and Knutson, 2010) (Fig. 2.2).

RL provides a framework to a mechanistic understanding of reward based learning. Next section will present the similarities between RL and the neuronal processes that are discussed here.

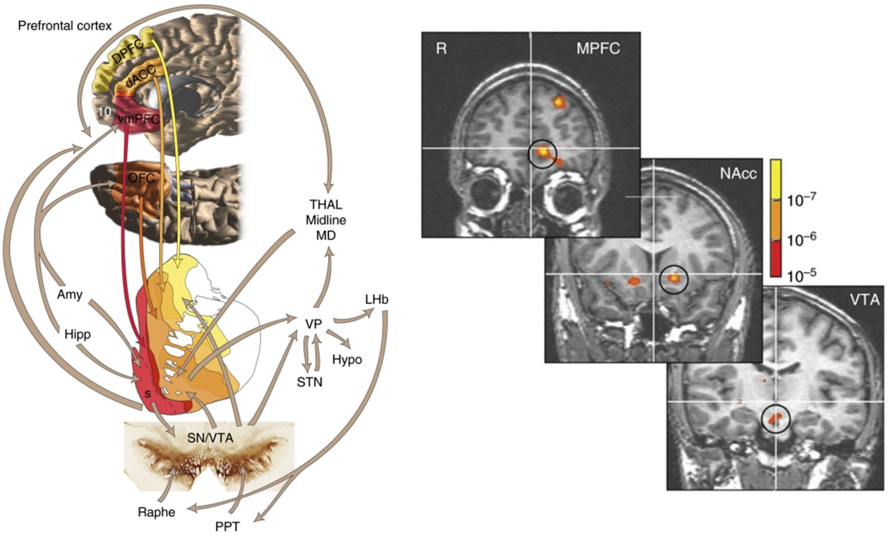


Figure 2.2: Projection sites and brain structures that are important in processing reward (left). Activations in medial prefrontal cortex (mPFC) , NAcc, and VTA during reward expectation in a monetary task (right). These structures form ventral cortico-basal loop (from (Haber and Knutson, 2010)).

2.1.5.2 Prediction Error Processing in the Brain

Bush and Mosteller (1951) introduced an RL framework to classical conditioning. They formalized the associative strength between the conditioned stimulus and the unconditioned stimulus in mathematical terms. Rescorla and Wagner (1972) extended this approach for a variety of classical or Pavlovian conditioning arrangements.

A major distinction between Rescorla-Wagner and TD learning is that TD is a real-time learning system which estimates the future reward without waiting until all the outcomes are observed, as first proposed by Sutton and Barto (1998). TD treats both conditioned cue and the unconditioned cue in the same way. On the other hand, in Rescorla-Wagner, associative strengths are built at the conditioned stimulus (CS) presentation and prediction error (PE) is calculated at the onset of the unconditioned stimulus (US) (Chase et al., 2015a).

In the well known experiment by Schultz Schultz et al. (1993), electrophysiology data gave some first insights about the dopaminergic activity in the midbrain to rewarding stimuli. The association between DA depletion and the impaired cognition was previously known as a result of lesions in frontal cortex (Brozoski et al., 1979) or dysfunctioning basal ganglia in Parkinson's patients (Cools et al., 1984). In the experiment, an awake monkey received apple juice if he reaches and presses the lever on the left side after the presentation of a start cue (Fig. 2.3). Simultaneous recording of midbrain dopaminergic neurons showed that there was no response at the onset of the start cue early in the experiment, but the response is observed at the juice delivery. Later in the experiment, as the monkey learned to predict receiving juice, the response was observed following the start cue rather than the juice delivery itself. They interpreted these results in terms of basic attentional and motivational processes underlying cognitive behavior. However, following experiments showed that these activities can be better explained by TD learning theories (Schultz et al., 1997). Indeed, the midbrain activation resembled a reward PE. Interestingly, when the reward (juice) is omitted, firing rates of dopamine neurons decreased below the baseline (Fig. 2.3).

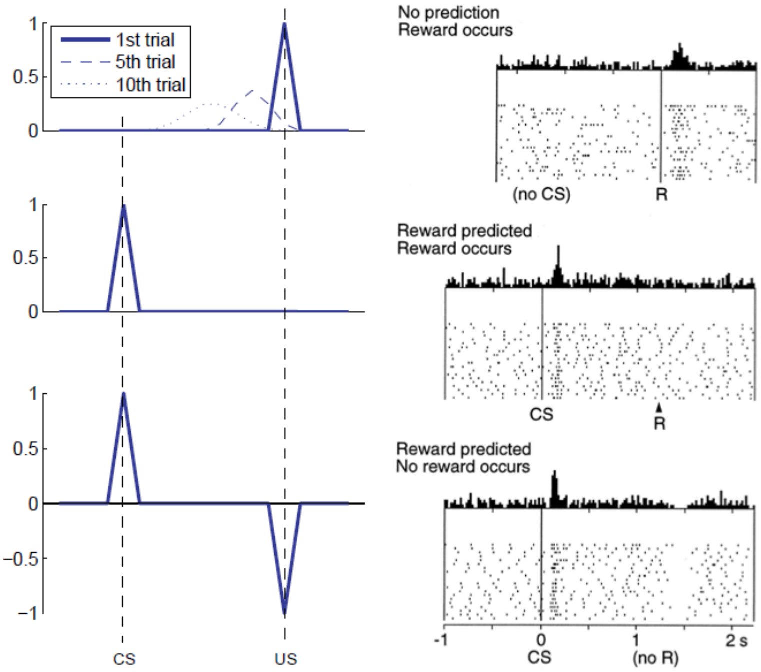


Figure 2.3: Midbrain DA neurons responses resemble PEs in TD learning. Electrophysiological recordings from monkeys during conditioning (right) showed that early in the experiment dopaminergic neurons respond at the onset of the reward. Once the animal learns the predictive value of CS that is a lever touch, there is no response observed at the reward onset, but response occurs at the onset of CS. When the reward is omitted, a depression is observed in the DA acitivity(from (Schultz et al., 1997)). The plots on the left show corresponding PE magnitudes at the US onset in a [-1,1] range depending on the event type: If reward is not expected PE is positive (top), when the reward is fully predicted, PE is 0 (middle). When the reward is omitted the PE is negative (bottom), (from (Niv, 2009))

Human fMRI studies support the electrophysiololgy findings. Predictability of a primary reward modulated activation of the striatum, a target of midbrain dopaminergic neurons (Berns et al., 2001; McClure et al., 2003). Striatum shows differential activations to rewarding and punishing feedback: A punishment feedback decreased the activation below baseline whereas a rewarding outcome sustained activation (Delgado et al., 2000). In a monetary incentive task with different probabilities of winning money, NAcc activity increased linearly with the reward probability during the expectation period, it coded PE during the outcome period (Abler et al., 2006). Also vStr represented expected value of a stimulus that is the reward magnitude times the reward probability, during anticipation, and the PE signal that is the difference between the actual outcome and the expected value, at the reward delivery (Yacubian et al., 2006).

Applications of temporal difference learning in fmri studies supported that these computations are performed in dopaminergic target regions. PE signal is encoded in vStr and OFC before learning at the time of presentation of reward, but after learning this activity shifted to the onset of the CS (O'Doherty et al., 2003). Bayer and Glimcher (2005) used regression analyses to predict midbrain dopamine activity and observed that dopamine firing rates increase when the reward is more than weighted average of previous rewards, but decrease when the current reward is significantly less than the weighted average of previous rewards suggesting that midbrain is encoding PE when the signal has a positive value. Monetary reinforcement schedules influenced striatal DA transmission as shown by using radioligand Positron Emission Tomography (PET) (Zald et al., 2004). Furthermore, Schonberg et al. (2007) showed that PE responses in striatum correlated with the behavioral performance and differentiate learners from non-learners.

Action value coding is an important function of the basal ganglia. Different modalities have identified cortical and sub-cortical regions that appear to play a role in reward based learning and decision making. Under an RL algorithm, recordings of striatal neurons revealed that these neurons represent action values and predict choice probability of actions (Samejima et al., 2005). In another fMRI study, OFC activity was correlated with reward magnitude, and the midbrain and vStr activity was correlated with TD prediction errror (Rolls et al., 2008). It is more likely that OFC encodes expected outcomes given the difficulties that OFC / vmPFC damaged patients have in decision making (Bechara et al., 2000; Camille et al., 2004). Source-reconstruced MEG data analyzed using a biophysically plausible network model identified ventromedial prefrontal cortex (vmPFC) as a region that performs value comparison during value guided decision making (Hunt et al., 2012). However, others found that distinct roles of learning and decision making process are attributed to different brain regions: Goal values are correlated with the activity in medial OFC, decision values are correlated with lateral OFC and the PEs are correlated with the vStr activity (Hare et al., 2008).

Although the results from these neurophysiology studies are correlational (Schultz, 2010), findings from optogenetics and pharmacological studies provided a causal link about the role of DA in reinforcement learning. The first causal evidence between DA, brain activity, and behavior in humans came from an fMRI study with a drug adminisitration. Enhancement of dopaminergic activity by L-DOPA (levodopa) caused a reduced reward PE expressed in striatum and a greater propensity to choose the most rewarding action compared to subjects treated with haloperidol which results in decreased dopaminergic function (Pessiglione et al., 2006). Furthermore, authors could demonstrate the behavioral patterns under different drug conditions by applying a standard action-value learning algorithm.

Advances in optogenetics provided causal and temporally precise control of dopaminergic activity. Optogenetic stimulation of VTA neurons identified the dopaminergic neurons as signaling reward PEs, and GABAergic neurons as signaling expected reward (Cohen et al., 2012). Action potential firing in stimulated dopaminergic neurons of VTA mediated behavioral conditioning suggesting that dopamine neuron activation alone is sufficient to provoke reward related behavior (Tsai et al., 2009).

Aberrant reward and PE processing have been reported in numerous clinical cases. Corticostriatal activity was diminsihed during reward anticipation and mPFC activity was reduced during reward outcome processing in patients with binge eating disorder (BED) (Balodis et al., 2014). Parkinson's patients differed in behavioral performance depending on being on- or off medication (Frank et al., 2004). SN responses were lower in Parkionson's patients to negative outcomes during deep brain stimulation (DBS) compared to positive outcomes (Zaghloul et al., 2009). Also reduced PE responses were found in striatum and midbrian in schizophrenia patients (Gradin et al., 2011). Therefore, the neural and behavioral effects of treatments e.g. antipsychotic drugs can be monitored by changes in the PE processing.

Two seperate basal ganglia pathways have been proposed to mediate rewarding and aversive learning behavior (Freeze et al., 2013), 'direct' and 'indirect' pathways, which are associated with different DA receptor types, D1 and D2, respectively (Fig. 2.4). Activation of the indirect pathway in mice elicited a Parkinsonian state with decreased locomotor initiations, whereas modulating the direct pathway increased locomotion (Kravitz et al., 2010). Stimulation and blocking of D1 and D2 receptors in the mPFC resulted in different patterns of behaviors in risk-based decision making (Onge et al., 2011), most interestingly D2 stimulation impaired decision making. Further, while D2 blockade increased preference for a risky choice, D1 blockade decreased this bias.

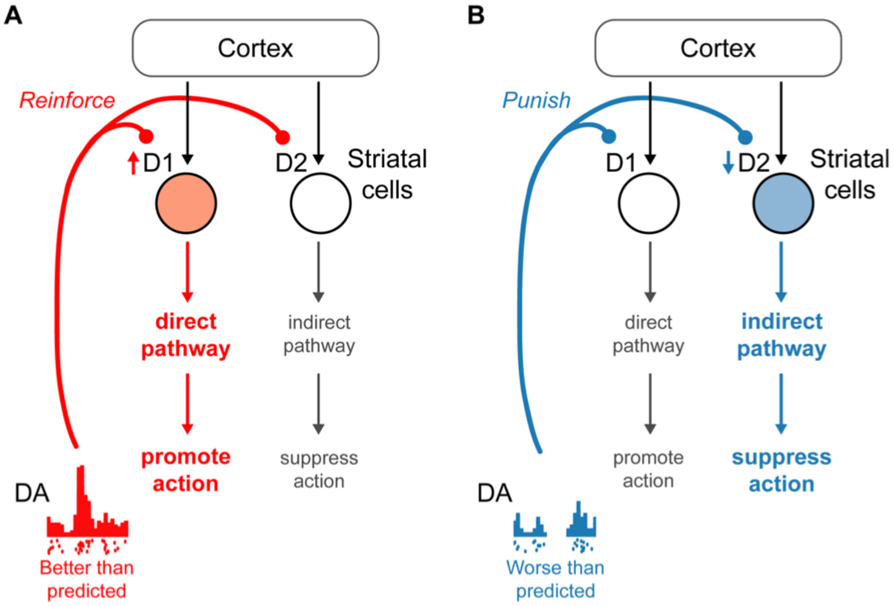


Figure 2.4: Separate pathways for positive and negative feedback in basal ganglia. There are two type of DA receptors in the striatum. (A) Positive reinforcement activates the direct pathway via D1 receptors: Firing DA neurons promotes the immediate selection of better than predicted action by activating the D1 receptors in striatum. (B) Negative or punishing outcome activates the indirect pathway by inhibiting D2 receptors. This results in reinforced cortico-striatal plasticity is altered either for selecting these actions or for avoiding them in future. (from (Bromberg-Martin et al., 2010)).

In the next section we will go through some important findings about how the brain is organized to perform these computations.

2.1.5.3 Genetic Influences on Reinforcement Learning

Individual differences in reinforcement learning due to genotype can be observed both at the neuronal and behavioral level. Experiments on DARP32 gene, which is critical for dopamine dependent striatal synaptic plasticity, knockout mice suggested that cortico-striatal synaptic plasticity is affected by D2 receptor stimulation (Calabresi et al., 2000). Anatomic and functional imaging studies on human also provided evidence for the role of dopamine in shaping fronto-striatal plasticity (Meyer-Lindenberg et al., 2007). Others have studied the association of genetic polymorphisms with avoidance learning due to the altered striatal D2 receptor function. (Frank and Hutchison, 2009) reported a gene-dose effect of C957T polymorphism, which is likely to be in linkage with Taq1A polymorphism, on relative avoidance in a probabilistic learning task. They further found a substantial direct effect of a promoter SNP, rs12364283, on avoidance. Also, participants with reduced expression presynaptic relative to postsynaptic D2 autoreceptors due to SNPs rs2283265/rs1076560 performed worse in avoiding the least rewarding option and better at choosing the most rewarding option. In a similar study, computational modeling revealed that effects of different genes can be identified by reinforcement learning parameters: While increasing expressions of COMT and DRD2 alleles were associatied with higher and lower learning rates of negative feedback, respectively, increasing expressions of DARPP-32 allele was related with lower learning rates for positive feedback (Fig. 2.5) (Frank et al., 2007) suggesting that these genes modulate integration of different feedback.

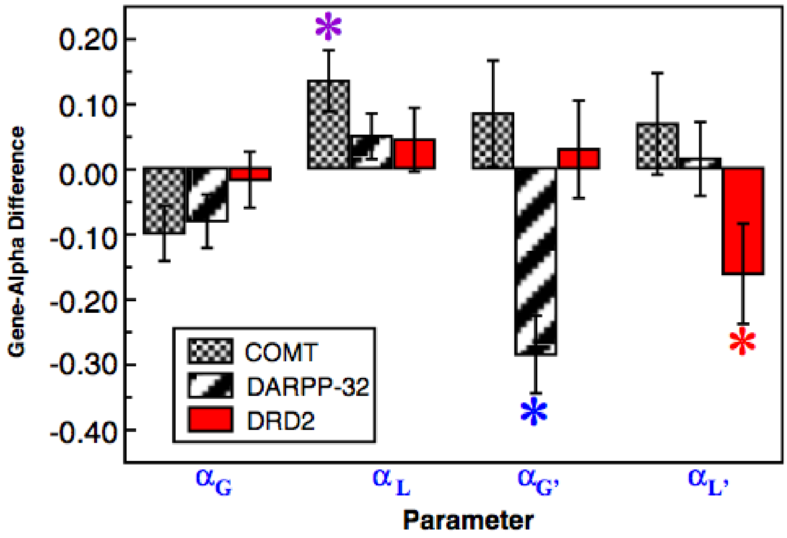


Figure 2.5: Genetic dissociations of three genetic polymorphisms on positive and negative learning rates for rewarding and punishing outcomes, respectively. Increasing COMT allele was associated with higher positive learning rate, increasing DARPP-32 allele was associated with lower positive learning rates, and finally increasing DRD2 allele was associated with lower negative learning rates (from Frank et al. (2007)).

2.2 Bayesian Learning and Decision Making

How does the brain deal with the uncertainty and how does the nervous system integrate cues that are informative but have different reliabilities? Cue combination studies can help understand how a perceptual system behaves and adjusts weights according to the reliability of different cues. In the next section we will look at the examples from the cue combination field mainly considering multisensory integration.

2.2.1 Cue Combination

An organism can use multiple sources of information in its environment to make better decisions, which is in most cases more beneficial than using single sources. Sensory cues such as visual, auditory, tactile, and olfactory are integrated to guide an agent's actions. Different mechanisms have been proposed to combine these sensory cues.

Linear models of cue combination propose a weighted sum of cue reliabilities so that more reliable cues will have greater influence on the action taken or the decision made. It is based on the assumption that cues are Gaussian distributed with the reliability, or precision π_i (inverse variance $\pi_i = 1/\sigma_i^2$) and are conditionally independent. Under the uniform priors the maximum likelihood estimate defines the optimal integration of means (Ernst and Banks, 2002):

$$< \hat{x} > = \sum_i w_i \hat{x}_i \tag{2.9}$$

$$w_i = \frac{\pi_i^2}{\sum_j \pi_j^2} \tag{2.10}$$

where w is the ratio of precisions and i refers to the cue.

To validate the optimal integrator model, cue combination studies tested combined-cue model against the performance under the single-cue model. While combining cues optimally, individual differences in perception influence weighting of cues with subjective reliability (Knill and Saunders, 2003). In case of correlations of the cues, weights should be corrected for the correlations in cue reliabilities (Oruç et al., 2003). Many studies have confirmed this standard cue integration model in human subjects. One of the earliest studies to show that humans integrate multisensory cues optimally was by Ernst and Banks (2002). They predicted that visual and haptic estimates should be combined based on maximum likelihood estimates and the variance of the final estimate should have a smaller variance than visual and haptic variances alone according to the following equation:

$$\sigma_{VH}^2 = \frac{\sigma_V^2 \sigma_H^2}{\sigma_V^2 + \sigma_H^2} \tag{2.11}$$

Standard cue combination studies usually ignore prior knowledge although in reality it has a large influence on decision making (Vilares and Körding, 2011). Others approached the problem from a Bayesian decision theoretic framework. In Bayesian decision theory this corresponds to multiplication of likelihood functions for each cue with the prior distribution of estimates before the sensory observation

$$p(x|s1, s2) \propto p(s1|x)p(s2|x)p(x) \tag{2.12}$$

where the probability distributions replaced the point estimates in weighted linear model. One key difference of Bayesian formulation to linear weighted model is the incorporation of prior knowledge. When the prior and likelihood terms are Gaussians then the estimate will correspond to the mean of the posterior density. Körding et al. (2007) showed that humans can perform causal inference not only in high-level cognition, but also in perception. In an auditory-visual localization task they used priors for visual and auditory cues in a Bayesian structural model and marginalize over these terms in order to calculate optimal estimation of the position, hence showing that by integrating nonlinear terms, i.e. interaction priors, human performance can be modeled more successfully. According to Bayesian nonlinear models, as the conflict between cues increases, subjects might down-weight the cues but not fully ignore the cues (Knill, 2007b). Others showed suboptimal performance in human subjects when dealing with perceptual estimation in uncertain environments (Landy et al., 2007). The visual system can change its model of the statistics of planar figures as shown by a prior model which can change from trial to trial similar to a Kalman Filter (Knill, 2007a). Unlike many cue combination models, models of Kalman filter do not have the assumption that variables of interest do not change over time (Vilares and Körding, 2011). Variables can change over time depending on the environmental factors.

Among the brain regions that have been studied for multisensory, or multimodal cue combination are the superior colliculus (SC) and the dorsal medial superior temporal area (MST). The cue combination rule was investigated in single neuron recordings in MSTd and it was found that a weighted linear combination rule described the responses with the weights depending on cue reliabilities (Morgan et al., 2008). Layers of SC consists of multisensory neurons and integrates visual, auditory and somatosensory information (Meredith and Stein, 1983). Humans can integrate visual and vestibular (inertial motion) cues in a statistically optimal fashion to increase precision of the final estimate of heading angles (Gu et al., 2008). In heading perception, MSTd neurons are tuned to both visual and vestibular information. Electrical microstimulation of MSTd provided a causal link between the neurons in this area and visual heading judgements (Gu et al., 2012).

Although there is not a unified theory of neuronal behavior for optimal cue integration, there have been some important theories. The approach of Ernst and Banks (2002) brought up the question as to whether the nervous system implements an MLE integrator, which is performed by the interactions among populations of visual and haptic neurons. However, the wide amount of evidence showing that humans perform near-optimal Bayesian inference implies that neurons encode and combine probability distributions. This type of representation of probability distributions instead of the value of a stimulus from decision making to motor control tasks is called probabilistic population codes hypothesized by Ma et al. (2006). They suggest that while cortex represents probability distributions, these distributions are transformed to estimates in motor cortex or in subcortical areas during decision-making.

Another proposed framework to explain the computational mechanisms that neurons implement to combine cues is called divisive normalization, where the activity of each neuron is divided by the net activity of all multisensory neurons to produce a final response (Ohshiro et al., 2011). Since this approach takes into account the interaction of the neurons in the population, it is a good candidate to explain the nonlinear effects observed in multisensory units such as SC and MSTd as we discussed above.

So far we have considered the cases of stationary environments, where the cue reliabilites do not change over time. In the next sections, we will consider models that also take into account the dynamics of the environment.

2.2.2 Bayesian Decision Theory

Computational models of learning can be classified as normative and descriptive. Normative models assume that the learner is ideal such that it is fully rational and learns perfectly accurate. On the other hand, descriptive models try to approximate the actual behavior which is not always 'so optimal'. "Observing the observer" is a descriptive framework that includes perceptual inference embedded in a generative model of decision-making (Daunizeau et al., 2010b). This approach provides a representation of sensory inputs and responses of a subject. It is based on the assumption that humans are Bayesian observers who have prior beliefs about the hidden states of the world and update their beliefs with each new piece of information.

According to this framework, a Bayesian observer implements two levels of processing. First level is a perceptual model of the environment $m^{(p)}$ that causes the sensory input. Second is a response model $m^{(r)}$ which is a mapping of sensory input to the observed responses. Under a perceptual model, the subject can form a probabilistic model of the environment:

$$p(u, x|m^{(p)}) = p(u|x, m^{(p)})p(x|m^{(p)}) \tag{2.13}$$

where the first term on the right hand side is the likelihood of the sensory input given the hidden states x under $m^{(p)}$. The last term is the prior beliefs about the hidden states x before any observations are made. In a decision making task, causal structure of states need to be learned and encoded in marginal posterior density according to Bayes' rule:

$$p(x|u, m^{(p)}) = \frac{p(u, x|m^{(p)})}{\int p(u, x|m^{(p)})dx} \tag{2.14}$$

Note that the updates follow a Markovian sequence: Current posterior belief depends only on the current input and past beliefs:

$$p(x|u^{(1,\ldots,k)}, m^{(p)}) \propto p(u^{(k)}|x, m^{(p)})p(x|u^{(1,\ldots,k-1)}, m^{(p)}) \tag{2.15}$$

Variational treatment of perceptual model introduces an approximate posterior over hidden states $q(x|\lambda)$ assuming that subjects track the mean and variance of these variables, which depend on parameters of the perceptual model and the sensory inputs $\lambda \equiv \lambda(u, \vartheta)$. Therefore, tracking the sufficient statistics can be performed by the subject in a Markovian way. A response model $m^{(r)}$ with parameters θ maps these representations to the observed responses y. Likelihood of observed responses can be factorized over the trials:

$$p(y|\theta, \vartheta, u, m^{(r)}) = \prod_k p(y^{(k)}|\theta, \vartheta, u, m^{(r)}) \tag{2.16}$$

A typical response model is in the form of a softmax function (see section 2.2.4). Inverting the response model with an approximation provides a solution to the inverse Bayesian Decision Theory. Details of the variational approximation of the perceptual and the response model are given in Daunizeau et al. (2010b).

In chapter 5, we will deal with an environment involving perceptual uncertainties. A perceptual model called Hierarchical Gaussian Filter (HGF) will be implemented (described in the next section). Therefore, a stochastic mapping of perceptual beliefs to actions (precision weighted response model) will be used in that chapter.

2.2.3 Hierarchical Gaussian Filter (HGF)

HGF (Mathys et al., 2011) is an extension of the IBDT models proposed in (Daunizeau et al., 2010a). It is a generative model that describes the relationship between an agent's beliefs and its environment. When combined with a response model, HGF is very convenient to study different aspects of human learning and decision making. It is proposed as an alternative to ideal Bayesian learners, because these Bayesian models comprise high dimensional integrals, giving rise to less plausible neuronal implementation.

Figure 2.6 represents the graphical model of HGF. Three levels are the state variables, x_1, x_2, x_3, of the environment that generates the input u. In theory, many other levels can be added to the hierarchy. For a binary setting, the first level state variable $x1 \in \{0, 1\}$, will generate input, u. The likelihood model for this causal structure is given by

$$p(u|x_1) = (u)^{x_1}(1-u)^{1-x_1} \tag{2.17}$$

The second level state variable x_2 is a parameter of the probability that $x_1 = 1$. This conditional probability is given with the following:

$$p(x_1|x_2) = s(x_2)^{x_1}(1-s(x_2))^{1-x_1} = Bernoulli(x_1; s(x_2)) \tag{2.18}$$

where s is the sigmoid function:

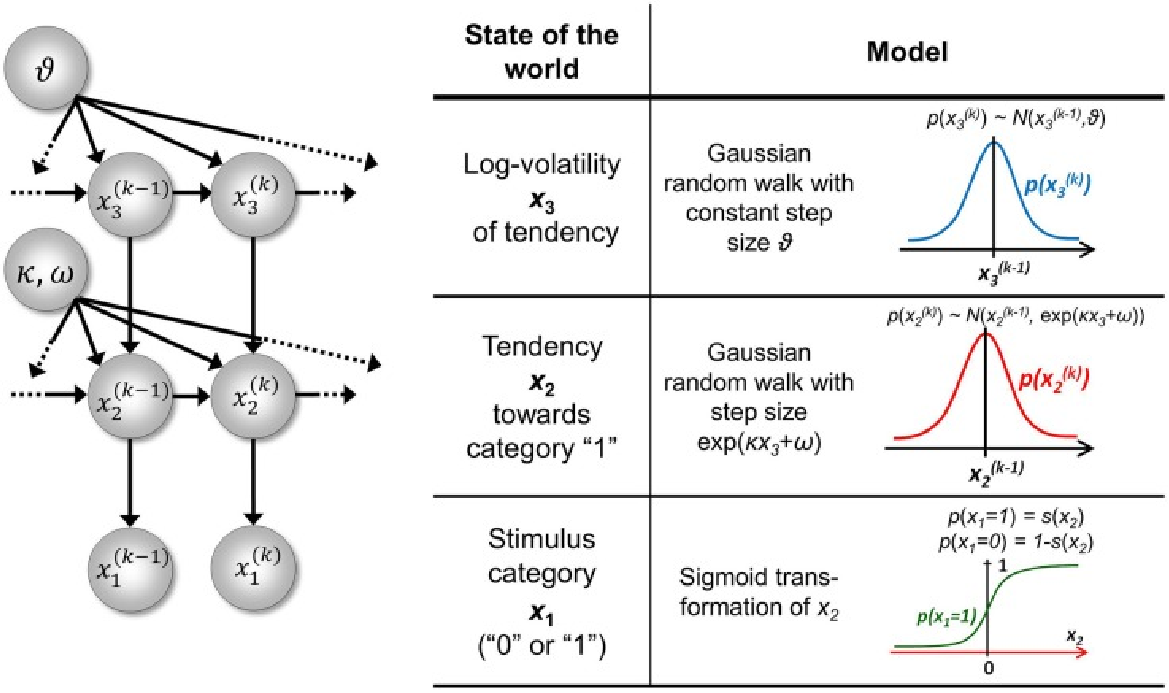
$$s(x) = \frac{1}{1+exp(-x))} \tag{2.19}$$

The second level is assumed to be a Gaussian random walk. Its mean is conditional on its previous value $x_2^{(k-1)}$. The third level is the log-volatility of the environment such that it forms the variance of the second level:

$$p(x_2^{(k)}|x_2^{(k-1)}, x_3^{(k)}, \kappa, \omega) = \mathcal{N}(x_2^{(k)}; x_2^{(k-1)}, \exp(\kappa x_3^{(k)} + \omega)) \tag{2.20}$$

$$p(x_3^k|x_3^{(k-1)}, \vartheta) = \mathcal{N}(x_3^{(k)}; x_3^{(k-1)}, \vartheta) \tag{2.21}$$

κ scales the influence of third level onto the second level, therefore called as coupling parameter. ω is the variance parameter for the second level independent from other levels. Finally, ϑ is the prior for the variance of third level which also follows a Gaussian random walk. The



State of the world	Model
Log-volatility x_3 of tendency	Gaussian random walk with constant step size ϑ
Tendency x_2 towards category "1"	Gaussian random walk with step size exp(κx_3+ω)
Stimulus category x_1 ("0" or "1")	Sigmoid trans-formation of x_2

Figure 2.6: Overview of the three level HGF. For any trial k, log-volatility $x_3^{(k)}$ follows a Gaussian random with constant step size ϑ). The second level $x_2^{(k)}$ is the tendency of the outcome. Its distribution is also Gaussian with step size $e^{(\kappa x_3^{(k)}+\omega)}$). The first level state variable $x_1^{(k)}$ is the binary stimulus category and the sigmoid transform of the second level so that $p(x_1 = 1) = s(x_2)$ (from (Mathys et al., 2011)).

joint probability for this hierarchical model is defined in the following:

$$
\begin{aligned}
&p(u^{(k)}, x_1^{(k)}, x_2^{(k)}, x_3^{(k)}, x_2^{(k-1)}, x_3^{(k-1)}, \kappa, \omega, \vartheta) \\
&\qquad = p(u^{(k)}|x_1^{(k)})p(x_1^{(k)}|x_2^{(k)})p(x_2^{(k)}|x_2^{(k-1)}, x_3^{(k)}, \kappa, \omega) \\
&\qquad p(x_3^{(k)}|x_3^{(k-1)}, \vartheta)p(x_2^{(k-1)}, x_3^{(k-1)}p(\kappa, \omega, \vartheta)
\end{aligned}
\tag{2.22}
$$

Learning can be defined as updating current beliefs with each new piece of information. The main idea of the model inversion is to calculate the posterior density at each time point after observing the new input. The joint probability of the input and the states at time point k $p(u^{(k)}, x_1^{(k)}, x_2^{(k)}, x_3^{(k)}, \kappa, \omega, \vartheta)$ can be calculated given all the previous inputs $u^{(1,\ldots,k-1)}$ by integrating over $x_2^{(k-1)}$ and $x_3^{(k-1)}$. Once the new input $u^{(k)}$ is observed, posterior probabilities can be calculated for each variable by marginalizing over the rest of the variables. This approach has been applied by (Behrens et al., 2007) and (Behrens et al., 2008) to implement an ideal Bayesian observer. In HGF, a variational Bayesian (VB) inversion to the generative model including mean field approximation returns the joint posterior distribution as a product of approximate marginal posterior distributions. For the binary state x_1, the approximate posterior is a Bernoulli distribution:

$$\begin{aligned} p(x_1^{(k)}|\kappa,\omega,\vartheta,u^{(1,\ldots,k)}) &\approx q(x_1^{(k)}) \\ &= Bern(x_1^{(k)};\mu_1^{(k)}) = (\mu_1^{(k)})^{x_1^{(k)}}(1-\mu_1^{(k)})^{1-x_1^{(k)}} \end{aligned} \tag{2.23}$$

The approximate posteriors for the second and third level are Gaussians so that they can be encoded in their first two moments, i.e., mean and the variance.

$$p(x_2^{(k)}|\kappa,\omega,\vartheta,u^{(1,\ldots,k)}) \approx q(x_2^{(k)}) = \mathcal{N}(x_2^{(k)};\mu_2^{(k)},\sigma_2^{(k)}) \tag{2.24}$$

$$p(x_3^{(k)}|\kappa,\omega,\vartheta,u^{(1,\ldots,k)}) \approx q(x_3^{(k)}) = \mathcal{N}(x_3^{(k)};\mu_3^{(k)},\sigma_3^{(k)}) \tag{2.25}$$

For brevity, we will not go through the variational inversion. For details of the variational inversion and the quadratic approximation to the variational energies please refer to (Mathys et al., 2011). Importantly, for the second and third levels, belief updates at time point k are

$$\mu_2^{(k)} = \mu_2^{(k-1)} + \sigma_2^{(k)}\delta_1^{(k)} \tag{2.26}$$

$$\mu_3^{(k)} = \mu_3^{(k-1)} + \sigma_3^{(k)}\frac{\kappa}{2}\frac{e^{\kappa\mu_3^{(k-1)}+\omega}}{\sigma_2^{(k-1)} + e^{\kappa\mu_3^{(k-1)}+\omega}}\delta_2^{(k)} \tag{2.27}$$

where $\delta_i^{(k)}$ the prediction errors at level $i \in (1,2)$, and $\sigma_j^{(k)}$ is the variance at level $j \in (1,2,3)$:

$$\delta_1^{(k)} = \mu_1^{(k)} - \hat{\mu}_1^{(k)} \tag{2.28}$$

$$\hat{\mu}_1^{(k)} = s(\mu_2^{(k-1)}) \tag{2.29}$$

$$\delta_2^{(k)} = \frac{\sigma_2^{(k)} + (\mu_2^{(k)} - \mu_2^{(k-1)})^2}{\sigma_2^{(k-1)} + e^{\kappa\mu_3^{(k-1)}+\omega}} - 1 \tag{2.30}$$

$$\sigma_2^{(k)} = \frac{1}{\frac{1}{\hat{\sigma}_2^{(k)}} + \hat{\sigma}_1^{(k)}} \tag{2.31}$$

$$\hat{\sigma}_2^{(k)} = \sigma_2^{(k-1)} + e^{\kappa\mu_3^{(k-1)}+\omega} \tag{2.32}$$

$$\hat{\sigma}_2^{(k)} = \hat{\mu}_1^{(k-1)}(1 - \hat{\mu}_1^{(k-1)}) \tag{2.33}$$

They are similar to the update rules given in reinforcement learning where the prediction errors, δ are weighted by learning rates to update current beliefs. In the second level, variance acts similar to a learning rate Eq. (2.26).

2.2.4 Bayesian Brain

A common assumption in sensory and perceptual processing in the brain is that the information is processed hierarchically. Lower-levels represent more details about the stimulus properties, where the higher levels represent an integrated information (Rauss and Pourtois, 2013). These levels communicate with each other via bottom-up or top-down processes. (Melloni et al., 2012) have proposed that in the visual system, bottom-up processes takes place in V1, top-down control is encoded in V2, and V4 encodes the interaction of these two processes.

According to predictive coding hypothesis of the brain, the units of the hierarchical structure pass messages recurrently through forward and backward connections. The high level prediction units pass down predictions. PE units receive these predictions and compare with the actual input to generate a new PE that is passed up to update predictions (or beliefs) in the higher units. As a result, predictive coding theory suggests that while predictions shape the online estimation of the state of the world, PEs affect plasticity and learning (Van de Cruys et al., 2014).

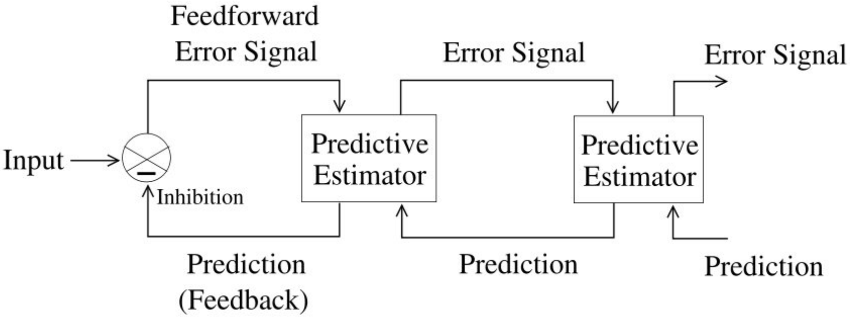


Figure 2.7: Predictive coding in the visual cortex. A general architecture describing predictive coding models consisting of input layer that signals the error to the higher level prediction units through feed-forward connections, whereas the prediction units signal the predictions of estimated neural activity (from Rao and Ballard (1999)).

The free energy principle formulates possible predictive coding mechanisms in the brain (Friston, 2009). It suggests that hierarchical message passing takes place through encoding sufficient statistics of the causes of sensory inputs (i.e. hidden states). According to this, actions and sufficient statistics of the states are optimized by minimizing free energy. The neural implementation of the free energy principle implies one important quantity, precision or inverse variance, which adjusts the influence of PEs on the belief updates in state (or prediction) units. This precision of PEs is thought to be encoded as a gain of the neuronal units representing PEs and influenced by new rewards. From a behavioral point of view, this is plausible because an optimal agent should decrease its belief update in an uncertain environment, which can be succeeded through adjusting the precision of PEs. Therefore, this adjusted PE is called precision weighted PE and is similar to the update in reinforcement learning theory.

A neuronal network model of the auditory cortex based on a predictive coding approach i.e. including a predictive layer and a PE layer, is tested with MEG explained many properties of mismatch negativity (Wacongne et al., 2012). A comparison of fMRI activity of the visual cortex provided more support for visual cognition as a predictive matching process rather than a feature detection system (Egner et al., 2010). A three-level hierarchical model of visual cortex has explained some properties such as extra-classical receptor field effects (Rao and Ballard, 1999). The model assumed that neurons with these properties signal the difference between the input and the prediction to the higher levels (Fig. 2.7).

Bayesian models of learning and decision making are promising methods to study hierarchical processing and to test Bayesian brain hypotheses. Learning the basic and more detailed features about the stimulus at different levels can be modeled with hierarchical Bayesian models. HGF has been used to model learning the probabilistic structure of audio-visual contingencies to derive PEs at multiple levels, which were then used in fMRI analysis. They found that while low-level PEs are encoded in visual areas and midbrain, high-level PEs were found in the basal forebrain which suggests that hierarchical coding of PEs might explain that different neuromodulatory mechanisms play role in learning different properties of stimulus, i.e. stimulus outcome vs. outcome probabilities (Iglesias et al., 2013).

2.3 Functional Magnetic Resonance Imaging

2.3.1 BOLD signal

Functional Magentic Resonance Imaging (fMRI) is a non-invasive imaging method to study brain function such as motion, vision, speech, emotion, etc. fMRI provides a high spatial resolution (order of a few millimeters) and a good time resolution of a few seconds. Furthermore, it is a useful method for presurgical planning, and diagnosis of disorders of brain function.

Radio frequency (RF) pulse excites the hydrogen atoms to a higher level energy state and aligns it with the magnetic field. Once the RF pulse is removed the atoms return back to lower level energy state. This process is called relaxation and can be described in two dimensions: Relaxation on the direction of B0 (or static magnetic field) is called longitudinal relaxation. Relaxation on the -xy plane that is perpendicular to BO is transverse relaxation. Both processes are exponential decays with time constants T1 and T2 , respectively. However, T2 is influenced by the field inhomogeneities caused by tissue properties, hence called T2*. It depends on the neural activity in an indirect manner (see below) and forms the basis of fMRI BOLD signal.

Neural activity in the brain changes the blood supply in its surrounding. When oxyhemoglobin looses its oxygen, it also looses its diamagnetic properties and becomes deoxyhemoglobin, a paramagnetic molecule. Measurement of blood oxygenation level dependent (BOLD) contrast as a signal by MR scanners is based on this principle. Increase in the neural activity will decrease the deoxyHb amount. This is followed by an increase in the BOLD response as a result of increased inhomogeneity in the magnetic field around the active brain tissue. Early studies of BOLD signal (Ogawa et al., 1990) showed that changes in the oxyhemoglobin / deoxyhemoglobin ratio induced by physiological events in the brain are detected by BOLD contrast in gradient-echo proton images. Since the observations of rapid changes in blood oxygen levels with gradient-echo echo-planar imaging (EPI) (Turner et al., 1991), it became an important method for time course studies of brain imaging.

Analyses of multi unit activity and local field potentials (LFP) recorded simultaneously with fMRI revealed that BOLD signal reflects LFPs and therefore corresponds to synaptic activity of a neural population rather than its spiking output (Logothetis et al., 2001). To link the causal relationship between neural activity and BOLD response we can consider the following theoretical equation (Logothetis and Wandell, 2004)

$$B(x) = \int_{n(x)} [A(u) + N_N(u)]H(u)P(x-u)du + N_M(x) \tag{2.34}$$

where B(x) is the measured BOLD response. The summation of the ongoing neural activity A(x), and the random neural noise, Nn(x) gives the total neural activity which is then multiplied with hemodynamic response efficacy $H(u)$ and point spread function P, since the efficacy of the coupling differs along the cortex, hence a function of location of activity x. Finally, the second term N_M is the instrumental noise.

Following the stimulus, properties of the blood flow changes, which is also known as haemodynamic response. Many software packages of fMRI data analysis provide a model of haemodynamic response function (HRF) . In this thesis we will be using canonical HRF whose shape depends on parameters such as delay of response and undershoot, dispersion of response and undershoot, etc. These physiological variables are parameterized with double gamma function as implemented in SPM package. A first order Taylor approximation of canonical response allows for modeling its dispersion and temporal derivatives as well. A more detailed explanation of hemodynamic model is given under DCM section.

Block designs can consist of blocks of identical trials or two or more alternating trials. Event related designs are used to model transient responses evoked by discrete stimuli. In this framework one needs to average over many trials to provide a good signal-to-noise ratio (SNR). This results in longer acquisition times compared to block designs where SNR is better due to sustained neural activation during a block. Although both approaches have their own advantages, experiments that include continuous events fit only to block designs. If the experimenter wants to investigate multiple events within trials then event-related designs are appropriate.

2.3.2 fMRI Data Analysis

First step of preprocessing of functional scans is slice timing or slice time correction. It corrects for the time delay between the slices of one volume acquired during one repetition time. The correction routine takes into account the order of slice acquisition, and to perform time shifting, it applies convolution to introduce phase shift in the frequency domain and transforms back to time domain. Temporal processing is followed by spatial processing which usually starts with a motion correction to estimate movement parameters. Subject movements in the scanner will induce distortions in the images and can cause mistakes in the final activation maps, if they are not included in the timeseries modeling. Rigid body realignment with six degrees of freedom (x, y, z, pitch, roll, yaw) estimates movement parameters which are then used to transform all the scans to the defined reference scan. Next, anatomical images are registered to functional images. This step is known as co-registration where the source image is moved to match a reference image. Then spatial normalization applied to warp images onto standard anatomical space with a template image such as Montreal Neurological Institute (MNI) template. Finally, smoothing is applied to each voxel where the hemodynamic response is convolved with a Gaussian smoothing kernel. It is recommended to select the full-width at half maximum (FWHM) of the kernel based on the voxel size and the expected effect size (Penny et al., 2011). This procedure is necessary for several reasons. Most importantly, during inter-subject averaging one needs to minimize differences in the anatomy, and a normal distribution of errors is required for the validity of parametric tests (Penny et al., 2011).

General linear model (GLM) is a statistical model commonly used in PET and fMRI data analysis. Time series of each voxel is fitted with the same GLM, hence called a mass-univariate approach. The GLM equation is

$$Y_i = x_{i1}\beta_1 + ... + x_{ij}\beta_j + ... + x_{iJ}\beta_J + \epsilon_i \tag{2.35}$$

where y_i is the i-th observation, x_{ij} is the j-th explanatory variable, and β is the parameter that need to be estimated (Friston et al., 1995). Errors are assumed to be independent and identically distributed (iid) with normal distribution $N(0, \sigma_2)$. In matrix notation it can be written as

$$Y = XB + \epsilon \tag{2.36}$$

Y and B are now the column vectors consisting of $y_1, ..., y_J$ and $\beta_1, ..., \beta_J$, respectively. X is the design matrix of size IxJ. For each trial type of interest there will be a separate regressor in the design matrix X. The trials of interest are coded as binary values at its onsets and then convolved with HRF. Typical method to estimate β coefficients is least squares estimation. When X is of full rank, i.e. when all columns are linearly independent, parameter estimation is given by ordinary least square (OLS)

$$\hat{B} = (X^T X)^{-1} X^T Y \tag{2.37}$$

$\hat{B}$ is the optimum set of parameters that minimizes sum of squares $\epsilon^T \epsilon$. In cases of linear dependency, $X^T X$ will be singular and have no inverse. Therefore pseudoinverse can be calculated, which will provide least square estimates with the minimum sum-of-squares.

After the parameter estimation, contrast weights which are linear combination of parameters, are constructed for hypothesis testing. Statistical inference on the subject level depends on the question. For example, a t-test can be used to test an activation in a voxel against the null hypothesis that there is no activation. Or a two-sample t-test can be used for testing differences in activations by comparing the means of two parameters with appropriate contrast vector. Resulting statistical images are called contrast images and can be carried to a second level analysis with another GLM. At this stage it is possible to test significant group activations or to compare regional activations between two groups for a certain condition. Another way to do statistical inference is using F-contrasts. Finally, due to the large number of voxels in the brain scans, multiple comparison problem arises. Height thresholding can overcome this statistical issue. Correction methods include Bonferroni, false discovery rate (FDR), or family wise error correction (FWE) in case of a field of voxels.

2.3.3 Brain Connectivity

The interactions between brain regions can be studied with connectivity models. They are network models where the nodes are brain regions and the links that connect those nodes are anatomical, or functional connections (Friston, 1994; Rubinov and Sporns, 2010). Three major concepts of connectivity have been recently distinguished: (i) structural connectivity, subsuming the anatomical (and physiological) basis for information transfer between individual network entities, (ii) functional connectivity, describing a correlative relationship between brain activities or physiological variables depending on brain activity (eg., blood flow, blood oxygenation, etc.), and (iii) effective connectivity, denoting a causal relationship between brain activities and directly expresses the issue of information transfer.

Anatomical connections are formed by synapses or fiber pathways in white matter. Anatomical connections of an individual can change through plasticity and ageing. A promising method for identifying these connections is tractography based on Diffusion weighted imaging (DWI) where the signal of anisotropic diffusion along axonal fibers is used to track anatomic connections. One can estimate parameters such as traces, Fractional Anisotropy (FA), Apparent Diffusion Coefficient (ADC) as correlates of structural connectivity. Despite all its advantages diffusion images lack of providing the direction of connections. Invasive tracing studies can identify which brain region projects to another.

Functional integration of the cortex can be determined with connectivity methods. They are based on statistical relationships between regional activations. These methods include:

- Decomposition methods: Principle Component Analysis and Independent Component Analysis
- Static and dynamic models of effective connectivity: e.g. Psychophysiological Interaction, Structure Equation Modeling, Dynamic Causal Modeling, Granger's Causality.

Decomposition methods such as principle component analysis (PCA) and independent component analysis (ICA) have been adapted for studies of brain connectivity. PCA decomposition detect temporally correlated signal, and it requires orthogonality between the timeseries. ICA, on the other hand, decomposes data into independent components, and is based on the assumption that its components are statistically independent and have non-Gaussian distributions. Although each method constitutes limitations, ICA approach has been very useful in analyzing spatial, frequency, and connectivity characteristics of statistically independent components of resting state functional data (Kiviniemi et al., 2003).

While functional connectivity refers to temporal correlations or covariance between the regional activities, it does not take into account the direction of the influence. The models that describe the directional influence of one brain region onto another are called effective connectivity models. For instance, psychophysiological interaction (PPI) is a very basic one that is based on the regression models. It considers the interaction between the neuronal activity in one region, x_k, with an experimental factor, g_p, and its effect on the neuronal activity in another region, given with the following statistical model (Friston et al., 1997)

$$x_i = x_k \times g_p.\beta_i + [x_k g_p G].\beta_G + \epsilon_i \tag{2.38}$$

where x_k and g_p are column vectors, $x_k \times g_p$ is element-wise product, and G is a matrix whose columns are effects of no interest. An example to this interaction is the contribution of V1 activity onto V5 activity under the attention to visual motion (Friston et al., 1997). Note that a PPI analysis is limited to a single source and a single target area at one time, and it does not take self connections into account.

Structural equation modeling (SEM) is another effective connectivity method. It makes assumptions about causal connections between regions based on neuroanatomical knowledge. In this method, after connectivity matrix is specified, path coefficients are optimized by matching the estimated covariance with observed covariance (McIntosh, 1994). Note that both SEM and PPI are static models of effective connectivity. In the next section we will look at a dynamical model of connectivity, DCM. Although there are other methods such as Granger causality (Granger, 1969), the main focus will be on DCM as it is applied in this thesis.

Finally, some studies have shown associations between functional and anatomical connectivity. For example, functional connectivity can be predicted by anatomical connectivity with a predictive power that is independent from the computational model implemented (Messé et al., 2015). Others have found evidence for structure-function relation by demonstrating the anatomical connectivity underlying spontaneous cortical dynamics (Honey et al., 2007). Finally, anatomic connectivity parameters derived from diffusion tensor images have been used as a prior to inform DCMs (Stephan et al., 2009b). This approach has proved that integrating anatomic information improves the estimation of effective connectivity.

2.3.4 Dynamic Causal Modeling

DCM is a an effective connectivity method that estimates the strength of neuronal coupling among brain regions that influence the activity directly or indirectly. It combines a neuronal model with a modality specific forward model such as haemodynamic model for fMRI, or electromagnetic models for EEG. The inversion of neuronal and a forward model allows for making inferences in the neuronal activity level which is superior to many other connectivity models, e.g. SEM or PPI. The estimation and inference methods are fully Bayesian and will be explained in this section. It was first introduced as an effective connectivity model and dynamic input-state-output with multiple inputs and outputs (Friston et al., 2003) based on a previous study that reports the same approach for a single region, i.e. Bayesian identification of hemodynamic models (Friston, 2002).

These couplings are established specific to experiment. Therefore, it assumes a known deterministic input which is the experimental stimulus in the form of either boxcar or stick functions. Each stimulus can enter the system in different ways. One way is to evoke responses in a certain region directly, for example visual stimulus entering the V1 can be of this kind. Another way that a stimulus can perturb the system is to modulate the connectivity strength between two regions. In its basic form, DCM considers the brain as a nonlinear deterministic system. The state variables consist of four states for hemodynamic model and one state for a neuronal model. Note that this formulation is modality dependent. In the case of event related potentials, one need to specify 8 state variables per region, and a linear model instead of hemodynamic model. Here the equations of DCMs for BOLD signal will be described. The neuronal state equation for N interacting brain regions with the neuronal states $z = [z_1, ..., z_N]$, the change in the neuronal activity, $\dot{z}$, is in the given form

$$\dot{z} = f(z, u, \theta^n) \tag{2.39}$$

u is the experimental input that perturbs the system and θ^n are the time-invariant parameters of the neuronal system. In the original DCM, bilinear state equations are formed by using a Taylor series expansion around the system's resting state $(z = 0, u = 0)$:

$$\begin{aligned} \dot{z} &\approx f(z,u)|_{z=0,u=0} + \frac{\partial f}{\partial z} z + \frac{\partial f}{\partial u} u + \frac{\partial^2 f}{\partial z \partial u} zu & (2.40) \\ &= [\, A + \sum_i u_i B^i] \, z + Cu & (2.41) \end{aligned}$$

where

$$\begin{aligned} A &= \frac{\partial f}{\partial z} z & (2.42) \\ B^i &= \frac{\partial^2 f}{\partial z \partial u} zu & (2.43) \\ C &= \frac{\partial f}{\partial u} u. & (2.44) \end{aligned}$$

The matrix A represents the intrinsic connectivity in the absence of an input. It represents connectivity at the baseline level. Each matrix B^i describes the effect of input u_i on the connectivity between the regions. Finally, C matrix represents the direct influence of external inputs on the regional activity. Matrix elements a_{ij} and b_{ij} correspond to backward and forward connections for $i \neq j$, and to self-connections for $i = j$. These partial derivatives are the neuronal parameters, $\theta^n = \{A, B^i, C\}$, that need to be estimated. The coupling parameters are rate constants of the neural populations, hence, they are in units of Hz.

Figure 2.8 demonstrates an example of a system of connected regions. The following equations describe the neuronal state equations for this architecture as a system of differential equations:

$$\begin{aligned} \dot{z}_1 &= a_{11} z_1 + u_1 c_1 + a_{13} z_3 + u_2 a_{13} b_{13} & (2.45) \\ \dot{z}_2 &= a_{22} z_2 + a_{21} z_1 + a_{23} z_3 & (2.46) \\ \dot{z}_3 &= a_{33} z_3 + a_{32} z_2 & (2.47) \end{aligned}$$

We can write these expressions in the matrix form:

$$\begin{bmatrix} \dot{z}_1 \\ \dot{z}_2 \\ \dot{z}_3 \end{bmatrix} = \left\{ \begin{bmatrix} a_{11} & 0 & a_{13} \\ a_{21} & a_{22} & a_{23} \\ 0 & a_{32} & a_{33} \end{bmatrix} + u_2 \begin{bmatrix} 0 & 0 & b_{13} \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \right\} \begin{bmatrix} z_1 \\ z_2 \\ z_3 \end{bmatrix} + \begin{bmatrix} c_1 \\ 0 \\ 0 \end{bmatrix} u_1 \tag{2.48}$$

Nonlinear DCMs (Stephan et al., 2008) were introduced in order to model fast changes in effective connectivity such as short-term synaptic plasticity (STP) that are driven by the history of synaptic inputs. One such mechanism is neuronal gain control which describes the changes in the gain of one neuronal unit as a multiplicative interaction of synaptic inputs from two other neuronal units. Those kind of nonlinear effects, also known as gating, can be

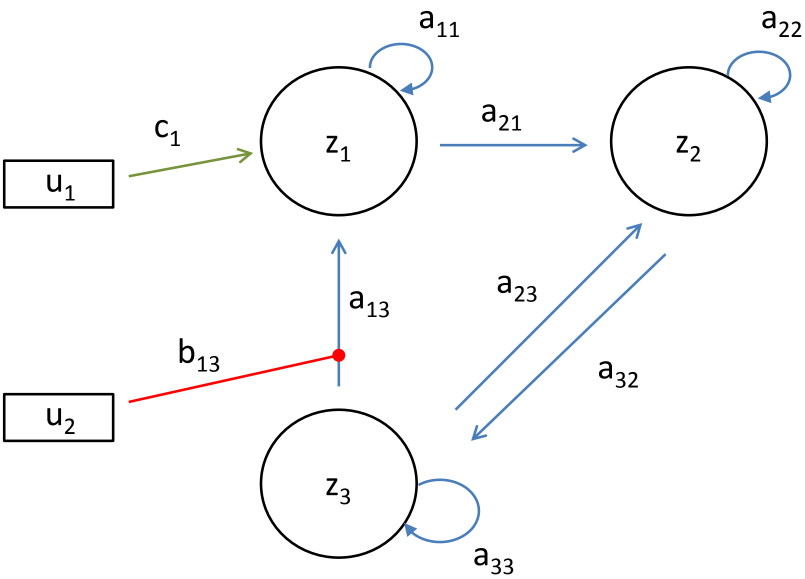


Figure 2.8: Interaction of three brain regions manipulated by experimental input. The relationship between regional activations, z_1, z_2 and z_3 and their connectivity strengths a_{ij} as well as the experimental inputs u_1, u_2 is given with a set of differential equations.

modeled by including an additional term from Taylor expansion of neuronal activity:

$$\begin{aligned}\dot{z} &\approx f(z,u)|_{z=0,u=0} + \frac{\partial f}{\partial z}z + \frac{\partial f}{\partial u}u + \frac{\partial^2 f}{\partial z \partial u}zu + \frac{\partial^2 f}{\partial z^2}\frac{z^2}{2} && (2.49)\\ &= [\,A + \sum_i u_i B^i + \sum_j z_j D^j]\, z + Cu && (2.50)\end{aligned}$$

where

$$D^j = \frac{1}{2}\frac{\partial^2 f}{\partial z_j^2}|_{u=0}. \tag{2.51}$$

DCM uses a hemodynamic model that is a nonlinear process known as as extended Balloon model (Friston, 2002) and converts neuronal activity to the observed BOLD signal. Parameters of the hemodynamic model are combined with neuronal state equations, $\theta = \{\theta^n, \theta^h\}$, to obtain a full forward model

$$\begin{aligned}\dot{z} &= f(z,u,\theta) && (2.52)\\ y &= \lambda(z) && (2.53)\end{aligned}$$

For any given values of θ, predicted BOLD signal $h(u,\theta)$ can be calculated and compared

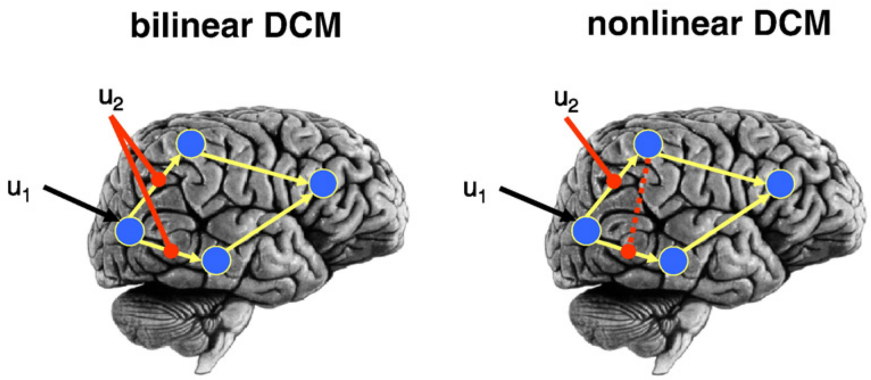


Figure 2.9: Bilinear and nonlinear DCMs. It is possible to model connectivity strengths that can be modulated by another region's activity (red dashed line) with quadratic terms in nonlinear DCMs (from (Stephan et al., 2008)).

to the value of y with the observation model:

$$y - h(u, \theta) = X\beta + e \tag{2.54}$$

where X is confounding effects with coefficients β, and e is the measurement error.

Given their priors, parameters are estimated using a fully Bayesian approach and expectation maximization (EM) algorithm. The goal of the EM algorithm is to find the maximum likelihood solution of a model by maximizing its expected value under the posterior distribution of the hidden variables. The estimation procedure returns the expected values $\eta_{\theta|y}$ and covariance $C_{\theta|y}$ of parameters under Gaussian assumptions. Finally, Bayesian inference is used for hypothesis testing: An effect, $c^T\eta_{\theta|y}$ is tested to be above a certain threshold γ with cumulative normal distribution:

$$p(c^T\eta_{\theta|y} > \gamma) = \Phi_N(\frac{c^T\eta_{\theta|y} - \gamma}{c^T C_{\theta|y} c}) \tag{2.55}$$

While this equation is used for inference about parameter space, we will describe in the next section the inference about the model space.

2.3.5 Inference on Model Space

Bayesian model selection (BMS) is a method for comparison of alternative models using the model evidence, $p(y|m)$ that is how likely to obtain the observed data y, given model m. It is obtained with the following integral over the model parameters θ:

$$p(y|m) = \int p(y|\theta, m)p(\theta|m)d\theta \tag{2.56}$$

where $p(\theta|m)$ is factorized over model priors, and $p(y|\theta, m)$ is the likelihood of observin the data given the model and its parameters. This likelihood term is related with the posterior distribution of model parameters $p(\theta|y,m)$ through Bayes rule:

$$p(\theta|y,m) = \frac{p(y|\theta,m)p(\theta|m)}{p(y|m)} \tag{2.57}$$

Since the integral in equation 2.56 is not analytically tractable, model evidence, $p(y|m)$ is optimized by using variational free energy under Laplace approximation (Friston et al., 2007), which introduces a free energy lower bound on the log model evidence (Beal and Ghahramani, 2003):

$$\log p(y|m) = F + KL(q(\theta)||p(\theta|y,m)) \tag{2.58}$$

Maximizing the free energy, F, will minimize Kullback-Leibler divergence between the approximate posterior and true posterior, $q(\theta) \approx p(\theta|y,m)$. It has been shown that the variational free energy scheme outperforms other model comparison measures such Bayesian Information Criterion (BIC) and Akaike Information Criterion (AIC) for comparing DCMs (Penny et al., 2010). This is expected since the penalty term is not just a function of free parameters, but rather depends on how much the approximated posterior for each parameter is deviated from its true posterior.

Finally, approximated model evidence can be used for model comparison. Depending on the model space assumption e.g. if the optimal model should vary across the subjects, one can select either fixed effect (FFX) or random effect analysis (RFX) . A simple method for fixed effect analysis is calculating Bayes Factor (BF). For two models m_i and m_j, BF is

$$BF_{ij} = \frac{p(y|m_i)}{p(y|m_j)} \tag{2.59}$$

Random effects analysis can be performed with variational Bayesian model comparison which treats each model as a random variable and it is more robust in case of outliers (Stephan et al., 2009a). Another approach for model comparison, especially in case of big number of models, model space can be partitioned. An example of this approach will be implemented in Chapter 4.

3 Functional MR Imaging of Reinforcement Learning

3.1 Introduction

Reinforcement learning (RL) is a powerful method to study learning systems in humans because it is not only applicable in many real-world scenarios, but also biologically plausible. It is different from unsupervised learning as the data is labelled, i.e. reinforcement is either a reward or a punishment. However, RL also differs from supervised learning because the error signal does not provide information about which action should be taken. Instead learning takes place based on the relative rewards of available actions. Also, RL can deal with large number of possible state-action pairs (Tesauro, 1995). These properties make RL suitable for studying human learning in real-world situations. Therefore, this chapter will evaluate its utility in functional neuroimaging and behavioral studies.

This study will consider RL models for a probabilistic selection task where the learning can be formulated as a Markov decision process. This task is similar to a two-armed bandit where the agent learns the return of each arm. In addition to a standard Q-learning model, another model candidate, which takes account of differential learning from positive and negative reinforcement, will be evaluated. Simulations will be performed to understand the behavior of different agents for the given reward schedule. The simulations will examine

- learning behavior of different agents with high and low learning rates,
- trading-off exploration and exploitation,
- whether the choice of initial action values change the course of learning,
- the effect of having separate learning rates for gains and losses in comparison to a single learning rate.

Next, models will be evaluated on a real data set to understand which model explains human behavior better. The selection of these models are based on the biological evidence from earlier studies (Jocham et al., 2011; Frank and Fossella, 2011). Although it is suggested that distinct neuronal structures are involved in model-based and model-free RL, this study will only consider model-free algorithms and their implementation. For a model-based RL example, please refer to Doll et al. (2016).

The RL informed general linear model is a powerful method for fMRI analysis. It enables us to understand many neuronal and behavioral processes quantitatively. This approach will be implemented on an empirical dataset in order to assess its capability in the following:

- identifying which regions are involved in processing prediction errors and different types of reinforcement,
- reflecting differences in the dopamine dependent brain and behavioral responses,

- making associations between neural and behavioral processes.

Performance in learning from negative reinforcement, also known as avoidance learning, has been associated with impairments in dopamine signalling in the brain. Underlying genetic causes are important elements for understanding these alterations and developing new treatments. RL provides a framework to study neurocomputational mechanisms that are modulated by genetics. It can capture the differences in the behavior in terms of learning rate and temperature parameter, or performance in learning from different types of reinforcement. Furthermore, the alterations can be observed in the neural correlates of these processes by combining RL with neuroimaging data (such as prediction error processing).

This study will evaluate the power of RL framework on an underexplored, but the strongest known, genetic factor of obesity. Variations in the fat mass and obesity-associated (FTO) gene predisposes humans to nonmonogenic obesity (Dina et al., 2007; Frayling et al., 2007) (monogenic diseases are result of a single mutated gene such as Huntington's disease, while non-monogenic disease are result of multiple genes in combination with environmental factors). It is linked to a broad spectrum of altered behaviors including: food choice, attention deficiency, impulse control, and substance abuse (Hess and Brüning, 2014; Sobczyk-Kopciol et al., 2011; Choudhry et al., 2013; Karra et al., 2013; Chuang et al., 2015). Moreover, recent analysis of *FTO*-deficient mice revealed that a lack of *FTO* specifically impairs dopamine receptor D2/3-mediated control of neuronal activation which affect dopamine-dependent regulation of locomotor activity and reward sensitivity (Hess et al., 2013). Consistently, behavioral alterations associated with *FTO* variants in humans have also been linked to altered dopaminergic transmission (Kenny, 2011). However, the underlying neurobiological mechanisms by which *FTO*, or obesity predisposing variants of the human *FTO* gene, affect behavior, remain elusive.

Another genetic factor influencing D2R signaling and body weight is the TaqIA restriction fragment length polymorphism (rs1800497), located in the ankyrin repeat and protein kinase domain-containing protein (ANKK)1 gene, downstream from the D2R gene (Neville et al., 2004). Healthy individuals who carry the A1 allele, compared with those who do not, show diminished striatal D2R density (Joensson et al., 1999) and reduced glucose metabolism in dopaminoceptive regions involved in reward processing (Noble et al., 1997). This genetic trait has been shown to moderate (1) increased likelihood of obesity (Noble et al., 1994), (2) food reinforcement and intake, especially in obese individuals (Epstein et al., 2007), and (3) the association between neural responses and weight gain (Stice et al., 2008).

Given that *FTO* regulates dopaminergic signaling in mice and *ANKK1* affects D2R signaling in humans, therefore we hypothesized that *FTO* and *ANKK1* gene variants may interact to control D2-dependent behavior and associated neural responses. Such an interaction would provide direct evidence that *FTO* gene variants modulate D2-dependent neurotransmission in humans. To evaluate the individual contributions and potential interaction of *FTO* and *ANKK1* gene variants in dopamine-controlled behavior, the effect of genotype on reward and avoidance learning was studied. fMRI was used to investigate whether rewarding outcomes engage DA signaling depending on genotype. Prior findings from *FTO*-deficient mice (Hess et al., 2013) suggested that a lack of *FTO* specifically impairs D2/3R-mediated autoinhibition of dopaminergic midbrain neurons. Furthermore, the *ANKK1* genotype modulates midbrain response to rewards in humans (Felsted et al., 2010), and reward prediction errors (PEs) are encoded by phasic dopamine release from neurons in the ventral tegmental area/substantia nigra (VTA/SN) (Schultz et al., 1997; Montague et al., 2004). For these reasons, our primary analysis focused on PE signals in the midbrain. The modeling approach presented here brings the benefits of quantifying the effects of genetics on the behavior and its link to cognition.

3.2 Methods

3.2.1 Reinforcement Learning Models

A standard action-value (Q) learning algorithm (Watkins and Dayan, 1992) was utilized for modeling reward based learning behavior. Similar to the two-armed bandit problem, there are two actions to choose from: A and B. The update rule for an action value in each trial, i.e. the expected reward of selecting a particular action A, is given by the following equation:

$$Q_{i+1}(A) = Q_i(A) + \alpha\delta_i \tag{3.1}$$

where i is the current trial, α is the learning rate, and δ is the prediction error (PE), which is computed for any trial i:

$$\delta_i = r_i - Q_i(A) \tag{3.2}$$

where r_i is the reward on trial i, and takes the values $\{0, 1\}$. Therefore, in case of a reward, δ will be positive because reward is modeled with a value of 1; by contrast, a nonrewarding trial is modeled with a value of 0, resulting in a negative PE. The learning rate α scales the impact of the PE (i.e., the degree to which PE is used to update the action value). The softmax decision rule assigns probabilities for selecting each action (Sutton and Barto, 1998). For example at trial i, the probability of choosing action A is:

$$p_i(A) = \frac{e^{Q_i(A)/\beta}}{e^{Q_i(A)/\beta} + e^{Q_i(B)/\beta}} \tag{3.3}$$

The parameter β reflects the subject's individual bias toward either exploratory or exploitative behavior. Please refer to background chapter for a detailed description of the softmax decision rule.

Two learning rates model: Previous studies on reinforcement learning have suggested that humans may differ in learning from positive or negative PEs (e.g., (Niv et al., 2012)). It has therefore been proposed that separate learning rates may mediate updates in response to positive and negative PEs, respectively (Frank et al., 2007; Gershman, 2015):

$$Q_{i+1}(A) = \begin{cases} Q_i(A) + \alpha_+[r_i - Q_i(A)], & \text{if } r_i = 1. \\ Q_i(A) + \alpha_-[r_i - Q_i(A)], & \text{if } r_i = 0. \end{cases} \tag{3.4}$$

Therefore, in the following sections both models will be evaluated.

3.2.2 Simulations

In order to understand the behavior of an agent that utilizes the above mentioned learning strategies, simulations were performed for a stimulus pair with a reward schedule of %80 − %20, i.e. choosing one stimulus led to a reward in %80 of the trials, whereas choosing the other stimulus led to a punishment in these trials. Simulations were presented as an average of 50 runs. The first simulations evaluated the influence of different learning rates and temperatures by setting α to the values $\{0.1, 1\}$, and β to the values $\{0.1, 3\}$. These values were selected based on the boundary values of a reasonable range that is informed from earlier studies. For that reason, a similar range was also used on the empirical dataset during model fitting. Different initial conditions of action values were tested for a comparison of convergence to the real values. A second set of simulations targeted the case of having two learning rates which is used in RL literature to model an agent that learns differentially from positive and negative feedback.

3.2.3 Participants

Ninety-two healthy volunteers (45 male) participated in the study. Participants were selected based on the genetic stratification of a larger sample (589 health individuals) and differed according to their *FTO* (rs9939609 T/A variant) and *ANKK1* (rs1800497 G/A variant) genotype but were matched for similar age (26 ± 0.45 years), body mass index (BMI) (23 ± 0.22), and general intelligence (Table 5.1); participants were further assessed by the Beck Depression Inventory II (Beck et al., 1996) to preclude any acute depression. For reasons unrelated to these criteria, 13 further subjects had to be excluded from data analysis: two participants due to malfunction of the MR scanner, another for an incomplete test phase as the participant experienced panic inside the scanner. 10 others participants were excluded because they did not perform the task satisfactorily: An elimination criterion regarding the performance in the test phase was used such that subjects whose correct responses on AB trials were less frequent than wrong responses (A<B) were eliminated. In total, 79 subjects were included in further data analysis (Table 5.1). All participants gave written informed consent to participate in the experiment, which had been approved by the local ethics committee of the Medical Faculty of the University of Cologne (Cologne, Germany).

Table 3.1: Descriptive (mean ± SEM) data of participants, gender, age, Wechsler Adult Intelligence Scale-Matrice Scale, BMI

		Gender		**Age /years**		**WAIS-MS**		**BMI**		**BDI-II**	
Genotype	*# Subjects*	*m*	*f*	*Mean*	*SEM*	*Mean*	*SEM*	*Mean*	*SEM*	*Mean*	*SEM*
$A1^-FTO^-$	20	8	12	25	0.7	12	0.5	22.5	0.6	7.0	1.2
$A1^-FTO^+$	21	9	12	27	1.1	11.2	0.4	23.9	0.9	7.6	1.0
$A1^+FTO^-$	16	6	10	26	1.0	11.6	0.1	22.4	0.5	10.8	1.4
$A1^+FTO^+$	22	9	13	26	0.9	11.0	0.1	22.5	0.4	8.0	1.2

3.2.4 DNA isolation and SNP genotyping

Isolation of DNA from buccal swabs was performed using the QIAamp DNA Blood Mini Kit (# 51106, QIAGEN) according to the manufacturer's instructions. Concentration and quality of the DNA were determined with an ND-1000 UV/Vis- Spectrophotometer (Peqlab). SNP genotyping for rs9939609 (*FTO*) and rs1800497 (*ANKK1*) was performed with 20 ng of DNA in triplicates using allelic discrimination assays (TaqMan SNP Genotyping Assays, Applied Biosystems by Invitrogen). The genotyping PCR was performed on a 7900HT Fast Real-Time PCR System (Applied Biosystems), and the resulting fluorescence data were analyzed with Sequence Detection Software version 2.3 (Applied Biosystems).

3.2.5 Reinforcement Learning Task

After informed consent was obtained, participants completed the probabilistic selection task developed by Frank et al. (2004) and formerly applied in the same form by Jocham et al. (2011). It consisted of two phases: an initial reinforcement learning (training) phase and a subsequent transfer (test) phase. Both phases were performed during one fMRI session. During the learning phase, participants were presented with pairs of symbols that were probabilistically associated with reward. In each of three pairs, one symbol was always better (i.e., associated with a higher reward probability) than the other, but the differences in the reward probability were unequal across the three pairs. Symbol pairs were presented in random order, and subjects had to learn to choose the more frequently rewarded symbols from these pairs. Immediately after each choice, the outcome (a smiling face indicating a reward or a frowning face for no reward, see Fig. 3.1a) was presented. The three stimuli pairs were animal figures and associated with 80%-20%, 70%-30%, or 60%-40% of positive feedback (see Fig. 3.1b). This setup provided a varied learning scenario, including difficult-to-learn and easier-to-learn trials. Each pair was presented 120 times; the whole session comprised 401 trials, including 41 null events (black screen). After this reinforcement learning session, subjects underwent a test phase where the stimuli consisted of all 15 possible combinations of the 6 animal figures presented during the learning session. In this test phase, the subject was asked to choose the better option (or i.e., choose A trials) or to avoid the worse option (i.e., avoid B trials, see Fig. 3.1b) based on the previous experience with the stimulus pairs.

3.2.6 Statistical Analysis

To statistically evaluate differences between *FTO* and *ANKK1* gene variants on choice behavior, unpaired t tests were performed; before this and to test their interaction, an ordinary one-way ANOVA was calculated. Before all statistical calculations were performed, a D'Agostino- Pearson omnibus normality test was completed to verify that the data were compatible with a normal distribution. A significance level of p = 0.05 was chosen in all statistical tests.

3.2.7 Parameter Estimation and Model Comparison

Both learning models described above were fitted to the participants' behavior in the reinforcement learning phase. The models learn the action values, Q(A), Q(B),..., Q(F) for each of the six stimuli, A to F (Eq. 3.1 and 3.4). Then, the probability of selecting a particular stimulus is calculated via softmax function (Eq. 3.3). In order to fit the free model parameters ($\theta_1 = \alpha$, β) and ($\theta_2 = \alpha_+$, α_-, β) to binary choice data y, maximum likelihood estimates were calculated.

$$\hat{\theta} = \arg\max_{\theta} log(p(y|\theta)) \tag{3.5}$$

A systematic grid search procedure examined both parameters, from 0.01 to 1 for α, and from 0.01 to 3 for β, with a step size of 0.01. For the two learning rates model, 3D matrices were used for faster computations (one dimension for each parameter).

The two forms of Q learning models were compared by using the Bayesian Information Criterion (BIC) and random effects Bayesian model selection. The BIC is based on Laplace

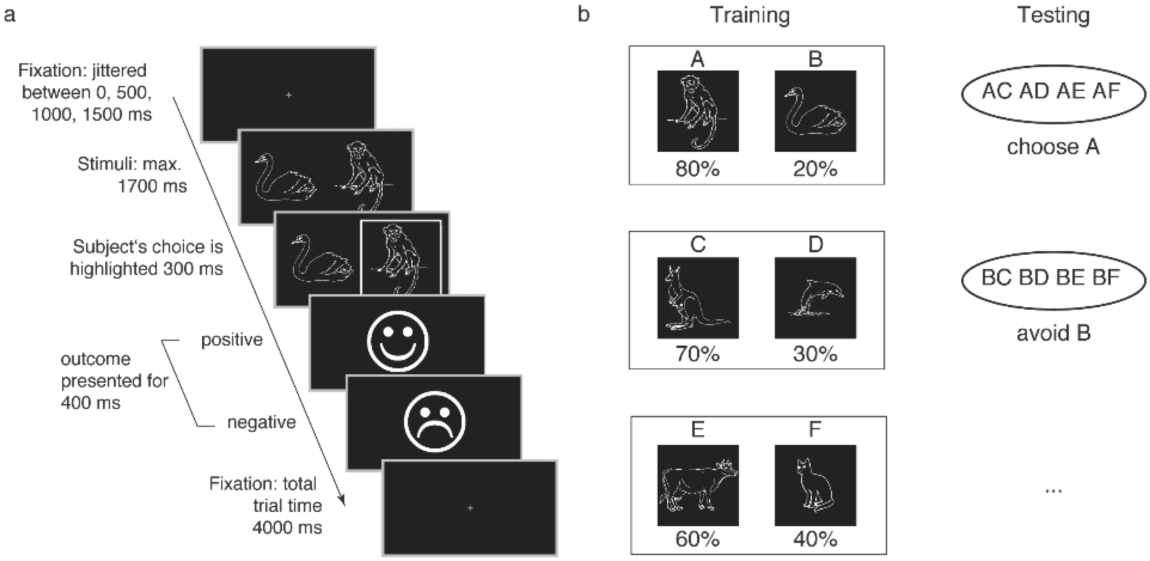

a
Fixation: jittered between 0, 500, 1000, 1500 ms
Stimuli: max. 1700 ms
Subject's choice is highlighted 300 ms
outcome presented for 400 ms
positive
negative
Fixation: total trial time 4000 ms
b
Training
Testing
A
B
80%
20%
C
D
70%
30%
E
F
60%
40%
AC AD AE AF
choose A
BC BD BE BF
avoid B
...

Figure 3.1: Probabilistic Selection Task (PST). a) Schematic task sequence, event order and durations within a trial. Following selection of one of the two stimuli, the choice was visualized by a white frame. This was immediately followed by positive or negative feedback, according to the task schedule. b) Pairs of stimuli associated with different reward probabilities (percent positive feedback). In the subsequent test phase, new combinations of the stimuli are presented in order to assess participants' performance on learning more from negative feedback or from positive feedback. Trials were identical to those from the learning phase, with the exception that no outcome was presented.

approximation to the model evidence, at the asymptotic limit $n \to \infty$:

$$BIC = -2(logL) + l * \log(n) \tag{3.6}$$

where n is the sample size and l is the number of free parameters. Thus, this criterion requires large number of observations compared to the number of free parameters. To compare the BIC values of models in our population, random effects Bayesian model selection was performed. This procedure exploits flat priors over the model frequencies $p(r|H_1)$ that is a Dirichlet distribution. Flat priors are obtained by setting Dirichlet concentration parameter $\alpha = 1$, so that expected value of k^{th} model is $E[r_k|H_1] = 1/K$. The common way in Bayesian model comparison is to report exceedance probabilities for each model i.e. $\phi_k = P(r_k > r_{k' \neq k}|y, H_1)$. This is the probability that the k^{th} model is more frequent than other models given the group data and the flat priors that assign a uniform probability to each model (Stephan et al., 2009a). However, a recent suggestion is to take into account the statistical risk of the differences in model evidences due to chance. This is also known as Bayesian omnibus risk and caused by the erroneous choice of priors H_1 over H_0. It calculates the chance likelihood of observed data (Rigoux et al., 2014):

$$P_0 = \frac{P(y|H_0)}{P(y|H_0) + P(y|H_1)} \tag{3.7}$$

where null H_0 is obtained by the limit of Dirichlet concentration parameter $\alpha \to \infty$. At this limit, only uniform distributions over the model frequencies are likely. Unlike H_1, H_0 implies that model frequencies are fixed (prior variance is zero) and equal to each other, i.e. $r_k = 1/K$. To account for this risk, Rigoux et al. (2014) introduced protected exceedance probability $\tilde{\phi_k}$ by computing a Bayesian average of exceedance probability:

$$\begin{aligned}\tilde{\phi_k} &= P(r_k > r_{k' \neq k}|y) && (3.8)\\ &= P(r_k > r_{k' \neq k}|y, H_1)P(H_1|y) + P(r_k > r_{k' \neq k}|y, H_0)P(H_0|y) && (3.9)\\ &= \phi_k(1 - P_0) + \frac{1}{K}P_0 && (3.10)\end{aligned}$$

While the random effects Bayesian model comparison usually takes into account the free energy, protected exceedance probability approach is adopted here for accounting the variability in the BIC that is also an approximation to the log model evidence as mentioned above.

3.2.8 fMRI Acquisition

Imaging was performed on a Siemens 3T Trio scanner (Erlangen, Germany; maximum gradient strength 40 mT/m). Functional time series of each subject were acquired with a TxRx head coil (Siemens, Erlangen, Germany). For the functional time series, 30 axial slices (field of view 192 mm x 192 mm, thickness 3 mm, 0.3 mm interslice gap, 64 x 64 pixel matrix) parallel to the commissural line (AC-PC) were acquired in a descending order from top to bottom using a single shot gradient echo-planar imaging sequence (EPI: TR = 2000 ms, TE = 30 ms, bandwidth = 116 kHz, flip angle 90°). Additionally, high-resolution T1-weighted images were acquired in a separate scanning session using a 12-channel array head coil with a whole-brain field of view (MDEFT3D: TR = 1930 ms, TI = 650 ms, TE = 5.8 ms, 128 sagittal slices, resolution = 1 x 1 x 1.25 mm^3, flip angle = 18°).

3.2.9 fMRI Data Analysis

Functional MRI data were analysed using statistical parametric mapping (SPM8; Wellcome Department of Imaging Neuroscience, London, UK) in Matlab 7.12 (Mathworks, Inc., Sherborn, MA). Following re-alignment of the functional images and co-registration of the structural image to the mean functional image, we segmented the structural image and normalized both functional and structural images to a standard template in Montreal Neurological Institute (MNI) coordinate space. The functional images were smoothed by applying an 8 mm full-width at half maximum Gaussian kernel and resampled to isotropic resolution. Additionally a high-pass filter with a cutoff of 1/128 Hz was applied to remove all slowly varying signals from functional data.

Preprocessed scans from the learning phase were analysed with general linear models (GLM), using maximum likelihood estimation for serially auto-correlated observations at the first level (Worsley and Friston, 1995), with SPM8 (version 4290). The design matrix comprised regressors for reward and punishment onsets as well as the motion parameters, and the positive and negative prediction errors separately derived from the model as parametric modulators (Fig. 3.2). Preprocessed scans from the test phase were modelled with a separate GLM at single subject level, where onsets for choose A and avoid B, as well as onsets for events of no interest (any stimuli except choose A and avoid B), and the motion parameters were included as regressors.

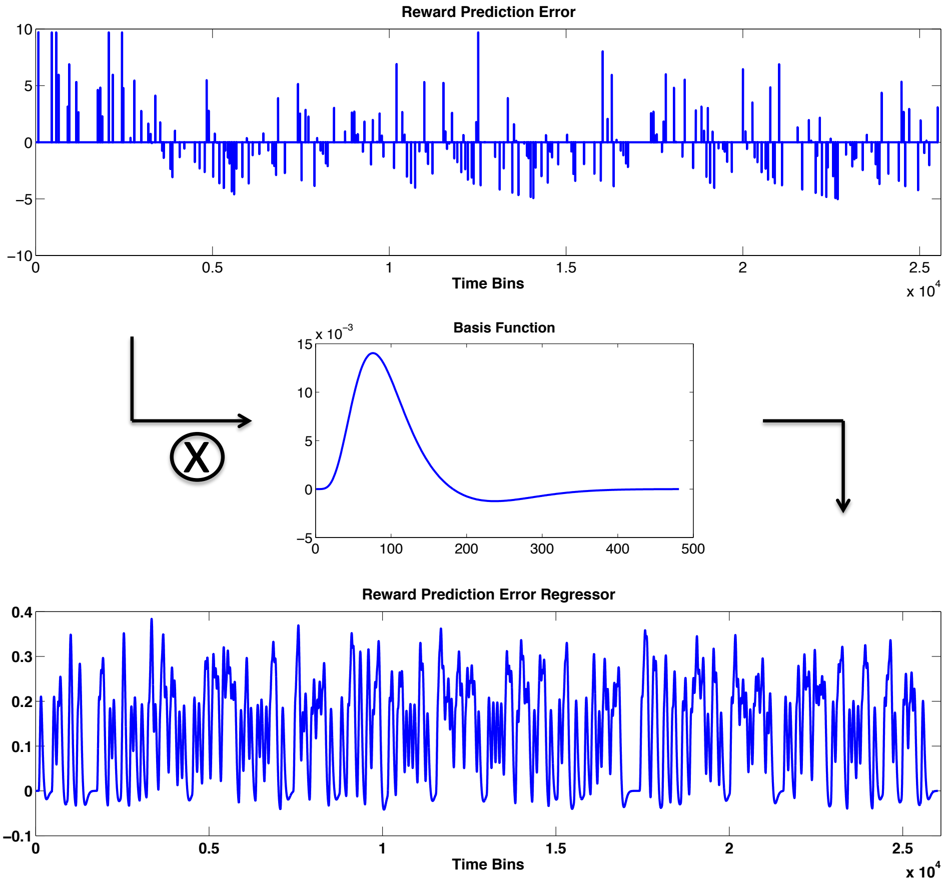


Figure 3.2: Parametric regressors. Convolution of prediction errors with hemodynamic response function.

3.2.10 Volume of Interest Analysis

Relevant single subject activations were further evaluated with volume-of-interest (VOI) analysis: Based on prior finding from *FTO*-deficient mice (Hess et al., 2013) that a lack of *FTO* specifically impairs D2/3R-mediated autoinhibition of dopamine neurons in the midbrain, an anatomical mask was applied for VOI analysis of the ventral tegmental area (VTA)/substantia nigra (SN) (Bunzeck and Düzel, 2006) (For the location of this mask, please see Fig. 3.11b which shows the activation within this mask).

After individual fMRI data from learning and transfer phases had been subjected to GLM analysis and relevant contrasts had been estimated, the peak effect size was searched within the VTA/SN VOI for each condition at the single subject level, using the RFXplot toolbox (Glaescher, 2009). To statistically test differences between *FTO* and *ANKK1* gene variants on VTA/SN activation, unpaired t-tests were carried out; to test their interaction an ordinary one-way analysis of variance (ANOVA) was calculated. Prior to all statistical calculations, a

D'Agostino-Pearson omnibus normality test was done to verify that the data were compatible with a normal distribution. A significance level of $P = 0.05$ was chosen in all statistical tests.

3.3 Results

3.3.1 Simulations

The course of reinforcement learning for different agents in an environment with a reward schedule of 80%−20% is shown in Fig. 3.3. The dramatic influence of the value of temperature parameter on exploration-exploitation behavior can be observed on the first and second plots. The first agent with $\beta = 0.1$ follows an exploitative course of action as it avoids exploring the other cue. As a result, the agent never learns its true value, i.e. estimated action value for B is 0 at the end of the training. In contrast, the second agent with a very high temperature parameter ($\beta = 3$) follows a more explorative strategy, hence, the estimated value of B approaches the real value that is 0.2. Further, a higher alpha ($\alpha = 1$) causes bigger steps in the value updates that are equal to the current value (Fig. 3.3, third plot). This results in an unstable course of actions and quickly diverging from the real value in comparison to the other agents whose updates converged slowly to the real values ($Q_A = 0.8, Q_B = 0.2$). Finally, choosing arbitrary values for initial conditions ($Q_{initial} = 0.5$ for both cues), did not influence the overall training performance, as it approached similar states in both initializations (Fig. 3.3, bottom plot). These results show that different values for α and β influence the policy of an agent in the same environment independent from the initial state.

Figure 3.4 depicts the results of the second set of simulations for a model with two learning rates. Substituting Eq. 3.2 in Eq. 3.4 resulted in an update of $-\alpha_- Q_i$ for each loss, and $-\alpha_+ Q_i + \alpha_+$ for each reward. A big value of α_- and a low value of α_+ will force action values to a smaller range by penalizing them largely for losses, while not allowing the agent learn "enough" from a reward (Fig. 3.4, middle plot). The opposite case is observed in the bottom plot where the action values are kept in a range higher than their real values. Having evaluated the learning of different agents for both models, the following results will present the application of both models in a real dataset to explain which strategy better describes human behavior.

3.3.2 Behavioral Results

Comparing behavioural performance on the probabilistic learning task between *FTO* genotypes revealed no significant difference during the choose A trials (Fig. 3.5a), while correct choices during the avoid B trials were significantly reduced in the *FTO+* compared to the *FTO-* group of participants (group x choice interaction: $F_{1,152} = 6.22, P = 0.014$). Similarly, and in line with previous studies (Klein et al., 2007), while correct choices during choose A trials did not differ significantly between *A1-* and *A1+* individuals (Fig. 3.5b), correct choices during the avoid B trials showed a tendency to be significantly reduced in *A1+* compared to *A1-* individuals (group x choice interaction: $F_{1,152} = 3.05, P = 0.08$). Interestingly, comparing the effect of combined *FTO* and *ANKK1* genotypes revealed a trend towards a reduction of correct choices during the choose A trial only between *FTO- A1-* versus *FTO+ A1+* carriers (Fig. 3.5e). However, there was a robust reduction of correct choices during avoid B trials in a gene dosage-dependent manner, i.e., correct choices to avoid B decreased in the presence of either the *FTO+* or *A1+* allele and carriers of the combination of both at-risk alleles performed significantly worse than carriers of the individual at-risk alleles (gene x gene interaction: $F_{3,74} = 2.88, P = 0.041$; Fig. 3.5f). These experiments indicate that *FTO* gene variants affect D2-dependent learning from negative outcomes and that group differences in learning behaviour are determined by the combination of both genotypes, which might point towards a genetic interaction of *FTO*- and *ANKK1*-regulated learning processes.

3.3.3 Reinforcement Learning Model Comparison

Statistical model comparison indicated that a model with two distinct learning rates was inferior to a model with a single learning rate (i.e., assuming two learning rates did not explain the behavioral data better than using a single learning rate when taking into account the added model complexity afforded by the additional parameter). Specifically, using the Bayesian Information Criterion (Schwarz, 1978) and random effects Bayesian model selection (Stephan et al., 2009a) we showed that the more parsimonious model with a single learning rate was favoured very strongly, with a protected exceedance probability of 0.995 (Rigoux et al., 2014). As a consequence, the results of this model were used for all subsequent analyses. For the winning model, the trajectories of action values for each stimulus of two example subjects are displayed in Fig. 3.6. The high and low learning rates of the two subjects resulted in bigger and smaller steps in updating their beliefs, respectively. On average, the subjects learned the real value of each stimulus at the end of the experiment (Fig. 3.7).

3.3.4 Whole Brain Activations

Prediction error (PE) processing: Group level statistical maps revealed significant activation in the midbrain, ventral striatum (vStr), anterior cingulate cortex (ACC) and insula for the ‘positive PE’ contrast (Fig. 3.10). The same regions were activated in response to ‘negative PE’.

Reward and punishment processing: As expected, we observed activations in the brain regions that are associated with reward processing (for a broad overview, please see the background chapter). Significant increases in BOLD activity to ‘reward-punishment’ contrast were found in the striatum and prefrontal cortex, whereas ‘punish- ment-reward’ contrast was associated with significant activation in the ACC, insula and midbrain (Figs. 3.8 and 3.9).

3.3.5 Volume of Interest Analysis

To address the effect of the *FTO* gene variants, the *ANKK1* gene variant, and their interaction, on neuronal activation, fMRI was used to investigate whether rewarding outcomes engaged DA neurons and if this was dependent on genotype. Here, the primary analysis focused on PE processing in the VTA/SN and therefore the volume of interest (VOI) analysis was based in this area.

fMRI measurements of VTA/SN activity revealed a significantly reduced positive PE response in a gene-dosage-dependent manner (i.e., the peak effect size associated with neural response of the positive PE in VTA/SN decreased in the presence of either the *FTO* or *A1* allele), and carriers of both risk alleles exhibited significantly reduced PE responses compared with noncarriers of the individual at-risk alleles (*FTO-A1-*; Fig. 3.11a).

3.3.6 Prediction Error Processing and Avoidance Learning

Strikingly, reduced PE responses in the dopaminergic VTA/SN (Fig. 3.11b) were associated with poorer ability to avoid negative outcomes during a later test phase (Fig. 3.12d), whereas learning to select the most rewarding stimulus (choose A) did not correlate with a positive PE response in VTA/SN (Fig. 3.12c). These results demonstrate that the *FTO* gene variants alter midbrain responses during reward learning, which is in turn associated with impaired avoidance learning. Again, a gene x gene interaction with *ANKK1* variants was found ($F(3,74) = 3.82, p = 0.013$) in the BOLD responses for prediction error processing. We did not observe a genotype effect on model parameters which were fitted with the single learning rate model (the winning model) separately for each participant given their choice behavior.

3.4 Discussion

In this chapter, some of the most widely used RL algorithms to study dopaminergic function were evaluated with simulations and an empirical dataset. Their capability in reflecting the differences in midbrain function was presented with RL informed fMRI analysis. This approach enabled us to make associations between prediction error processing in the brain and the performance in learning from negative reinforcement, both of which are influenced by the underlying genetic causes.

RL algorithms have proven to converge even for a large number of states like TD-Backgammon with 10^{20} states (Tesauro, 1995). The task that was used in the experiment had one state and two possible actions in each trial. Simulated behavior was presented in order to gain more insight about model parameters and convergence towards the true action values. The simulations were run on the easiest-to-learn trials of the probabilistic selection task, where the reward schedule was $80\% - 20\%$. Plots were generated as a result of 50 runs. The first set of simulations utilized the update rule with a single learning rate (Eq. 3.1). The agent with a low temperature ($\beta = 0.1$) avoided choosing the low reward frequency option except for the first trials. This can be explained by the lasting effect of first experiences on the subsequent behavior during training, which might result in the underweighting of rare events due to an underestimating of small probabilities, as suggested by Shteingart et al. (2013). On the contrary, the agent with a higher temperature parameter continued exploring both actions despite the discrepancy in their reward frequency. This is explicitly seen from the equation 3.3 where a higher β results in an equally likely choice of action for both options. Having shown that different parameters result in unique trajectories, it can be suggested that different learners can be identified on empirical data.

To evaluate the influence of initial conditions on the dynamics of learning, simulations were repeated with an initial action value of 0.5. After the early trials, learning behavior converged to the same curve as in the zero initialization. This observation suggests that initialization does not influence the speed of convergence to the real values considerably such that an agent can learn starting with no knowledge of the environment. A final conclusion to draw from this set of simulations was that the last agent with a high learning rate did not learn, but constantly updated its belief with big steps. Therefore, in such environments it is not recommended to set a high learning rate. Another alternative could be to use a dynamic learning rate. This approach is demonstrated in Chapter 5. The second set of simulations utilized separate learning rates for rewards and punishments, α_+ and α_-, respectively. Having the same value for both learning rates yielded a learning trajectory similar to the one learning rate model. However, the agent became too sensitive to gains or losses when the learning rate for one type of feedback was much bigger compared to the other learning rate ($\alpha_+ = 7\alpha_-$). Nevertheless, models with two separate learning rates are not trivial. In fact, it has been shown that using adaptive learning rates for positive and negative prediction errors by a running average of reward history, can improve the agent's performance (Cazé and van der Meer, 2013). In addition to the biological motivation of the separate basal ganglia pathways hypothesis, it can also provide a quantitative tool for studying psychological traits such as sensitivity to reward or punishment.

Following the simulations, the cognitive strategy that was adhered to by the participants, was presented. When RL models are adapted to study human learning, the space of possible models needs to be defined based on biological evidence. Also, model comparison needs to be performed carefully, as the resulting prediction errors will influence the consecutive fMRI analysis. Although having two learning rates did not add computational complexity, the added model complexity needs to be addressed during model comparison because, in general, the more number of parameters there are, the greater the chance of overfitting is. To account for the differences in the number of free parameters, the Bayesian information criterion was used for an approximation to the model evidence. The BIC is a valid approximation to log model evidence for the data and models presented here, as the sample size or number of trials (360), is much larger than the number of model parameters (2 and 3). However, it has been shown that the BIC can overestimate model complexity (Rigoux et al., 2014). To provide a statistically sound model comparison procedure, the statistical risk of the models having the same frequency (see methods section) is discounted by using protected exceedance probabilities. Comparing these two models with the real dataset, favored the model with one learning rate. Therefore, this model was used to inform fMRI analysis.

Activation likelihood meta analyses have shown that prediction error maps can be sensitive to model-fitting procedures (Chase et al., 2015b). During RL model fitting, log likelihood was maximized for all pairs of stimuli at once, which assumed that the participant learned the values of different pairs with the same learning rate and temperature parameter. Although this might not be the case in real life, judging by the fact that the estimated action values were close to the real values at the end of the experiment (Fig. 3.7), one can conclude that this fitting procedure performs reasonably well.

Finally, the power of combining RL and fMRI in capturing differences at the midbrain activation due to genotype was evaluated on a genetically matched population. Variations in the *FTO* gene have been robustly linked to obesity across multiple studies and ethnicities (Frayling et al., 2007). The underlying mechanism explaining how the *FTO* gene product contributes to obesity-related behaviors has remained largely unclear. Based on recent evidence that *FTO* regulates D2/3R signaling in mice, it was tested whether obesity-predisposing variants of *FTO* in humans, would influence D2R-dependent behavioral and neural responses during a reward and avoidance learning task. Furthermore, we examined whether they would interact with variants of *ANKK1*, which is also associated with obesity and D2R signaling (Noble et al., 1994; Stice et al., 2008) to influence behavioral, neural, and perceptual responses during a reward learning and in response to reward. The present behavioral and fMRI analyses indeed revealed an increasing effect of these gene variants and thereby suggest a role for *FTO* variants in regulating reward learning in humans. Both pharmacological and genetic studies have linked the ability to learn from positive and negative feedback to dopamine dependent neurotransmission within the basal ganglia neurcircuitry (Hikida et al., 2010; Frank and Fossella, 2011; Chowdhury et al., 2013). The direct pathway, populated mostly by D1R expressing neurons, is critical for optimizing behavior based on positive outcomes and error signaling. The indirect pathways, using mostly D2/3R-expressing neurons, are critical in optimizing behavior based on negative outcomes and error monitoring (Frank and Fossella, 2011). Consistent with the hypothesis that *FTO* and *ANKK1* variants produce synergistic effects on D2/3 receptor neurotransmission, individuals possessing both at-risk alleles performed significantly worse on negative, but not positive, outcome learning. Also, reduced responses were found in the VTA/SN associated with PE signaling in a gene-dosage dependent fashion, with reduced responses in carriers of a single at-risk allele and further reductions in carriers of both at-risk alleles. This diminished midbrain response during the generation of PEs was associated with the magnitude of impaired performance in negative outcome learning. No associations were observed here with positive outcome learning, which was unaffected by genotype.

A limitation of this study concerns the imaging method fMRI. One interpretational caveat concerning all fMRI studies of midbrain activity or connectivity, is that BOLD signals from the midbrain are not guaranteed to reflect the activity of dopaminergic neurons as the midbrain is heterogeneous in cellular composition and also contains GABAergic (Steffensen et al., 1998; Korotkova et al., 2004), and a small proportion of glutamatergic neurons (Morales and Root, 2014), such an anatomical complexity is paralleled by a functional complexity because dopaminergic neurons can co-release glutamate or GABA (Pignatelli and Bonci, 2015). However, as demonstrated by multimodal investigations of the correspondence between striatal DA release and midbrain BOLD activity in response to reward PEs or novel stimuli (Düzel et al., 2009), for paradigms specifically probing (reward) PEs, one may be relatively confident that phasic BOLD responses mainly arise from the dopamine neuron activity. Additionally, the genetic effects that were investigated here have an established biological relation to dopamine signaling; the *ANKK1* gene is known to affect D2R density (Pohjalainen et al., 1998; Joensson et al., 1999), and *FTO* affects DRD2-signaling in mice (Hess et al., 2013).

In conclusion, a detailed analysis of learning behavior with a Q-learning algorithm as well as fMRI data revealed genetic differences of neurocomputational processes. These differences were reflected in the dopaminergic midbrain when fMRI analysis was informed with model derived prediction errors. This study suggests that computational models of learning provides a powerful tool to understand neuronal and behavioral processes and to test related hypotheses quantitatively.

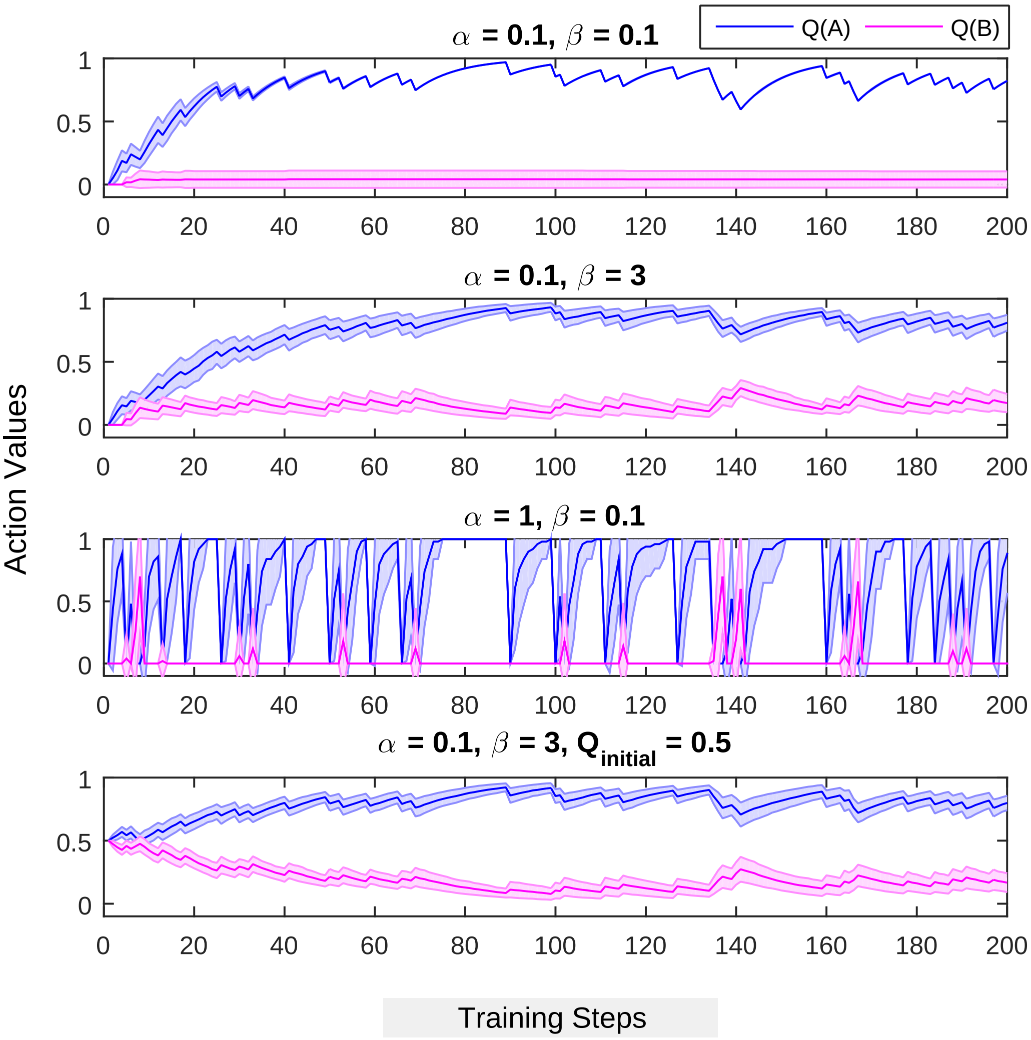


Figure 3.3: Simulated action values (Q) with the single learning rate model for different learners. The blue lines show the learning course of stimulus A which has a winning probability of 0.8. Similarly, the red line shows the learning course for stimulus B which has a winning probability of 0.2. Each agent follows a unique trajectory in the course of training depending on the given values of learning rate and temperature parameter (top three plots). Bottom: The choice of initial values of Q ($Q_{initial} = 0.5$ for both cues), did not impact learning. The trajectories were similar after the first 40 steps for both initializations. The results are average of 50 runs, and the shaded areas represent the standard error of the mean.

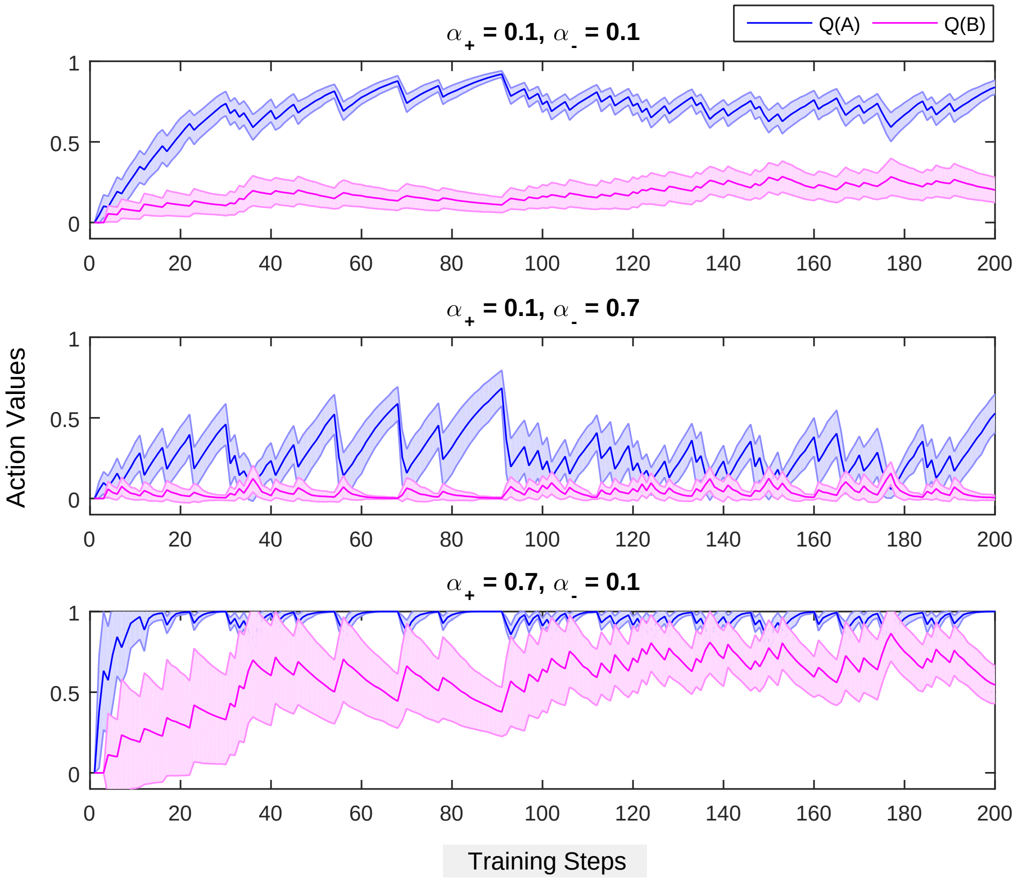


Figure 3.4: Simulations for the Q-learning model with two learning rates, α_+ and α_-. The blue lines show the learning course of stimulus A which has a winning probability of 0.8. Similarly, the red line shows the learning course for stimulus B which has a winning probability of 0.2. Top: When the learning rates are the same, agent behaves similar to an agent with a single learning rate. Middle and bottom: Disproportionate values of two learning rates do not allow the agent to learn the real values of both actions (see text for explanation). Temperature parameter was set to 0.7 in all three cases. The results are average of 50 runs, and shaded areas represent the standard error of the mean.

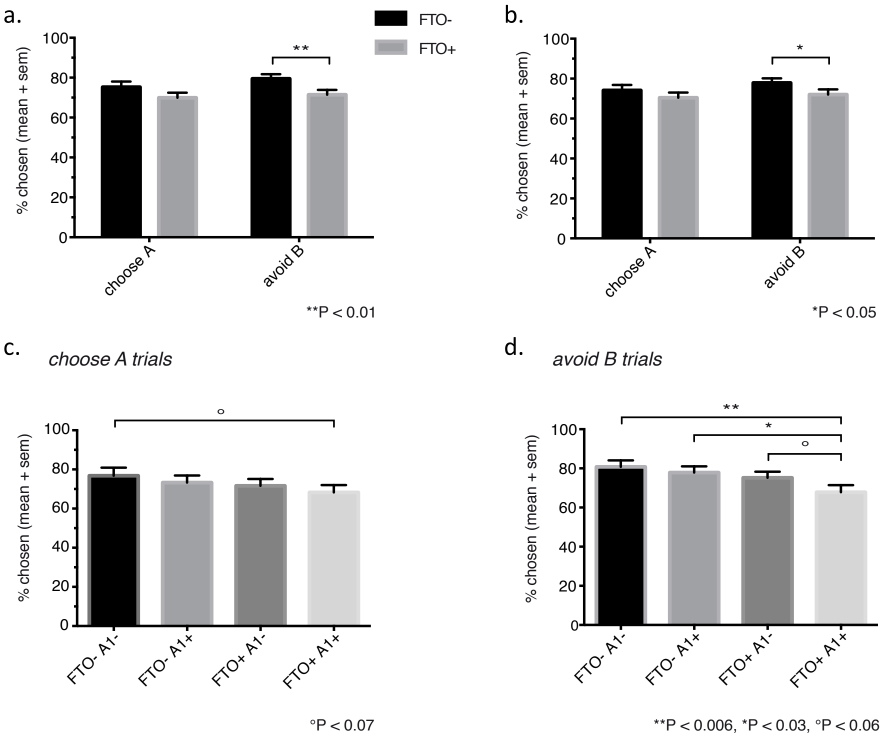


Figure 3.5: Results of the behavioural post test. (a) Choosing the better option A and avoiding the worse option B differs between the *FTO* groups; correct choices during avoid B trials are significantly reduced in the *FTO+* group ($p = 0.009$), but there is no significant reduction in choose A trials. (b) Behavior also differs between groups defined by *ANKK1* genotype: choose A trials did not significantly differ between *A1-* and *A1+* individuals, while correct choices during avoid B trials were significantly reduced in *A1+* individuals ($p = 0.026$). (c) Combined *FTO* and *ANKK1* genotypes do not show statistically significant differences on choose A trials, but a trend towards a reduction of correct choices on these trials between *FTO-A1-* and *FTO+A1+* carriers ($p = 0.065$). (d) Reduction of correct choices during avoid B trials in a gene dosage-dependent manner; choices decreased in the presence of either the *FTO+* or *A1+* allele, and carriers of the combination of both risk alleles performed significantly worse than non-carriers. Values are mean ± SEM.

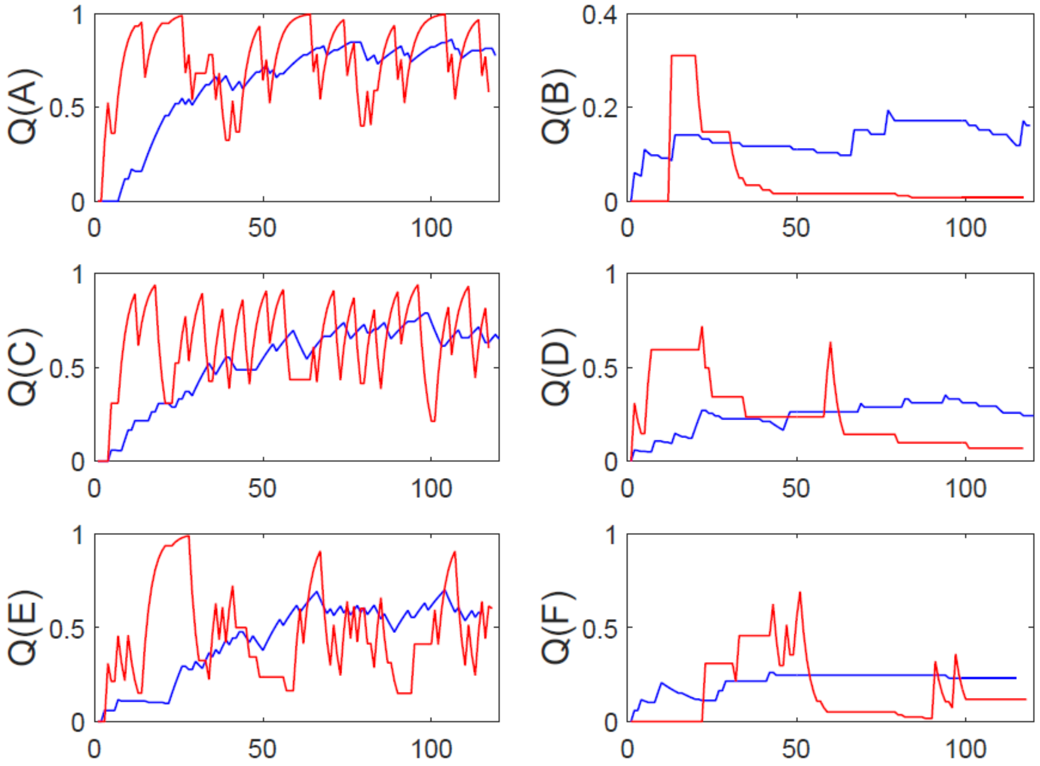


Figure 3.6: Trajectories for action values. Q-value of each stimulus across the trials is plotted for two subjects with high (red) and low learning rates (blue), $\alpha = 0.31$ and $\alpha = 0.06$, and with similar decision temperatures, $\beta = 0.23$ and $\beta = 0.27$, respectively.

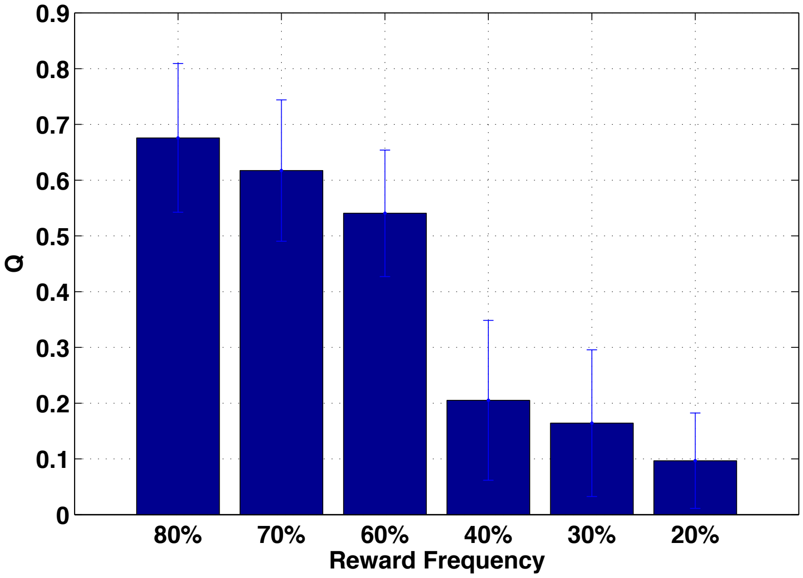


Figure 3.7: Learned action values at the end of the training phase. Bars and lines indicate means and standard deviations across the subjects.

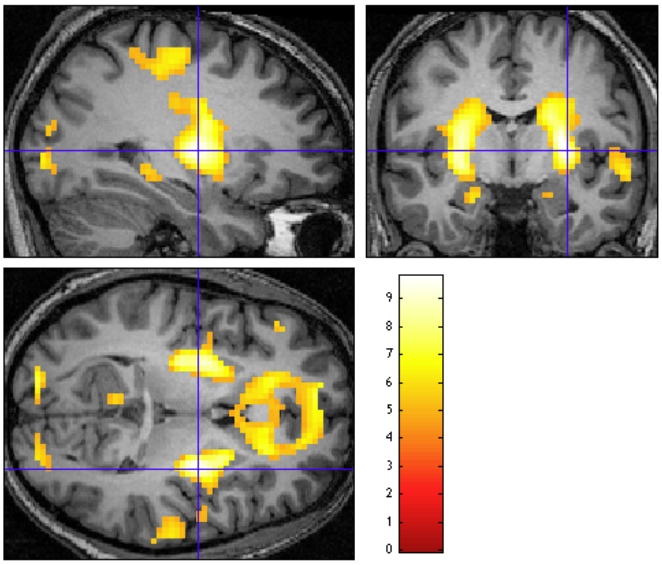


Figure 3.8: Brain regions activated at Reward > Punishment contrast. The crosshair is at the global maximum in the right striatum [30, -7, 7], $p<0.05$ (FWE-corrected).

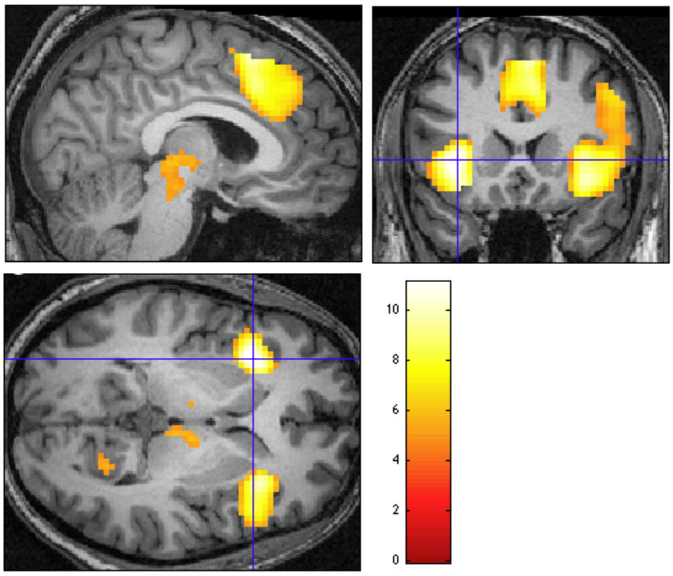


Figure 3.9: Brain regions activated at Punishment > Reward contrast. The crosshair is at the global maximum in left insula [-33, -20, 4] for coronal and axial slices. Sagittal slice on top left corner shows midbrain activation ($x = 9$), $p<0.05$ (FWE-corrected).

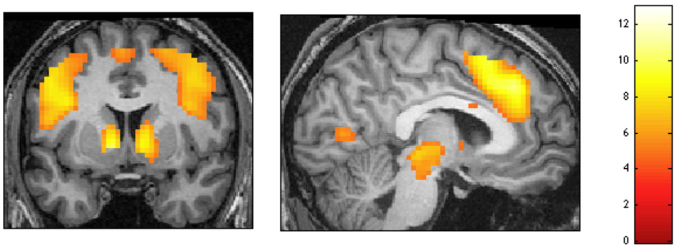


Figure 3.10: Brain regions activated at positive prediction error contrast. Left: The coronial axis displays vStr and insula activation ($y = 6$). Right: Sagittal slice ($x = -4$) displays the activation in the ACC and midbrain, p<0.05 (FWE-corrected).

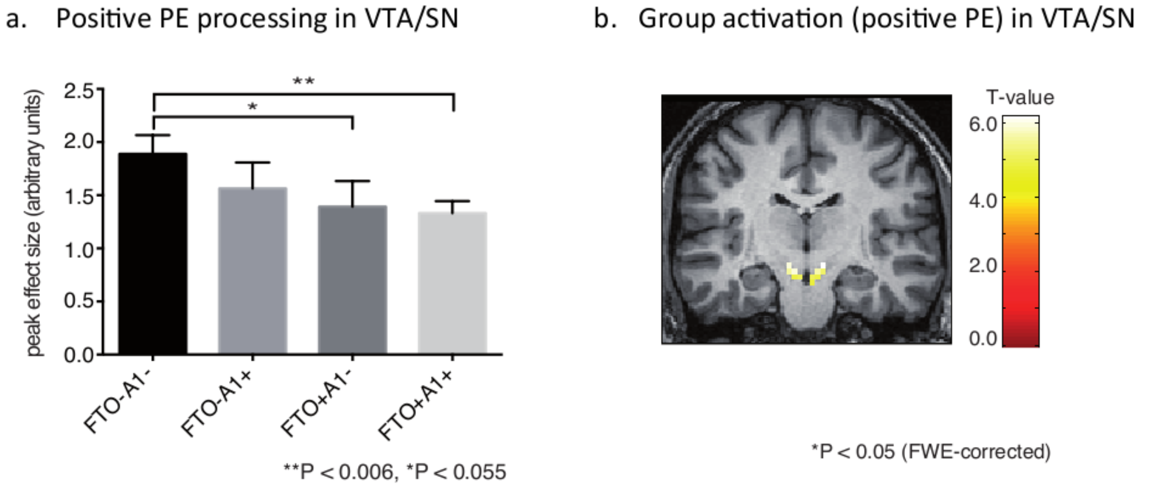


Figure 3.11: Differences of VTA/SN peak activation in response to positive prediction errors with regard to the interaction of *FTO* and *ANKK1* gene variants. Values are mean ± SEM. (b) Positive prediction error responses within the VTA/SN VOI ($Y = -18$), p<0.05 (FWE-corrected).

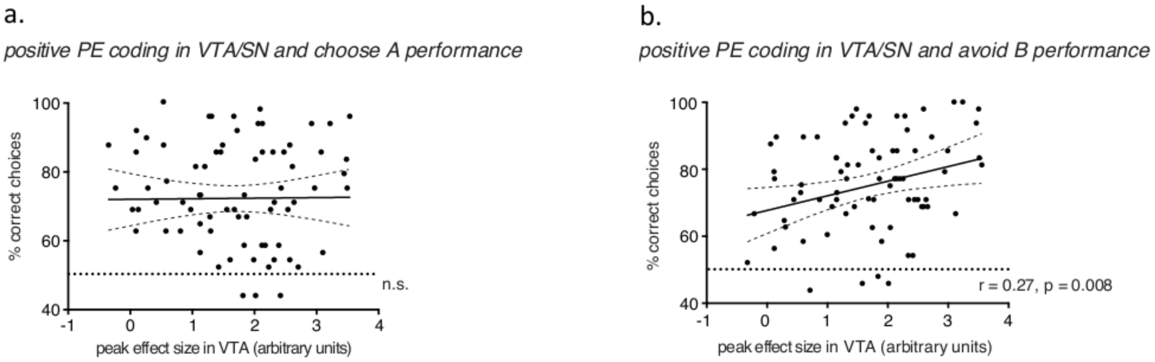


Figure 3.12: Association of prediction error representaion in midbrain and avoidance learning. a) VTA/SN peak activation to positive prediction errors correlated with correct choices on choose A and b) avoid B trials; the dashed lines denote the 95% confidence interval for the linear regression (solid line).

4 Dynamic Causal Modeling of Reinforcement Learning Circuitry in the Brain

4.1 Introduction

Reinforcement learning theory states that reward prediction errors are used to update action values about the available rewards in an environment (Sutton and Barto, 1998). From a neuronal perspective, it is thought that midbrain dopaminergic neurons encode and signal reward prediction errors (Schultz et al., 1997) and that they regulate cortico-striatal synaptic plasticity (van den Bos et al., 2012; den Ouden et al., 2010). In this chapter Dynamic causal models (DCMs) will be presented as a framework for assessment of the connectivity changes during reinforcement learning.

One key explanation to this mechanism is the free energy principle: "A system can minimize free-energy by changing its configuration to change the way it samples the environment, or to change its expectations." (Friston and Stephan, 2007) so that synaptic plasticity during learning reflects prediction error dependent changes in functional connectivity. There is evidence that prediction error modulates synaptic plasticity between auditory and visual areas during associative learning, as assessed by bilinear DCM (den Ouden et al., 2009), and the connectivity between cortical and motor regions by prediction error related activity in striatum, as revealed by nonlinear DCM (den Ouden et al., 2010).

DCMs are used for modeling causal interactions of neuronal populations. State equations describe the change in the neuronal activity of a brain region, which can be a function of different factors: (i) the activity of the region itself, (ii) a driving input onto this region, (iii) output of the activity of another region weighted by the synaptic connectivity strength. This connectivity strength can be modulated by another experimental input, or by the activity of another region. Bilinear state equation (Friston et al., 2003) for each region is given by equations 2.40 and 2.41. To model the modulatory effect of a regional activation onto the connectivity strengths, a forth term is included, which yields the nonlinear state equation (Stephan et al., 2008) (see equations 2.49 and 2.50), where D is the matrix describing the strength of this nonlinear modulation. After integrating state equations and combining them with hemodynamic model (Friston et al., 2000), posterior means and covariances of all the model parameters are estimated with a Bayesian inversion scheme (Friston et al., 2003). In this study, bilinear and nonlinear DCMs will be implemented to model the modulatory effect of prediction errors on the midbrain activity as well as the modulatory effect of prediction error coding in midbrain onto projection sites. Biological underpinnings of gating mechanisms in the reward processing network of the brain have been reported (Park et al., 2012; D'Ardenne et al., 2012). This study provides further evidence that nonlinear DCMs are useful connectivity models for testing the gating role of midbrain acitvation on the connections within a reinforcement learning circuit.

Little is known about the modulation of reward network and the integration of reinforcement learning signals in the human brain. It is yet unknown how these regions exert influences onto each other during reward outcome processing and how the prediction errors modulate the connectivity within this network. Therefore, in this chapter computational modeling of behavior will be combined with a causal model of brain regions. The goal of this work is to present DCM as a method for understanding the causal relationship in a reward network by using fMRI data from a probabilistic selection task (please refer to Chapter 3 for details of the experiment).

Finally, functional coupling can vary between individuals. For example, during a reinforcement learning task functional connectivity between brain regions, where the activation reflected the Q value, differed in learners compared to non-learners (Horga et al., 2015). Genetic factors can affect the levels of dopamine, which in turn will influence the connectivity of the reward and prediction error processing regions during reward based learning. It will be tested if the connectivity strengths can reflect genotype differences and the associated learning behavior. It is hypothesized that (i) the effective connectivity between these regions changes during reinforcement learning, in addition to the midbrain activity derived by prediction error input, and (ii) carriers of *FTO* and *ANKK1* allele have an altered connectivity strength compared to non-carriers, (iii) which in turn predicts the difference in learning.

4.2 Methods

4.2.1 Model Construction

The details of the experiment and genotyping are documented elsewhere (see chapter 3). A simple three-region DCM was constructed in order to infer connectivity strengths in a basic reward circuit, including mesolimbic and mesocortical efferents of the dopaminergic midbrain (VTA/SN). Specifically, the effective connectivity between these regions were quantified: (i) VTA/SN, whose dopaminergic neurons encode reward prediction errors by phasic dopamine release, (ii) ventral striatum or nucleus accumbens (NAcc), which plays a central role in reward processing and receives massive dopaminergic input from the midbrain, and mPFC, which is crucial for evaluating contextual aspects of reward and is involved in adaptive coding of reward prediction errors. The main questions in the constructions of DCMs were (i) where the driving input ‘reward’ enters the system, and (ii) the midbrain activity modulates connections in the network. Bayesian model selection (BMS) was used to investigate different variants of this three-region DCM. The set of alternative models, i.e. model space, is described below.

4.2.2 Time Series Extraction

A combination of anatomical and functional constraints were used to extract regional time series. For an anatomical definition of VTA/SN and NAcc, masks were applied (provided by Düzel et al. (2009)) and the Harvard-Oxford Subcortical Structural Atlas, respectively). For time series extraction from VTA/SN and NAcc, a 3 mm sphere was defined within the anatomical masks and around the peak voxel of each subject's 'positive prediction error' (for VTA/SN) and 'reward > punishment' contrast (for NAcc). For mPFC, we defined a 3 mm sphere around the peak voxel, which was limited to a search region of 6 mm distance from the group level 'reward > punishment' contrast maximum $[-3, 53, 1]$ and anatomically constrained by a mPFC mask created with the tool neurosynth (`wagerlab.colorado.edu`). All time series were adjusted by an F-contrast for effects of interest, thus removing confounds such as head movements (represented by a linear combination of realignment parameters).

4.2.3 Model Space

For defining the inputs used in the DCM analysis, a further GLM was constructed where rewarded and unrewarded trials were merged into the same regressor ('trial'), followed by two regressors of positive and negative prediction errors. Note that the a new GLM was required since we were interested in the region where the input enters the system whether it is a reward or a punishment. Based on the activations in the GLM analyses as well as the longstanding literature about reward neurocircuitry, a simple three-region model was formed. The key idea of this model is that activity in VTA/SN encodes trial-wise prediction errors (i.e., a bilinear modulation of VTA/SN self-connections with trial-wise prediction errors; encoded in the B matrix), and that the efferent connections of VTA/SN convey this prediction error signal to dopaminoceptive target regions (here: NAcc and mPFC), either as a direct via the endogenous connections of the model (A matrix) or in a nonlinear fashion (D matrix).

Two structural elements (A and B matrices) were identical for all models. Model assumptions were a fully connected model, i.e. bidirectional connections between SN/VTA, NAcc and mPFC, and that prediction errors modulated the self-connections of VTA/SN in a trial-by-trial fashion. Two other model components C and D matrices varied across models, resulting in a model space with a 2x4 factorial structure. First, the driving input 'trial' either enters the midbrain or drives all three regions (Fig. 4.1). Second, in each model midbrain activity modulated respectively: (a) none of the connections; (b) mPFC <==> NAcc reciprocal connections; (c) VTA/SN ==> mPFC and VTA/SN ==> NAcc; or (d) VTA/SN to both mPFC and NAcc self-connections (Fig. 4.2). In total, 8 alternative models per subject were estimated.

Figure 4.3 displays an example model selected from the model space to display connections and their strengths. The neural dynamics of the three regions $\dot{z}_1, \dot{z}_2, \dot{z}_3$, where z_1: VTA/SN activity, z_2: NAcc activity, and z_3: mPFC activity, can be described with this matrix equation:

$$\begin{bmatrix} \dot{z}_1 \\ \dot{z}_2 \\ \dot{z}_3 \end{bmatrix} = \left\{ \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix} + u_2 \begin{bmatrix} b_{11}^{(2)} & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} + u_3 \begin{bmatrix} b_{11}^{(3)} & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} + z_1 \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & d_{23}^{(1)} \\ 0 & d_{32}^{(1)} & 0 \end{bmatrix} \right\} \begin{bmatrix} z_1 \\ z_2 \\ z_3 \end{bmatrix} + \begin{bmatrix} c_{11} & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} u_1 \\ u_2 \\ u_3 \end{bmatrix} \tag{4.1}$$

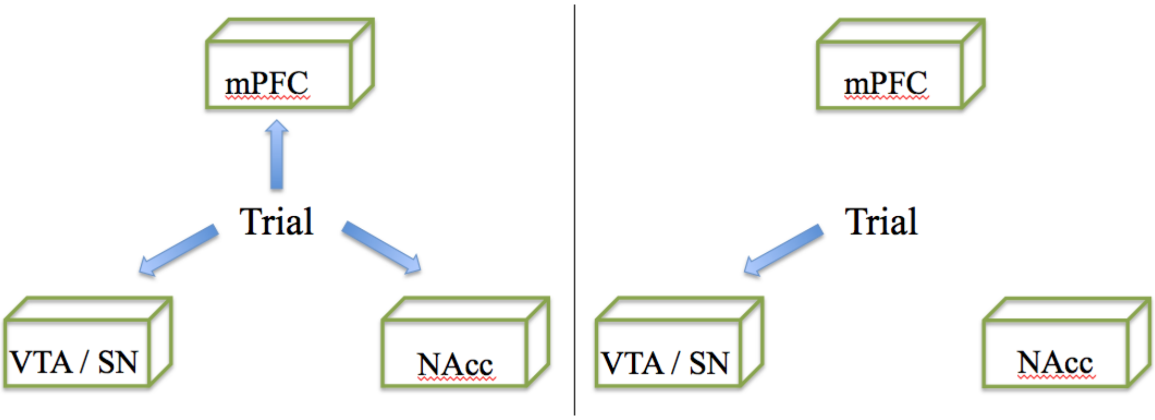


Figure 4.1: Driving input configurations. Driving input 'trial' (blue arrorw) enters the network at all three regions (left), or only at midbrain (right). This alternation would allow us to capture whether reward and punishment entered the network more than midbrain only, by comparing model evidence.

4.2.4 Bayesian Model Selection and Bayesian Model Averaging

Random effects BMS with Gibbs sampling was used for model comparison, which yields a posterior probability for each of the tested models (Stephan et al., 2009a) (see Fig. 4.4 for exceedance probability and posterior probability for each model). Models were grouped into 4 families based on the D matrix, i.e. how midbrain activity modulated different connections as explained above. Each family consisted of 2 nested models differing only by their driving input configuration (C matrix). Family wise model comparison showed that the models without any quadratic influence of the VTA/SN on other connections (family 'a', where $D = 0$) best described the data ($xp = 1$). The two models of the winning family were then merged using Bayesian model averaging, so that the parameter estimates of each model considered are weighted by the posterior probability of that model (Penny et al., 2010):

$$p(\theta_n | Y, m \in f_k) = \sum_{m \, \inf_k} q(\theta_n | y_n, m) p(m_n | Y) \tag{4.2}$$

where θ_n is the parameter vector for subject n, m is the model in the subset or family f_k. The $q(\theta_n | y_n, m)$ is the approximate distribution to the true posterior over parameters. This posterior term is a mixture of Gaussians, hence has a complicated form. Therefore, the marginal posterior for each parameter is obtained through sampling (for details see (Penny et al., 2010)). The resulting parameter estimates provided a basis for examining both the functional contribution of each connection to the network and genetic effects on connectivity strengths within the modeled reward circuit. Specifically, post-hoc t-tests were performed between carriers and non-carriers on the subject-wise parameter estimates provided by BMA (18): (i) A1+ vs. A1- group, (ii) *FTO+* vs. *FTO-* group, and (iii) the interaction of both. Finally, the association between the performance in avoidance learning with connectivity between midbrain and NAcc was tested by means of correlation analysis.

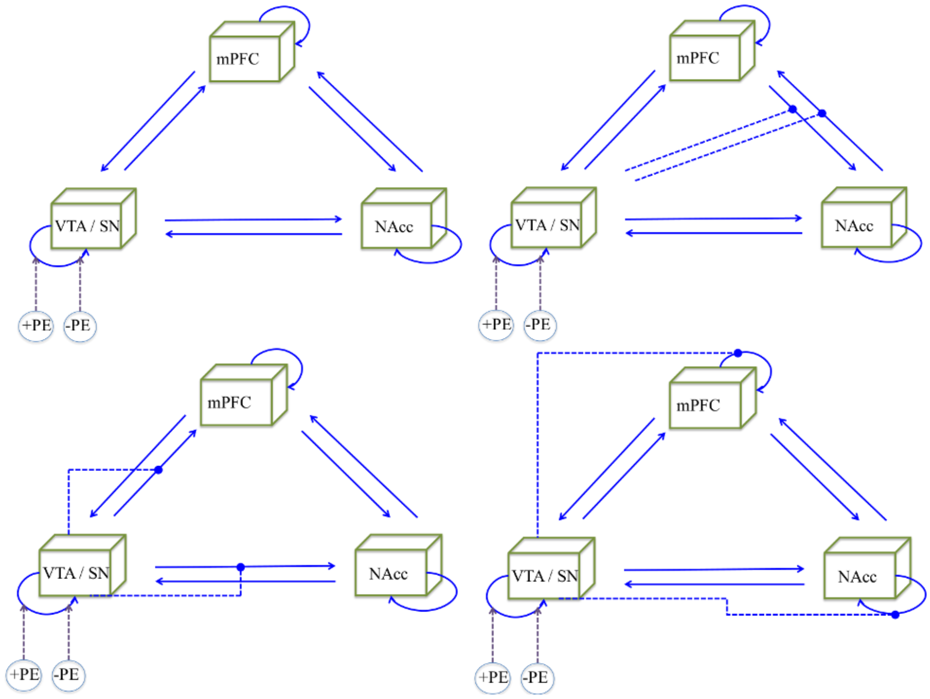


Figure 4.2: Nonlinear influence of midbrain activation. In addition to the bilinear model (top left), we have included nonlinear DCMs, with possible gating effects of VTA/SN, where the prediction error related activation in midbrain modulate connections between NAcc and mPFC (top right), connections from midbrain to NAcc and mPFC (bottom left) or the self-connections of NAcc and mPFC (bottom right). Each model also tested the potential inhibition / exhibition of self-connection of midbrain by positive and negative prediction errors. These 4 models formed the basis of model space partition such that in each family, the models differed only in terms of the driving input configuration, i.e. either midbrain or all 3 regions, hence, in total 2 x 4 = 8 models were tested per subject.

4.3 Results

4.3.1 Model Selection

Among the 8 models tested, there was not a certain winning model based on the Bayesian model selection with Gibbs sampling. Figure 4.4 shows the results of the model selection procedure. Model 1 and 5 have the highest exceedance probabilities, $\phi_1 = 0.26$ and $\phi_5 = 0.60$. After examining the models individually, the model space was partitioned into 4 families based on the D matrix. The family partitioning is displayed in Figure 4.5A (top). Each family differed in terms of the C matrix which identifies where the stimulus ‘trial’ entered the reward circuit, i.e. either VTA/SN or all three regions. Family wise model comparison procedure yield that the family with no non-linear modulations by VTA/SN ($d_{ij} = 0, \forall\, i, j$) had the highest frequency in the population (Fig. 4.5A, bottom). The member models of the winning family Model 1 and 5 are displayed in Fig. 4.5B. Following the family wise model comparison, parameter averaging was then applied for each of the connections from the 2

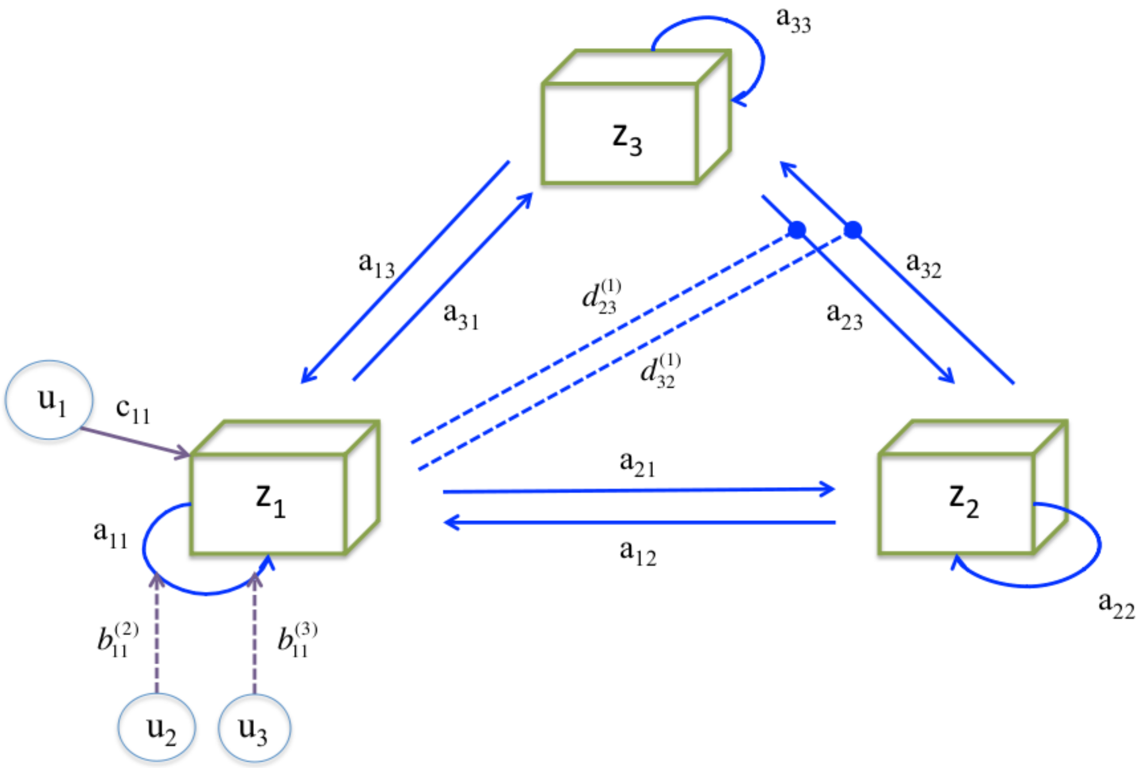


Figure 4.3: An example nonlinear model. This graphic represents all the connections of one nonlinear model where the midbrain activity (z_1) modulates connections a_{23} and a_{32}. The strength of connections are given by quadratic terms $d_{23}^{(1)}$ and $d_{32}^{(1)}$. It belongs to the family of models where the input (u_1) drives z_1. Intrinsic (or endogenous) connection strengths are given by a_{ij}. Finally, the self-connections a_{11} are modulated by inputs u_2 and u_3, through bilinear terms $b_{11}^{(2)}$ and $b_{11}^{(3)}$, respectively.

nested models, Model 1 and 5. Figure 4.6 shows the predicted time series (solid lines) for each of the brain regions of a subject by Model 5 which explained the observed responses (dotted lines) best for this subject.

4.3.2 Effective Connectivity Strengths in an RL Network

T-tests were performed on the averaged connections to understand the functional contribution of each connection to the reward network within the population. Table 4.1 shows that all endogeneous connections (a_{ij}) were significant except the connectivity from the midbrain to NAcc and from NAcc to mPFC. Furthermore, driving input 'trial' into VTA/SN region was also significant (p=0.002).

4.3.3 Genotype, Connectivity and Behavior

Possible modulations of *FTO* and *ANKK1* variants on the effective connectivity between reward-responsive mesolimbic and mesocortical regions were investigated. T-test between

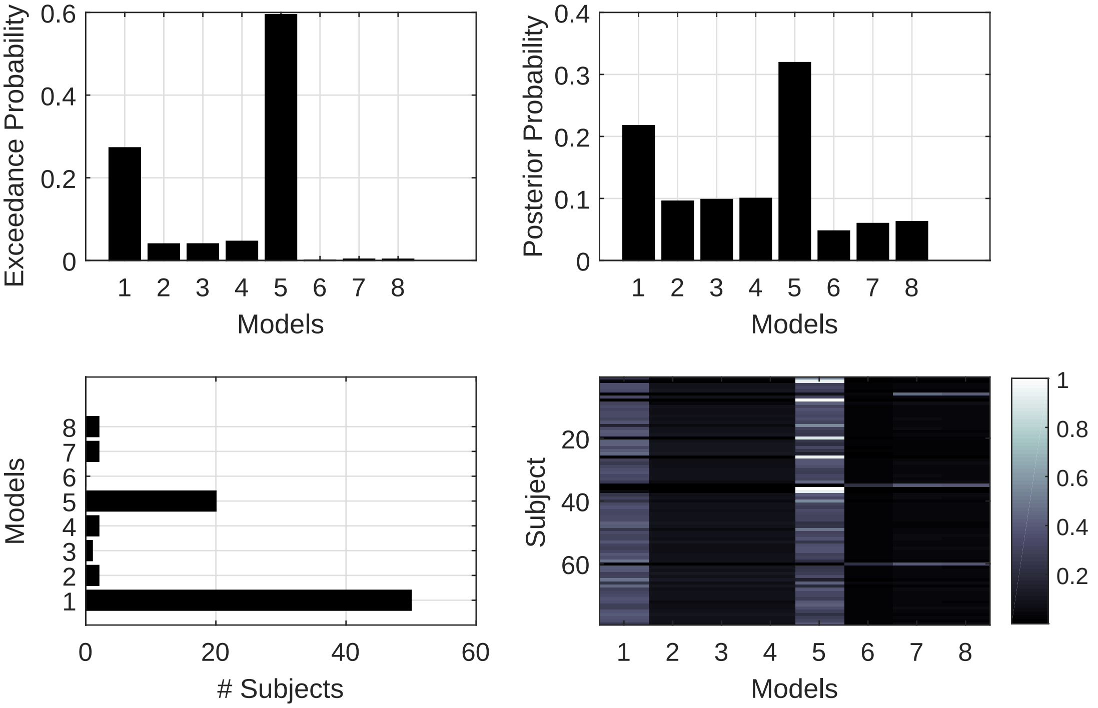


Figure 4.4: Bayesian Model Comparison. Model 1 and Model 5 had the highest evidence as demonstrated by exceedance probability ($\phi_1 = 0.26$ and $\phi_5 = 0.60$, top left), and posterior probability (0.22 and 0.32, top right). 63% of subjects favored the Model 1, and 25% of subjects favored Model 5 (bottom left), Model attribution shows posterior probability per subject and per model.

carriers and non-carriers on the Bayesian averaged parameters suggested a modulatory effect of *FTO* on the connectivity from VTA/SN to NAcc (p = 0.055; Fig. 4.7b, left), and from NAcc to mPFC (p = 0.017; Fig. 4.7b, middle). Fig 4.7a displays the significant differences with solid lines for *FTO* effect (left), and *ANKK1* effect (right). Strikingly an association of effective connectivity and behavior was observed: Increased connection strengths between VTA/SN and NAcc were associated with poorer ability to avoid negative outcomes (Fig. 4.8).

4.4 Discussion

In this chapter, we showed that DCM is a convenient state-space model for testing directional influences among the regions of a reward network. The prediction errors that were derived from reinforcement learning model were incorporated in the connectivity analysis to construct a plausible biophysical model. Since construction of alternative models and the regions that are included in each model are based on prior theoretical knowledge, this study proposed that using the information from computational models of learning provides a more elaborate DCM analysis for testing competing hypothesis.

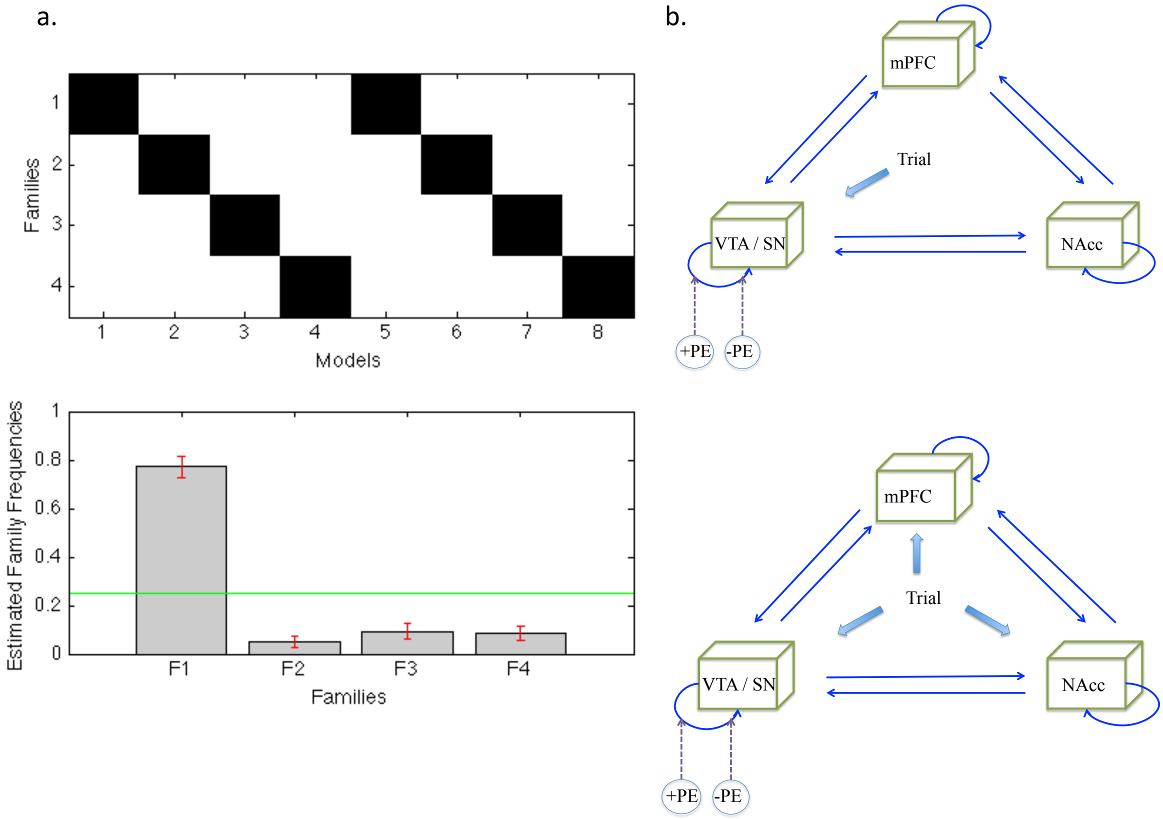


Figure 4.5: Family-wise Model Comparison. A) Family partition of model space (top), and estimated family frequency (bottom). The winning family is the first one (F1) with estimated family frequency of 0.77. B) The winning family F1 includes nested models Model 1 (top) and 5 (bottom).

While bilinear DCMs allowed for inferring the hidden parameters of the reward network, nonlinear DCMs provided testing for the gating hypothesis of midbrain activations via the second order derivative term. This method has shown to be useful in providing insights in clinical applications such as altered gating mechanisms in schizophrenia (Dauvermann et al., 2013). There are other forms of DCMs which can be applied for mechanistic modeling interactions. For example, two-state DCMs (Marreiros et al., 2008) can be useful for modeling inhibitory and excitatory populations in midbrain. However, this would require a better spatial resolution to identify these subgroups in such small structures.

Since DCM is not an explorative technique, defining of model search space is crucial. Here the analyses were based on a three-region model. Although there are other brain regions involved in during reinforcement learning, DCM has been validated and shown to be sensitive to group effects when it is applied to simpler models (Schuyler et al., 2010). Moreover, both functional and anatomical constraints were used for increasing the specificity of the time series extraction. One can also propose that having a large number of subjects (79) increases the reliability of the results presented in this work.

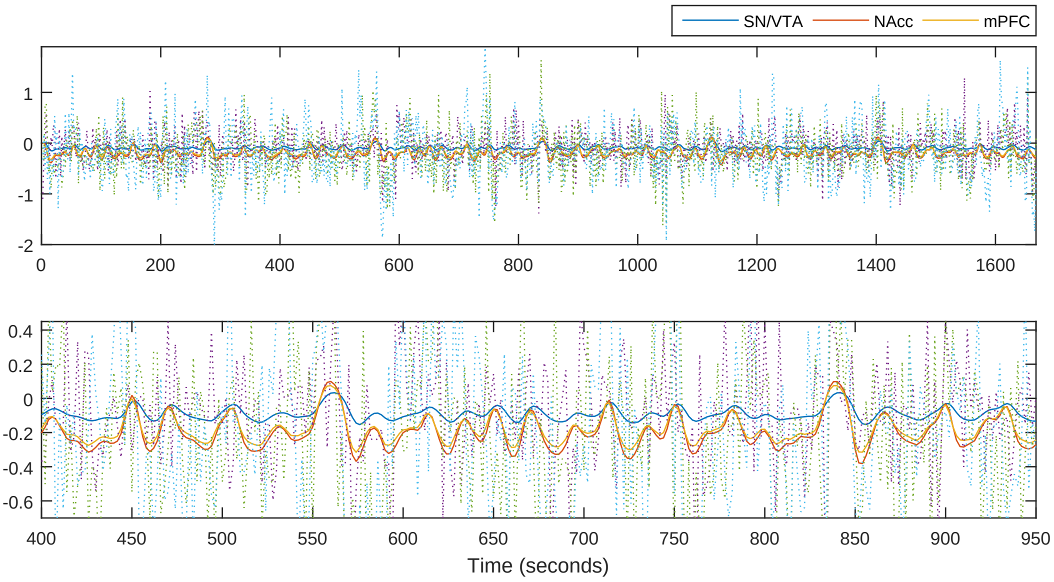


Figure 4.6: Observed and fitted time series for subject 3251. Explained variance is 14%. Top plot shows the observed responses (dotted lines) and predicted time series by DCM (solid lines) for each of the three regions. Bottom plot zooms in a time interval (400-950 s) for a detailed visualization.

In this study, nonlinear and bilinear models were compared. These models have different likelihood functions and priors, i.e. the elements of D matrix that have non-zero prior covariance for modulatory effect of midbrain activity in the nonlinear models. Variational Bayesian approximation to the log model evidence is an effective approach in comparing different type of DCMs as it accounts for both the posterior covariance among the parameters and the effective degrees of freedom (Stephan et al., 2008). Also, the family wise model comparison allowed us partitioning the model space with respect to the modulation of the network by midbrain activity, where each family comprised nested models. This approach is particularly useful when the number of models in each family are large (Penny et al., 2010).

It is accepted that the explained variance should not be below 10% for the time series predicted by DCM, and the estimated connection strengths should not be below 1/8 Hz. These cases indicate a convergence problem (see `spm_dcm_fmri_check.m` in spm package for details). While the connectivity strengths were larger than 1/8 Hz, the explained variance that is observed in this study (see Fig. 4.6) was low. This is a sign of low signal-to-noise ratio, which is often observed in cognitive experiments and event-related designs in comparison to sensory experiments and block designs. These points should be taken into consideration in the future studies of causal models that are based on learning paradigms.

DCMs with nonlinear interaction terms were included to test the gating role of midbrain activation on the connections to the projection sites. Family wise model comparison revealed that the model family without any nonlinear effect has explained the interactions best. In addition, there was not a clear winner within each family suggesting that the input trials entered the network at more than one region. Random effects analysis of the Bayesian averaged MAP estimates of connection strengths from the winning model family on the whole

Table 4.1: Random effects analysis on averaged parameters

	mean	*stdev*	*p*
Endogeneous Connections			
VTA/SN => VTA/SN	-0.492	0.029	2.2 x 10^{-6}
VTA/SN => NAcc	0.062	0.594	0.360
VTA/SN => mPFC	-0124	0.528	0.040
NAcc => NAcc	-0.493	0.036	2.2 x 10^{-6}
NAcc => VTA/SN	0.066	0.190	0.003
NAcc => mPFC	-0.016	0.269	0.600
mPFC => mPFC	-0.492	0.066	2.2 x 10^{-6}
mPFC => NAcc	0.093	0.178	1.3 x 10^{-5}
mPFC => VTA/SN	0.055	0.098	3.7 x 10^{-6}
Modulation by PE			
$+PE$: VTA/SN => VTA/SN	-0.004	0.361	0.92
$-PE$: VTA/SN => VTA/SN	0.002	0.119	0.87
Driving Input (Trial)			
Trial => VTA/SN	0.154	0.430	0.002
Trial => NAcc	0.038	0.379	0.379
Trial => mPFC	-0.217	0.368	1.304 x 10^{-6}

sample indicated that feedback information, i.e. the driving input, entered the network at mPFC and VTA/SN. However, *FTO* allele carriers vs non-carriers differed in terms of the driving input to NAcc that was found by post hoc t-tests. Moreover, *FTO* allele altered the connectivity from VTA/SN to NAcc as well as from NAcc to mPFC, whereas a significant influence of *ANKK1* allele onto the connection from VTA/SN to mPFC was observed. Note that directions are not trivial given the known dopaminergic projection sites. These findings show that the non-significant parameter estimates (i.e., trial -> NAcc; VTA/SN -> mPFC; NAcc -> mPFC; VTA/SN -> NAcc) was due to the opposite effects of the carriers and non-carriers of both alleles, which were both equally distributed in this sample.

It needs to be emphasized, however, that connectivity results are reported at uncorrected levels and should thus be considered with some caution. Generally, correcting connectivity estimates based on any generative model, such as DCM, for multiple comparisons is a non-trivial issue because of the posterior dependencies of model parameters that are ubiquitously encountered in biological systems (Gutenkunst et al., 2007). These dependencies make conventional correction methods, such as Bonferroni correction, very conservative (Stephan et al., 2010).

In conclusion, we showed that reinforcement learning models can be combined with DCM analysis for an effective model space construction. Also the differences in the neuronal network during reinforcement learning can be revealed by DCM analysis. Therefore, combining computational models of learning and dynamic models of brain networks can together capture connections that are affected by differences in DA level due to genetic polymorphisms and the associated learning performance.

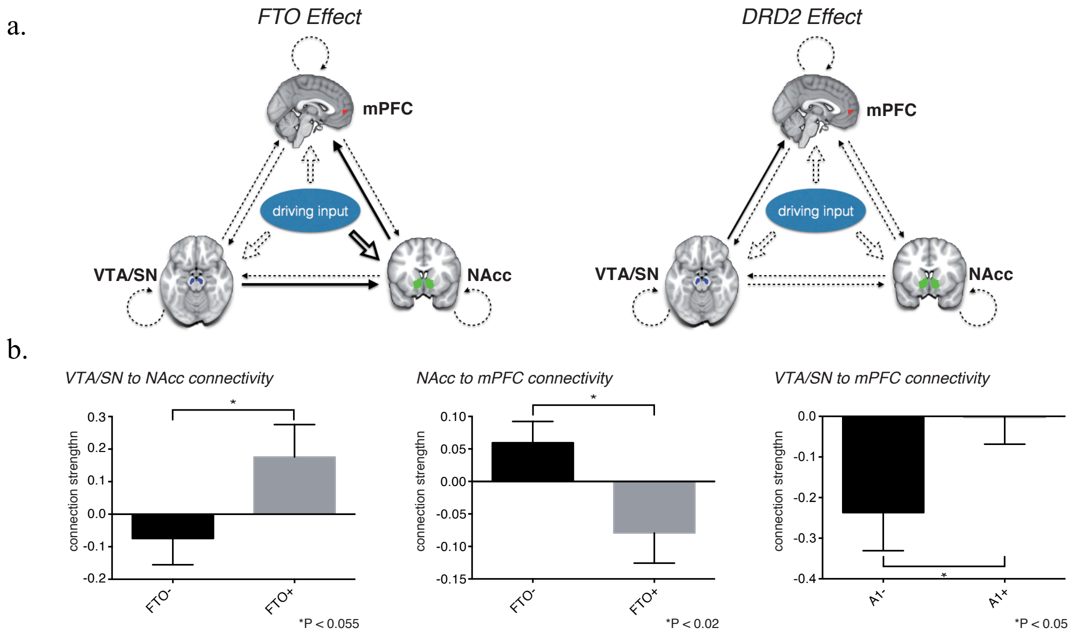


Figure 4.7: Influence of genotype on the brain RL network. A) Basic layout of a DCM for investigating modulation of reward-responsive regions by the *FTO* gene variant and the *ANKK1* gene variant. Solid connections indicate connections, which are significantly ($p < 0.05$) altered by genetic status. Dotted connections do not show a significant genetic effect. B) Average strengths of the connection from VTA/SN to NAcc (left) and from NAcc to mPFC (middle) under both *FTO* gene variants, as well as from VTA/SN to mPFC (right) relative to groups defined by *ANKK1* genotype. Values are mean ± SEM.

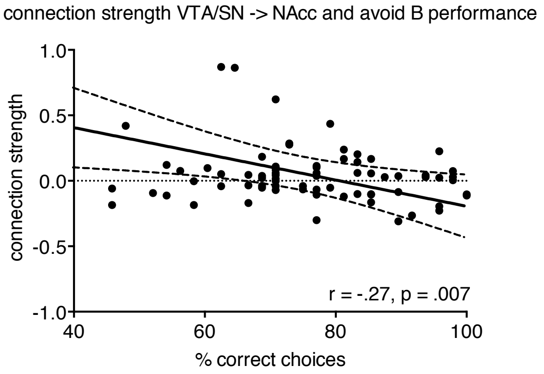


Figure 4.8: Correlation between connection strength from VTA/SN to NAcc and performance of avoidance learning. Dashed lines indicate the 95% CI for the linear regression.

5 Hierarchical Bayesian Modeling to Study Differences in Learning

5.1 Introduction

When a learning agent is faced with a volatile environment, it requires more complex algorithms than basic reinforcement learning to model its behavior. In an environment where the contingencies change, it is necessary to consider many features of available sources of information. Hierarchical Bayesian models provide a framework to integrate volatility in the form of a probabilistic dependency graph. These normative methods assume that humans are optimal Bayesian learners. However, there are a few limitations to the assumption that people learn optimally such as computational complexity and biological implausibility. It is not surprising that there are considerable inter individual differences in the way people learn from new information. This becomes particularly important to applications of computational models in psychiatry, where the individual differences are vital for understanding altered perceptual and cognitive mechanisms. Also, when complex integrals are considered these algorithms can be computationally expensive for large number of parameters. Bayesian approximate algorithms such as Hierarchical Gaussian Filter (HGF) (Mathys et al., 2011) can address these limitations. It is suitable to model learning and decision making in more complex, unstable (and therefore more realistic) environments because it can be adapted with cue combinations and volatility manipulations. This study will build on the modeling approach presented in Diaconescu et al. (2014), by extending the HGF algorithm in a parallel manner for learning the features of different cues.

Cue combination studies indicate that individual differences in the perception influence the weighting of the precision of multiple cues, during the formation of a combined belief (please see the background chapter for the relevant literature). In this chapter, a precision weighting response model will be applied to quantify any bias towards one cue over another. This response model further takes into account the volatility of the environment such that the temperature parameter is a function of subjective volatility estimates rather than being fixed. Also, the relationship of the environment with the agent, who has a dynamic learning rate and a dynamic temperature parameter, will be examined during learning from different cues.

Although using Bayesian modeling in combining beliefs about different cues is not a novel idea, HGF algorithm has never been implemented in a parallel learning system before, i.e. when the features of more than one cue need to be learned at the same time. This requires the importance of an evaluation of a rich parameter space. In this study, possible parameter identifiability issues will be addressed using simulations and model inversion diagnostics.

The proposed model will then be tested on a real dataset. Its evidence will be compared to two other models in order to validate that subjects are learning both cues and that they differ in the way they learn. The latter comparison is performed via the fixation of prior variance of model parameters, which corresponds to the assumption that all subjects are Bayes optimal learners. It is particularly important to evaluate alternative models before making inferences on the parameters of a certain model in a psychiatric context where the differences are considered as the source of individual perceptual processes.

Computational modeling provides a mechanistic approach which helps to bridge the gap between altered behavior and brain responses. Importantly, recent progress in computational modeling has convincingly demonstrated that Bayesian models can be used to formally investigate perceptual and cognitive mechanisms that underlie social behavior when explicit social advice is provided to study participants (Diaconescu et al., 2014). In particular, it has been shown that humans employ hierarchical generative models to make inferences about the changing intentions of others when attention is explicitly directed towards them and that they integrate estimates of advice accuracy (i.e. the correctness of the advice, which can be valid or misleading depending on the conflicting interests of the players) with non-social sources of information when making decisions. In Bayesian terms, this integration corresponds to an optimal weighting of social and non-social cues in terms of their relative precision.

Learning from social cues is an important component to understand many psychiatric conditions. For example, individuals with autism suffer from striking impairments in everyday life social situations, which begs the question which and how processes other than basic perceptual mechanisms may come into play (Hamilton, 2013). Currently, a prominent theoretical proposition suggests that the autistic spectrum might be specifically characterized by deficits of predictive coding or Bayesian inference (Pellicano and Burr, 2012; Sinha et al., 2014). Predictive coding formulations of perception propose that expectations in higher brain areas generate top-down predictions that meet bottom-up stimulus-related signals from lower sensory areas. The discrepancy between actual sensory input and predictions of that input is described as a prediction error. With regard to autism, it has been proposed that autistic traits might be related to higher sensory precision, i.e. a stronger reliance on (bottom-up) sensory evidence as compared to (top-down) prior beliefs, which can lead to a failure of automatically contextualizing sensory information in an optimal and socially adequate fashion (Friston et al., 2013; Lawson et al., 2014). From a predictive coding perspective, there are two possible pathologies. First, there could be deficits in predicting and inferring the mental states of others or, alternatively, these inference or representations are unable to influence behavior because they are afforded an impoverished weight or precision.

This study will apply hierarchical Bayesian modeling to behavioral data (Fig. 5.1) from a novel version of a probabilistic reward learning paradigm. The hidden variables of the model will be used to predict individual autistic traits. The paradigm included a social gaze cue about whose relevance no explicit information was provided in order to investigate autistic trait-related differences in the extent to which healthy individuals integrate and use this piece of social information during task performance. In light of the evidence discussed above, we hypothesized that autistic traits are related to differences in the extent to which individuals are influenced by social cues (i.e., their precision), rather than a general inability to process social cues and putatively underlying mental states. On the behavioral level, this should result in higher total task scores for individuals lower in autistic traits as they should be more easily able to exploit the additional social information. In terms of the underlying cognitive processes, we hypothesized that this behavioral advantage might be subserved by differences in the effect that social information can have on decision-making, which, in turn, would be inversely related to autistic traits. We further predicted that using the social cue should be more difficult under volatile conditions and differentially so for individuals with higher autistic traits. The experiment used in this study allows to extend the precision vs prediction hypotheses with social cue integration.

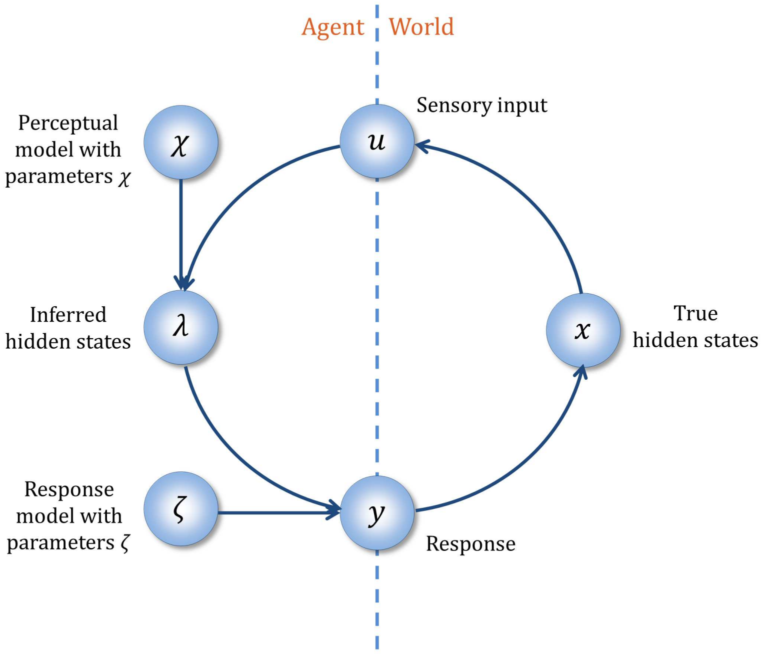


Figure 5.1: Model Framework. The agent is interacting with its environment through the sensory inputs that are a result of the true hidden states of the World. These true states x, can be inferred by the agent via a perceptual model. Finally, inferred states λ are formed into decisions or responses with a response model. Both the perceptual and response model has agent specific parameters χ and ζ, respectively (from `http://www.translationalneuromodeling.org/tapas/`).

5.2 Methods

5.2.1 Parallel Learning Systems

The "observing the observer" (OTO) approach provides a complete mapping from experimental stimuli to observed responses by making inversion of the perceptual model, $m^{(p)}$ and the response model $m^{(r)}$ (Daunizeau et al., 2010b): $u \to \lambda \to y$ (Fig. 5.1). An extension of this approach is a generative model called Hierarchical Gaussian Filter (HGF) which accounts for deterministic and probabilistic relationships between the environment and perceptual states (Mathys et al., 2011) (see Background chapter for details). In this chapter the HGF will be implemented to model the behavior of an agent who is learning the hidden states of two

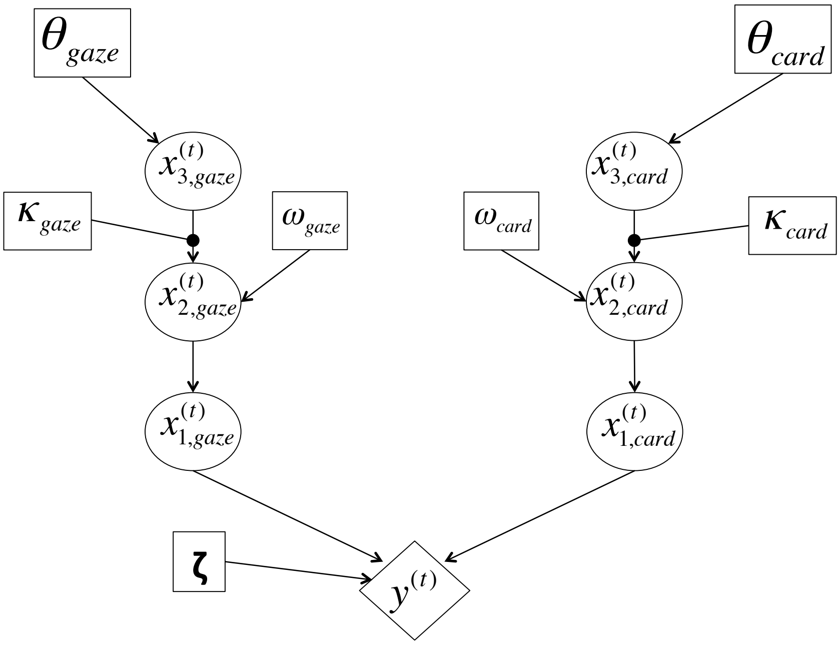


Figure 5.2: Graphical depiction of two parallel learning systems that were assumed to influence the choice behavior. For any trial t, $x_3^{(t)}$ follows a Gaussian random walk such that $p(x_3^{(t)}) \sim \mathcal{N}(x_3^{(t-1)}, \theta)$. The first level state variable $x_1^{(t)}$ is the accuracy at that trial and is a sigmoid transform of the second level state variable $x_2^{(t)}$ that also follows a Gaussian distribution: $\mathcal{N}(x_2^{(t-1)}, e^{(\kappa x_3^{(t)}+\omega)})$, where the variance term includes 2 parameters: ω is the fixed component of the step size variance, and κ accounts for the coupling between the third and the second level. The response model parameter ζ represents the weight on the precision of the inferred gaze accuracy.

sources of information. The learning environment consists of a social and a non-social cue. Congruency of response with advice, i.e. the advice given by the social cue (the gaze), was modeled using HGF combined with a response model as implemented in Diaconescu et al. (2014). This approach allows the estimation of two hierarchically coupled hidden states that describe subjects' learning about the environmental statistics, namely the probability and the volatility of the card and gaze cues, based on their responses.

The structure of the perceptual model is depicted in Fig. 5.2 as a graphical model. The state of the world is presented with three levels that evolve as Gaussian random walks where the step-size (or variance) is determined by a subject-specific parameter of the level above: The value of the state at each trial is dependent on the trial before in a Markovian fashion. The third level x_3 represents the learning about the volatility of the two stimulus categories with a variance determined by a fixed parameter θ.

$$p(x_{3,i}^{(t)}) \sim N(x_{3,i}^{(t-1)}, \theta_i) \tag{5.1}$$

where i denotes the cue type (gaze or card). The second level x_2 represents the tendency of the two stimulus categories. Its evolution in time is determined by the volatility (or x_3) and two learning parameters: κ describes the coupling between the two levels and ω represents the log-volatility of x_2 or the tonic part of the learning rate.

$$p(x_{2,i}^{(t)}) \sim N(x_{2,i}^{(t-1)}, e^{\kappa_i x_{3,i}^{(t)}+\omega_i}) \tag{5.2}$$

The first level state variable x_1 is a sigmoid transformation of x_2, and represents the probability of each stimulus category with a Bernoulli distribution.

$$p(x_{1,i}^{(t)}) \sim Bernoulli(x_{1,i}^{(t)}; s(x_{2,i}^{(t-1)})) \tag{5.3}$$

At each trial t, x_1 is binary: $x_1 \in \{0,1\}$ Two parallel perceptual models are implemented for gaze and card separately since the subject should implement one learning system for each of these sources of information (Fig. 5.2). The three level hierarchical perceptual model describes the state changes with time. For the gaze cue, the first level $x_{1,gaze}^{(t)}$ is the accuracy of the current advice, i.e. whether the gaze is directed towards the correct card or not on trial t. The second level $x_{2,gaze}^{(t)}$ is the current tendency of the gaze to give accurate advice. The third level $x_{3,gaze}^{(t)}$ is the current volatility of the second level, i.e. the change in the intentions of the gaze. For the card cues, first level $x_{1,card}^{(t)}$, describes the accuracy of the green card, i.e. whether the green card is correct or not. The second level is the tendency of green card being correct $x_{2,card}^{(t)}$, and the third level is the volatility of the tendency $x_{3,card}^{(t)}$, i.e. the change in the tendency. The variational approximation to the posterior distribution of the state variables estimate them in terms of their sufficient statistics $\{\mu_{k,i}, \sigma_{k,i}\}$ for each level $k \in \{1,2,3\}$ and cue type $i \in \{gaze, card\}$. The first level subjective beliefs $\mu_{1,i}$ are weighted by their precision $\pi_{1,i}$ to form the basis of a response model (of the observed behavior) as explained in detail below.

5.2.2 Precision Weighted Response Model

The HGF was applied to derive subject-specific accuracy and volatility estimates for card and gaze cues in a parallel manner. On a given trial t, subjects generated a combined belief $b^{(t)}$ after weighting the posterior expectation of inferred card and gaze accuracies, $\mu_{1,card}^{(t)}$ and $\mu_{1,gaze}^{(t)}$ to generate actions in the following manner:

$$w_{gaze}^{(t)} = \frac{\zeta \pi_{1,gaze}^{(t)}}{\zeta \pi_{1,gaze}^{(t)} + \pi_{1,card}^{(t)}} \qquad w_{card}^{(t)} = \frac{\pi_{1,card}^{(t)}}{\zeta \pi_{1,gaze}^{(t)} + \pi_{1,card}^{(t)}} \tag{5.4}$$

$$b^{(t)} = w_{gaze}^{(t)} \mu_{1,gaze}^{(t)} + w_{card}^{(t)} \mu_{1,card}^{(t)} \tag{5.5}$$

where w_{gaze} and w_{card} are effective precisions of gaze and card cues, ζ is the weight on the precision of inferred gaze accuracy or the additional bias towards the social cue; $\pi_{1,gaze}^{(t)}$ and $\pi_{1,card}^{(t)}$ are precisions (inverse variances) at the first level for gaze and card accuracies, respectively. Since the first level estimates are assumed to follow a Bernoulli distribution, one can calculate the precision at each trial by

$$\pi_{1,gaze}^{(t)} = \frac{1}{\mu_{1,gaze}^{(t)}(1-\mu_{1,gaze}^{(t)})} \qquad \pi_{1,card}^{(t)} = \frac{1}{\mu_{1,card}^{(t)}(1-\mu_{1,card}^{(t)})} \tag{5.6}$$

The probability of the choice behavior was assumed to be a unit square sigmoid function:

$$p(y^{(t)} = 1 | b^{(t)}) = \frac{(b^{(t)})^{(\beta)}}{(b^{(t)})^{(\beta)} + (1 - b^{(t)})^{(\beta)}} \tag{5.7}$$

where β is a function of the third level volatility estimate or μ_3:

$$\beta = exp(-\mu_{3,gaze}^{(t)}) + exp(-\mu_{1,card}^{(t)}) \tag{5.8}$$

5.2.3 Participants

In light of the evidence which suggests that autistic traits are distributed on a continuum across the general population and are known to show identical etiology across the diagnostic divide (Robinson et al., 2011), healthy participants were chosen based on their score on the German translation of Autism-Spectrum Quotient (AQ) questionnaire (Baron-Cohen et al., 2001). This experimental approach of studying autistic traits in neurotypicals makes it possible to infer about the etiology of autistic traits without potential confounds from a variety of co-morbid conditions often noted in patients with autistic spectrum disorders. In order to capture the extremes of the distribution and have a balanced proportion of participants with high and low AQ scores, 36 subjects were pre-screened and invited based on their AQ scores up to 25 (19 males; aged 20 to 37 years; mean age = 26.25 years). It has been shown that AQ has a good discriminative validity at a threshold of 26 (Woodbury-Smith et al., 2005). Participants did not have any history of neurological and psychiatric disorders and were invited by using preexisting database of the MPI for Metabolic Research comprising healthy native German volunteers. The distribution of AQ scores were as follows: range = 7 - 23, mean = 15.72, SD = 5.09. All participants gave informed consent before the beginning of the experiment. A description of subjects' traits is shown in 5.1.

Table 5.1: Descriptive (mean ± SEM) data of participants, gender, age, AQ, systemizing quotient (SQ), empathy quotient (EQ), IQ (verbal).

		Gender		**Age /years**		**AQ**		**SQ**		**EQ**		**IQ (Verbal)**	
AQ group	*# Subjects*	*m*	*f*	*Mean*	*SEM*	*Mean*	*SEM*	*Mean*	*SEM*	*Mean*	*SEM*	*Mean*	*SEM*
High AQ	18	9	9	25.5	0.7	20.4	0.5	27.1	2	41.1	2.1	101.9	2.1
Low AQ	18	10	8	27	1	11.1	0.4	23.9	2.1	44.3	2.6	103.2	2.7

5.2.4 The Experiment

The card game used in this study, which had been originally designed as two cards with associated winning probabilities (Behrens et al., 2007), was combined with a face cue presented in the center of the screen (Fig. 5.3A). The eye gaze direction of the face was manipulated to change during each trial and to then be directed towards one of the cards, before participants were allowed to make their choice. As a result, there were two things that need to be learned in the task. First, whether the reward is associated with the green card or the blue card.

Second, whether the gaze shift is directed towards the card that is rewarded. The probability of whether or not the face actually looked towards the winning card on a given trial (i.e. gaze accuracy) was systematically manipulated in accordance with a probabilistic schedule as well. Both the card and gaze accuracies were, thus, varied independently of one another across the experiment (Fig. 5.3B-C). The phases in which the trials have cues with unstable accuracy are referred as volatile phases. In the first half of the experiment (trials 1 to 60), card accuracy was stable and high, whereas in the second half (trials 60 to 120) it followed a volatile phase. For the gaze accuracy the volatile phase took place during trials 30 to 70. The probabilistic schedule for the gaze accuracy was reversed for the half of the subjects to avoid block order effects. The positions of the cards (left or right) were determined randomly.

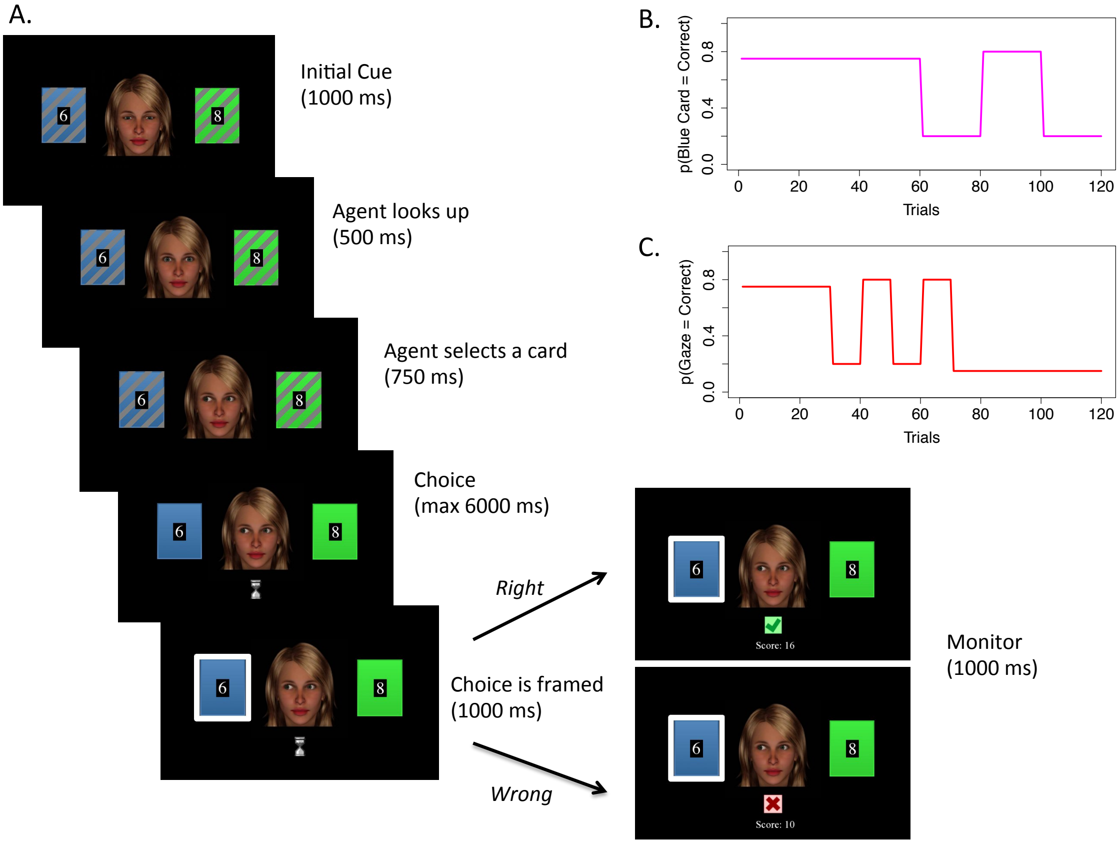


Figure 5.3: The experimental design. (A) Subjects can make a choice once the lines on both cards disappeared. If the choice is right on that trial a green tick is displayed and the reward value of the right card is added to the total score. If the choice is wrong a red cross is displayed and the score remains the same. Probabilistic schedules: (B) Probability of the blue card being correct (i.e., card accuracy), and (C) probability of the gaze showing the correct card (i.e., gaze accuracy).

In the instructions subjects were informed about the fact that the cards have winning probabilities, which can change during the experiment and which are independent of the reward magnitude that is displayed on them. They were instructed that they could earn an extra amount of money depending on their performance in the game. In the paradigm, only one of the two cards was correct on each trial. If participants chose the correct card. the reward value of the respective card was added to the total score. If they chose the wrong card, the score remained the same. The reward values were random (numbers between 1 to 9) to avoid subjects from associating them with the winning probabilities of the cards. Finally, participants were informed about the presence of a face on the screen, which was explained by stating that it was supposed to make the visual display more interesting. Participants did not receive any other information about the face in an attempt to keep the instruction about the gaze cue as implicit as possible. After the experiment, subjects filled out a brief questionnaire, which included ratings on a scale from 0-100 about the perceived difficulty of the task, about whether / how helpful participants' found presentation of the face as well as answering some yes / no-questions such as whether they could make out a rule about the change in the winning probabilities of the cards. This questionnaire is relevant, because it allows assessing to what extent participants were actually aware of the gaze cue being informative.

5.2.5 Simulations

In order to assess the model parameters several simulations were performed. First, the probability of taking the advice was simulated for agents with a different gaze precision weighting parameter, i.e. ζ, but with the same perceptual model parameters, $\kappa = 1, \theta = 0.25, \omega = -4$. Perceptual model parameter values were informed from previous studies using the HGF (Diaconescu et al., 2014), (Iglesias et al., 2013). The ζ took the values $\{0, 0.5, 1, ...3\}$ to cover a range of agents with low and high weightings. Second, to address parameter identifiability issues, four sets of simulations were run for each parameter. Decisions were simulated by changing one parameter at a time. The ζ took the values $\{0, 1, 2, 3\}$, ω took the values $\{-9, -6, -3\}$, κ took the values $\{0.2, 0.7, 1.2, 1.7\}$, and finally θ took the values $\{0.05, 0.15, 0.25, 0.35\}$. The decisions were simulated 200 times for each parameter setting and for the experimental input that was used in this study. Third, we examined dynamic learning rates, which are used in the update equations at the second and third level, of the simulated agents. Derivation of the update equations are given in Mathys et al. (2011), and we extended them for each cue i:

$$\mu_{2,i}^{(t)} = \mu_{2,i}^{(t-1)} + \frac{1}{\pi_{2,i}^{(t)}} \delta_{1,i}^{(t)} \tag{5.9}$$

$$\mu_{3,i}^{(t)} = \mu_{3,i}^{(t-1)} + \frac{1}{\pi_{3,i}^{(t)}} \frac{\kappa_i}{2} \frac{e^{\kappa_i \mu_{3,i}^{(t-1)} + \omega_i}}{\frac{1}{\pi_{2,i}^{(t-1)}} + e^{\kappa_i \mu_{3,i}^{(t-1)} + \omega_i}} \delta_{2,i}^{(t)} \tag{5.10}$$

Two learning rates can then be defined:

$$\alpha_i^{(1)} \equiv \sigma_{2,i}^{(t)} \tag{5.11}$$

$$\alpha_i^{(2)} \equiv \frac{1}{\pi_{3,i}^{(t)}} \frac{\kappa_i}{2} \frac{e^{\kappa_i \mu_{3,i}^{(t-1)} + \omega_i}}{\frac{1}{\pi_{2,i}^{(t-1)}} + e^{\kappa_i \mu_{3,i}^{(t-1)} + \omega_i}} \tag{5.12}$$

5.2.6 Model Inversion

Maximum a posteriori estimates of the parameters are obtained using an approximate variational Bayesian scheme. The update equations take the form of precision-weighted prediction errors following a form similar to an extended Kalman Filter and are therefore analytically tractable. Beliefs at every level in the hierarchy are updated with a step size equivalent to the prediction error times a ratio of precisions (precision of the data in the numerator and precision of the prediction in the denominator, Eq. 5.10). The HGF can be downloaded as a part of the software collection tapas (http://www.translationalneuromodeling.org/tapas/). For the details of the update equations and the variational Bayesian inversion scheme see Daunizeau et al. (2010b) and Mathys et al. (2011). Note that, we duplicated the update equations in the code since we have two sources of information, i.e. the card and the gaze. All the parameters ($\theta_i, \kappa_i, \omega_i, \zeta$) and the state variables ($x_{i,k}$) were estimated in terms of their sufficient statistics for each subject by using a quasi-Newton optimization algorithm (Newton Broyden-Fletcher-Goldfarb-Shanno method; Nocedal and Wright, 2006) as implemented in HGF version 3 running on MATLAB 7.12 (The MathWorks, Inc., Natick, MA).

5.2.7 Model Space and Model Selection

The priors of the model parameters are given in Table 5.2. While ω is estimated in its native space, θ and κ have lower bound zero and they are estimated in *logit* transformed space ($logit_a(x) = ln(\frac{x}{(a-x)})$) to constrain them with an upper bound a:

$$logit_{a_{\theta_i}}(\theta_i) \sim N(\mu_{logit_{a_{\theta_i}}}, \sigma_{logit_{a_{\theta_i}}}) \tag{5.13}$$

$$logit_{a_{\kappa_i}}(\kappa_i) \sim N(\mu_{logit_{a_{\kappa_i}}}, \sigma_{logit_{a_{\kappa_i}}}) \tag{5.14}$$

where $a_{\theta_i} = 0.5$ and $a_{\kappa_i} = 2$. Both upper bounds are selected based on the assumptions under the derivation of update equations (see the configuration file `tapas_hgf_binary_3l_config.m` for details).

In addition to the original model that was explained above (Model 1), an alternative model (Model 2) was used, which proposes that participants ignored the accuracy of the gaze completely and based their predictions only on the card accuracy. In Bayesian terms, the prior mean and variance were set to different values to obtain a second version of belief-to-response mapping whereas the perceptual model remained the same in both models. Prior mean and variance for ζ in log space for Model 1 was ($log(0.5), 16$), and for Model 2 it was $(-\infty, 0)$. ζ is transformed to its native space in the response model calculations (Eq. 5.4). The model space was augmented with a third model that consists of fixed perceptual parameters (κ, θ, ω). This model is a normative one, assuming that all agents behave in a Bayes optimal way. When a perceptual model is paired with a response model that uses volatility-dependent mapping of beliefs to decisions (Eq. 5.7 and 5.8) all three learning parameters need to be estimated. Therefore, for the third model, a decision noise response model was used, which includes inverse decision temperature parameter β (Eq. 5.7) that is now a subject specific, free parameter and do not depend on volatility. The priors for all three models are shown in Table 5.2.

After inverting each model for each subject, Bayesian model selection (BMS) was used. It is a well-established, powerful technique for model comparison. It computes a conditional

Table 5.2: Priors on model parameters that varied across the models (HGF with volatility (Model 1), HGF reduced card with volatility (Model 2), normative HGF with decision noise (Model 3).

	Model 1		**Model 2**		**Model 3**	
Parameters	Prior Mean	Prior Variance	Prior Mean	Prior Variance	Prior Mean	Prior Variance
$\theta_{gaze}, \theta_{card}$	0.25	16	0.25	16	0.25	0
$\kappa_{gaze}, \kappa_{card}$	0	16	0	16	0	0
$\omega_{gaze}, \omega_{card}$	-4	16	-4	16	-4	0
ζ	$log(0.5)$	16	$-\infty$	0	$log(0.5)$	16
β	-	-	-	-	$log(48)$	16

		Predicted Class	
		Y=1	Y=0
Actual Class	Y=1	TP	FP
	Y=0	FN	TN

Table 5.3: Confusion matrix

density of the model probabilities, given the log-evidences of all subjects (Stephan et al., 2009a). Inference was performed on the model space by comparing the exceedance probability of each model.

Finally, balanced accuracy was calculated for all the models for evaluating their performance. Balanced accuracy is a metric to evaluate the performance of a binary classifier. In our dataset, the two classes were $y = 1$, if the subject chose the card indicated by the gaze, and $y = 0$, if the subject chose the other card. Table 5.3 shows a confusion matrix, which is a table that displays the performance of a classifier in terms of true positives (TP), false positives (FP), false negatives (FN), and true negatives (TN) (19). Balanced accuracy takes the performance in each class into account:

$$Balanced\ accuracy = \frac{TP}{2(TP+FN)} + \frac{TN}{2(TN+FP)} \tag{5.15}$$

5.2.8 Multivariate Regression

Given the assumption that the autistic spectrum can be characterized by differences in the extent to which individuals weight social information rather than an inability to process them, we predicted that autistic traits (as measured by AQ scores) would be associated with the degree to which individuals weight the gaze cue while making decisions. More precisely, the model parameter ζ was of particular interest and expected to be negatively correlated with AQ scores. A large value of ζ therefore signals that a participant preferentially bases his/her decisions on the social advice, i.e. the gaze cue, compared to other cues during decision-making. To test this, a multivariate regression analysis was applied on the AQ scores by using the model parameters of the winning model as predictors (see below for model space).

Since the parameter ζ is estimated in log-space, it was included in the regression in log-space as well. All correlations were performed with bootstrapping (2000 bootstraps) and 95% confidence intervals. To demonstrate the specificity of the significant predictors to the AQ scores, a full model was designed including the following other variables: gender, age, systemizing quotient (SQ), empathy quotient (EQ), and IQ (verbal) scores.

5.2.9 Other Behavioral Measures

In the post-test questionnaire, the participants were asked to rate between 0 (not at all helpful) and 100 (extremely helpful) how helpful they had found the face. For this, a mean rating of 37.63 was obtained, which could be taken to suggest that there was some level of awareness.

Since the ability to exploit the additional social information should contribute to the overall task performance, the relationship between AQ scores and individual total task scores was assessed. Autistic individuals have a low tolerance for unpredictable situations (Robic et al., 2015). Therefore, it could be argued that people with high autistic traits infer differently about the gaze validity in high volatility compared to low volatile phases than people with low AQ. As a result, the influence of the environment, i.e. the changing probability and volatility of the gaze cues, on performance was compared between the two AQ groups, which were obtained using a median split procedure (median AQ = 15). The association between autistic traits and advice taking behavior on volatile low probability gaze cue trials (circled phases in Fig. 5.11B) was evaluated separately by means of correlation analyses.

5.3 Results

5.3.1 Simulations

Different precision weightings

Responses were simulated for different agents who have the same perceptual model parameters $\kappa = 1, \omega = -4, \theta = 0.25$, but different precision weightings. This results in the same learning trajectory (not shown here), but different choice probabilities $p(y|m)$ due to the different range of observation model parameter ζ. Figure 5.4 shows the sensitivity of ζ to the output, i.e, probability of taking the advice for different values of ζ covering the range from low ($\zeta = 0$) to high ($\zeta = 3$) weightings.

Parameter recovery

Decisions were simulated for different values of κ, ω, θ, and ζ and the model was inverted for each dataset to test whether the estimation procedure can recover these values. Figure 5.5 shows the means and standard deviations from 200 simulations for each parameter. It can be observed that while the model can identify the differences in ζ, ω and κ to a very good extent, it could not capture θ values equally well. This is an expected result, as it gets more challenging to recover the parameters in the higher levels of hierarchy due to the increasing errors or decreasing signal-to-noise ratio.

Influence of volatility on (dynamic) learning rates

Figure 5.6 shows the trajectories of learning rates at the second and third levels for gaze and card cues. Since the second level estimates $\mu_{2,i}$ is in logit space, we transform the first

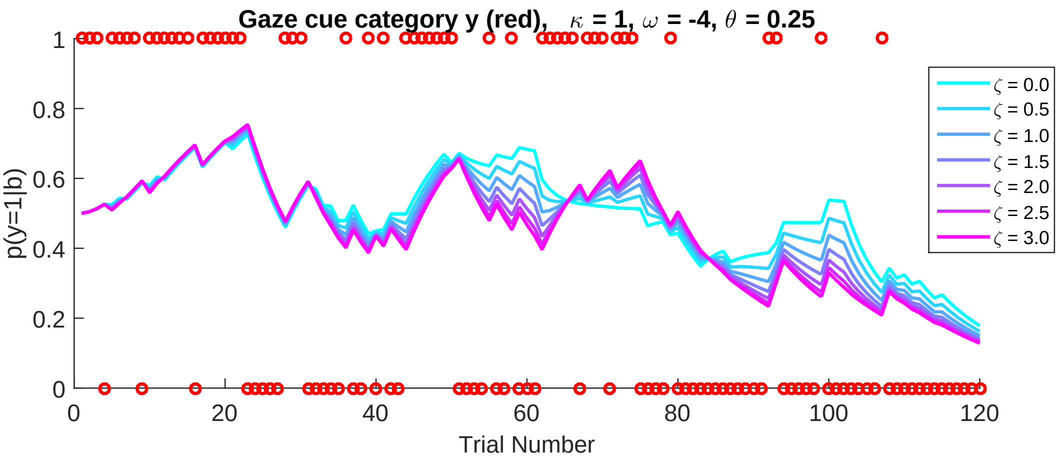


Figure 5.4: Simulations for an agent with same perceptual but different social cue weighting values. Probability of taking the gaze advice $p(y = 1|b)$ under different precision weighting parameters (light blue to purple) for a given input (red). The input is coded binary such that u = 1 if the gaze shows the winning card, u = 0 otherwise. The agent has the same perceptual parameter values, $\kappa = 1, \omega = -4, \theta = 0.25$.

learning rate $\alpha_i^{(1)} = \sigma_{2,i}$ to the first level following the approach in Iglesias et al. (2013):

$$q(\alpha_i^{(1)}) = s(\mu_{2,i})(1 - s(\mu_{2,i}))\alpha_i^{(1)}. \tag{5.16}$$

5.3.2 Empirical Dataset

Model Comparison

The model selection is based on the model evidence, which is a principled measure of the balance between model fit and model complexity (Friston et al., 2007). Model comparison was in favor of Model 1 (exceedance probability of 0.9408), suggesting that a hierarchical Bayesian model in which participants weighted both social and reward-related information best described subjects' responses. The exceedance probabilities for Model 2 and Model 3 were 0.0384 and 0.0208, respectively. Also, balanced accuracy was used to compare performances of the three models. A balanced accuracy of 0.61 for the winning model (Model 1), 0.54 for Model 2, and 0.58 for Model 3 was found (Fig. 5.7).

Learning multiple cues in a volatile environment

The mean and standard deviation of MAP values of parameters across participants are displayed in Fig. 5.4. Trajectories of perceptual model from an example subject is shown in Fig. 5.8.

Model Diagnostics

Potential identifiability issues regarding the model parameters were assessed. The posterior covariance matrix was computed for each participant. This is estimated by calculating the Hessian at the maximum a posteriori (MAP) estimates. The negative inverse of the Hessian

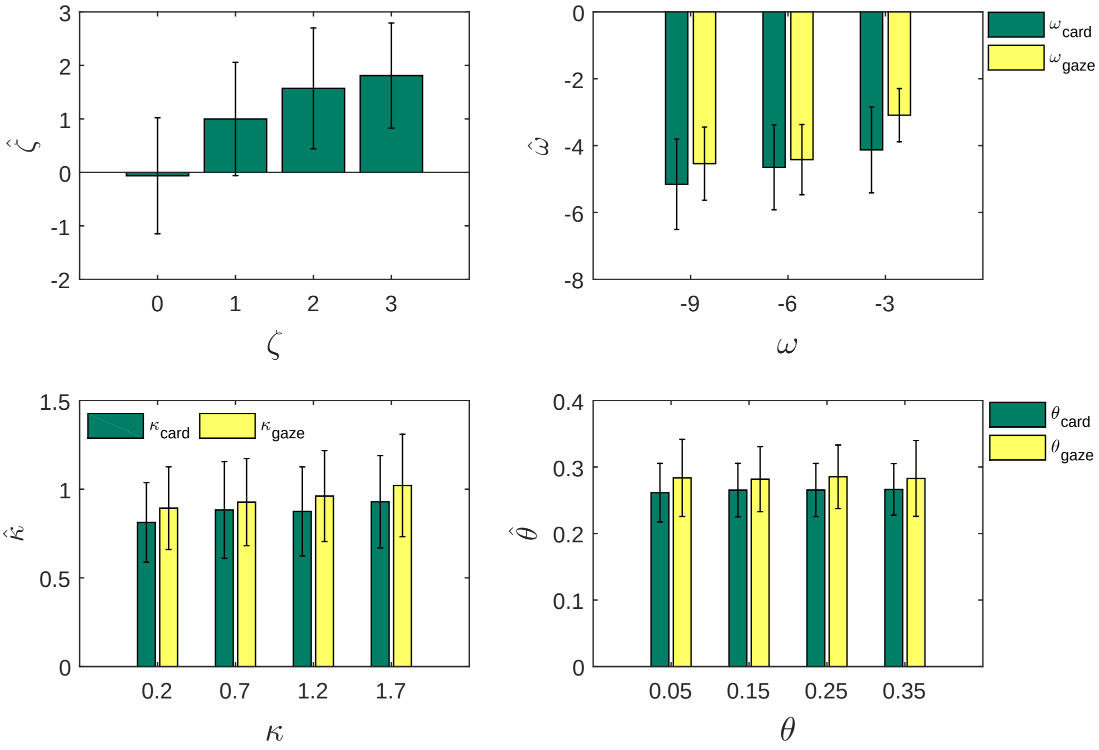


Figure 5.5: Simulations for parameter identifiability. Decisions were simulated 200 times for each agent with one different parameter value at a time (x-axis). Parameters were estimated (y-axis) for each simulation. While the estimation procedures captured differences in ζ, ω, and κ, it did not perform equally well for θ. Bars show mean estimates, and lines show standard deviations.

is the parameter covariance (Mathys et al., 2014). Some, but not high, correlations between κ and ω were observed, with a correlation of -0.34 (Fig. 5.9).

5.3.3 Construct Validity

Model Parameters as Predictors of AQ scores

A multivariate regression was conducted to investigate an assumed relationship of gaze-cue related model parameters ($\theta_{gaze}, \kappa_{gaze}, \omega_{gaze}, \zeta$) of the winning model and AQ scores. The analysis shows that ζ values, i.e. weighting the gaze, did significantly predict the AQ scores ($\beta = -1.60$, t(31) = -2.77, p = 0.0095). Other parameters θ_{gaze} ($\beta = 12.37$, n.s.), κ_{gaze}($\beta = -0.66$, n.s.), and ω_{gaze}($\beta = -0.18$, n.s.) were not significant predictors (F(4, 31) = 2.02, $R^2 = 0.21$). As negative coefficients in the multivariate regression analysis do not mean that there actually is a negative correlation between the response and the predictor, the direction of the association between AQ scores and the advice weighting parameter ζ was explored by performing a correlation analysis (Fig. 5.10, left). The Pearson's correlation coefficient between zeta parameter and AQ scores was -0.42 with 95% confidence intervals $(-0.66/-0.19)$.

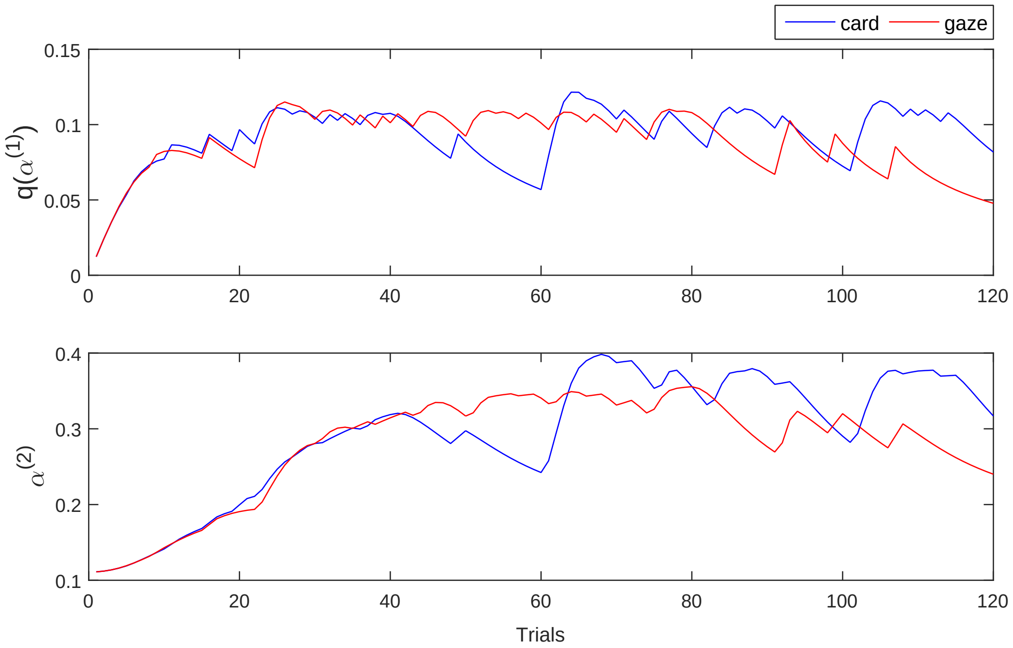


Figure 5.6: Dynamic learning rates. The agent updates its belief with time varying learning rates. While $\alpha^{(2)}$ depends heavily on the environmental volatility $e^{\kappa_i \mu_{3,i}^{(t-1)}+\omega_i}$, $q(\alpha^{(1)})$ also depends on the variance parameter from the first level σ_1. Please refer to definitions of variance parameters in the background section (Eq. 2.31). The volatile phase for the gaze cue is between trials $30-70$, and for the card cue it is between trials $60-120$. The agent had the following values for the perceptual model parameters: $\kappa = 1; \omega = -4; \theta = 0.25$. Note that a higher κ value would increase the influence of volatility estimate $\mu_3^{(t)}$ on the size of the update (Eq. 5.11)

.

A hierarchical regression was carried out in order to assess the unique contribution of ζ. The first regression model included all variables except ζ, i.e. $\theta_{gaze}, \kappa_{gaze}, \omega_{gaze}$, and the second regression model included all variables. Model comparison performed by means of ANOVA revealed that including ζ in the regression significantly improved the fit of the model to the data $F(1, 31) = 7.65, p = 0.0095$.

To address the specificity of the zeta parameter to the AQ scores, a full model was designed: A multivariate regression analysis including explanatory variables of AQ, IQ, SQ, EQ, Age, and Gender, was used to predict social gaze weighting parameter ζ. The analysis shows that only AQ scores significantly predict the parameter zeta ($\beta = -0.11, t(31) = -2.18, p = 0.037$). Other descriptive scores IQ ($\beta = 0.02$, n.s.), EQ ($\beta = -0.02$, n.s.), SQ ($\beta = 0.003$, n.s.), Age ($\beta = 0.09$, n.s.), Gender ($\beta(male) = 0.46$, n.s.) were not significant predictors (F(6,29) = 1.814, $R^2 = 0.27$).

Relationship of total scores, AQ and ζ

Individual differences in AQ scores were significantly correlated with participants' total scores (r = -0.39 with 95% confidence intervals, $-0.68/-0.13$; Fig. 5.10, middle). Also, a rela-

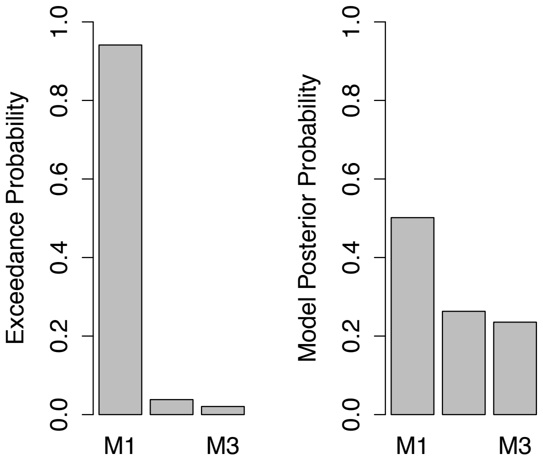


Figure 5.7: Bayesian Model Comparison. Exceedance probabilitiy (left) and posterior probability (right) for all models in the model space. Model selection procedure favors the Model 1 ($xp = 0.9408$).

Table 5.4: Mean and standard deviation of the maximum a posteriori (MAP) estimates of the winning model parameters. (ζ is in log-space).

	θ_{card}	θ_{gaze}	κ_{card}	κ_{gaze}	ω_{card}	ω_{gaze}	ζ
Mean	0.3010	0.2859	0.9528	0.9041	-4.8329	-4.5240	0.9284
Std	0.0340	0.0630	0.3514	0.3805	1.5252	2.0813	1.5006

tionship between total scores and ζ was observed such that a more pronounced weighting of advice was related to higher total scores (r = 0.50 with 95% confidence intervals, 0.20/0.86; Fig. 5.10, right).

Association between AQ scores and the utility of misleading advice in a volatile environment

The scores obtained in each phase of gaze accuracy, i.e. stable high accuracy, stable low accuracy, volatile high accuracy, volatile low accuracy, were used as a direct measure of behaviour. Figure 5.11A illustrates the performance in each phase of the experiment. A significant difference between two groups during the volatile low probability phase was observed (Welch's t-test: t(33) = 2.21, p= 0.034). This phase is marked with blue circles in Fig. 5.11B. Similarly, AQ correlated with the number of trials where the subjects took the advice ($r = 0.52, 95\% CI = 0.29/0.75$) in the same phase (Fig. 5.11C). Therefore, even during the volatile phases of the experiment the low AQ group was able to take advantage of the misleading advice by avoiding it.

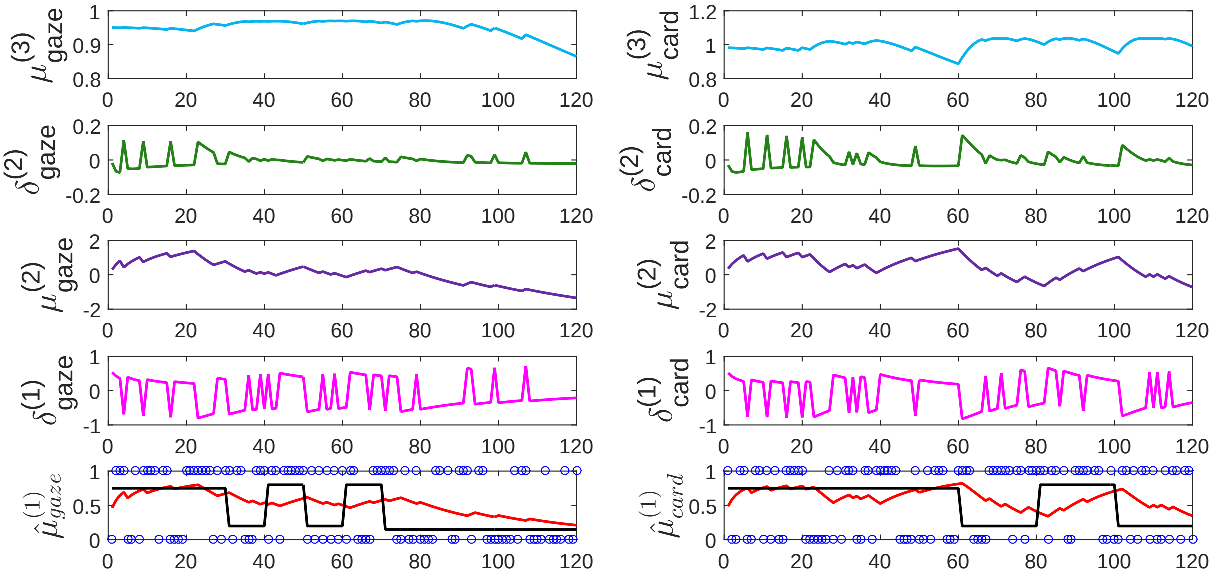


Figure 5.8: Learning trajectories from an example subject for gaze (left) and card cues (right). For the gaze cue: $\mu^{(3)}$ is the posterior mean of the third level variable and describes the volatility of the gaze accuracy (left top, light blue). As expected, it decreases towards the end of the experiment, due to the stability in the outcome in the last phase of the experiment (left bottom, black line), and it is higher in the volatile phase. Subject's estimate of gaze probability $\mu_{(1)}$ (bottom plot, red), which is a sigmoid transform of second level estimates $\mu_{(2)}$ (middle plot, purple), is close to the true probability of the outcome (bottom plot, black line). Blue dots (bottom plot) are the actual outcomes ($y = 1$ for correct gaze advice, $y = 0$, for incorrect gaze advice). Model estimates for card cue perform similarly well (right plots). The first level and second level prediction errors $\delta^{(1)}$ and $\delta^{(2)}$ that are used to update estimates at each trial are plotted on magenta and green plots, respectively.

5.4 Discussion

In this study, hierarchical Bayesian modeling was implemented to study the learning behavior of an agent in a complex environment. The learning task included two cues whose features (probability and volatility) had to be learned and combined by the agent for an optimal performance. The HGF and the precision weighted response model allowed us to model individual differences in learning and decision making of different agents. The performance of this approach was evaluated with simulations and a real dataset from 36 subjects.

The three level HGF is a perceptual model with a Bayesian structure. The priors on the means and variances of the parameters were informed by previous research, which showed that the HGF performs well for these prior values. The parameters that were estimated in the logit space (κ_i, ω_i) were bounded theoretically: The lower bound was zero to keep them in the positive range, and upper bounds depended on the assumptions of variational approx-

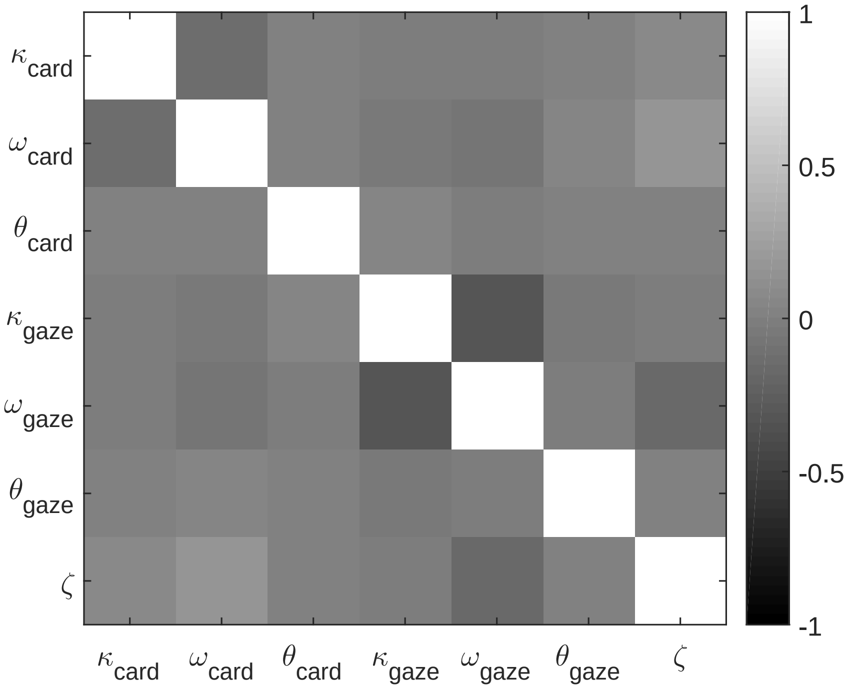


Figure 5.9: Correlations among the parameters. Correlation matrix was computed from posterior covariance matrix (see text for details). The highest correlation was observed between κ_{gaze} and ω_{gaze}, with a correlation of -0.34.

imation (Mathys et al., 2011). Using simulations we saw that these values are meaningful for the experimental input presented here such that agents with these parameter values could predict the probability and volatility of both cues appropriately. Further, keeping the prior variances relatively large ($\sigma^{(0)} = 16$) allowed for estimating parameters that span a range so that the behavior of different learning agents could be inferred.

The approach that is presented here can be considered as a *latent-mixture model* as it assumes that data is generated by two processes whose properties are not observed or latent (Lee and Wagenmakers, 2014). The two processes are the individual's perceptual models about each cue and the data is the choice behavior that is generated by combining beliefs about the cues via the response model. The precision weighting response model integrated the first level estimates or accuracy, but it also takes into account the volatility of the cues. Therefore, the stochasticity in the decision making became a function of volatility of the card and the gaze cues, unlike the temperature parameter that is used in softmax decision rule (see Background). Equation 5.8 indicates that volatility of each cue has an additive effect on the temperature parameter, which results in an agent who behaves most deterministic when both cue probabilities are stable.

Differences in the parameter values of the HGF can be captured for a single cue independent from the optimization procedure, e.g. Markov Chain Monte Carlo and variational Bayes (Mathys et al., 2014). The quasi-Newton optimization method that was used in this study is

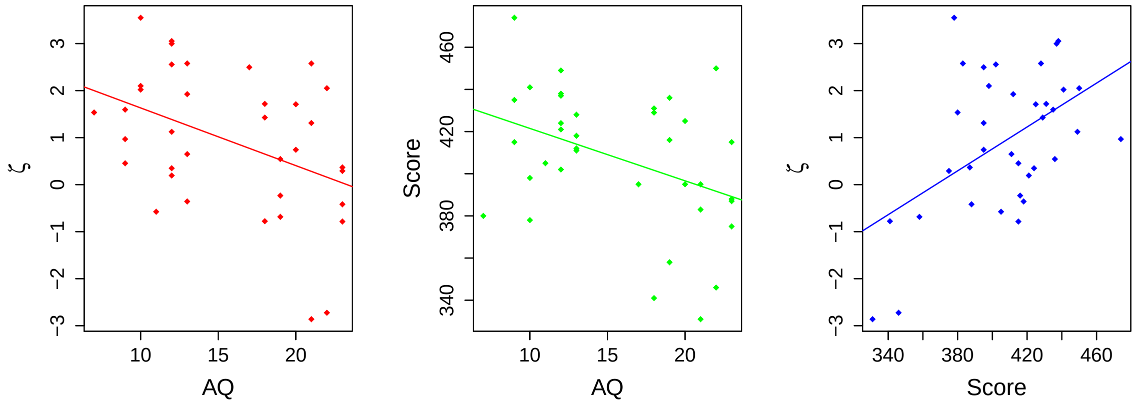


Figure 5.10: Relationship of total scores, AQ and ζ. Left: Negative correlation between ζ parameter and AQ traits (r = -0.42, $95\%CI = -0.66/-0.19$). Middle: Negative relationship between subjects' AQ traits and their score at the end of the task (r = -0.36, $95\%CI = -0.68/-0.13$). Right: Positive correlation between total score and ζ parameter (r = 0.5, $95\%CI = 0.20/0.86$).

a robust algorithm, but it can be susceptible to local minima in comparison to other global optimization methods such as the Gaussian process optimization (GPO), (Lomakina et al., 2015). As far as the complexity of the parameter space is considered, this could be an issue. Therefore, the performance of the estimation procedure needs to be evaluated with different algorithms in the future. Nevertheless, simulations showed that when parameters have similar values for different cue types, they could be recovered by the estimation procedure for the values within their corresponding range (Fig. 5.5).

In reinforcement learning, using an adaptive learning rate can outperform the models with a fixed learning rate to handle the complexity in volatile environments. The analogy between the HGF and reinforcement learning can be seen in the update equations (Eq. 5.9 and 5.10). At both second and third levels, the agent updates its belief in every trial with a step size of learning rate times the prediction error from the level below. The learning rate for volatility update $\alpha^{(2)}$ increases during the stable phase, but this increase stops in the volatile phase. This is an expected result because the weight on the prediction errors should be higher in the stable phase where the agent can use a new piece of information with more certainty (Fig. 5.6 bottom). However the influence of environmental uncertainty $e^{\kappa_i \mu_{3,i}^{(t-1)}+\omega_i}$ is less pronounced for the learning rate $\alpha^{(1)} = \sigma_2$ (Fig. 5.6 top). This is due to the definition of variance parameter σ_2 (Mathys et al., 2011), which has counter effects of the environmental uncertainty with the uncertainty at the first level σ_1 (see Eq. 2.31).

We observed that parameter estimation can capture a wide range of perceptual and response model parameters to a good extent. This was an important finding for the utility of the HGF in capturing individual differences, which is important for the clinical application. A final set of simulations was run for testing the parameter identifiability in case the parameters are very different for each cue ($\theta_{gaze} = 7\theta_{card}, \kappa_{gaze} = 8\kappa_{card}, w_{gaze} = 3w_{card}$) such that the values were selected at the boundaries of parameter ranges. However, the HGF could

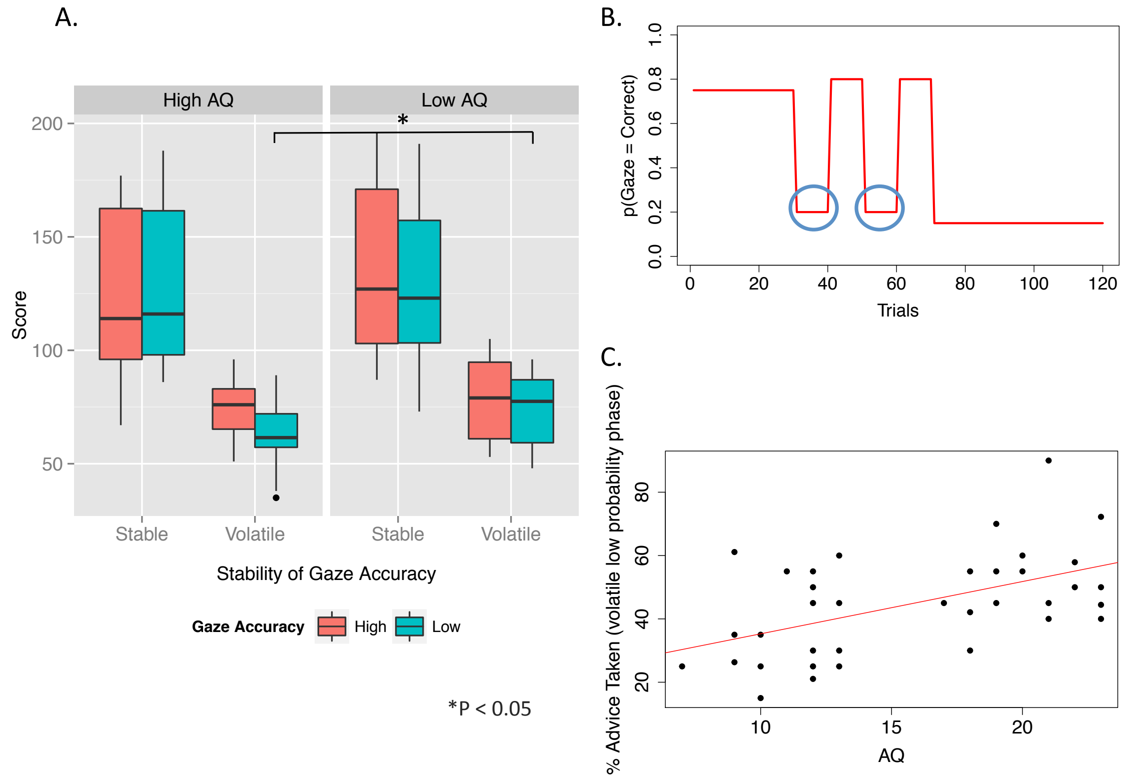


Figure 5.11: The influence of structure of the environment on the behavior. (A) Scores obtained by high and low AQ groups in different phases of the experiment based on the features of the gaze cue (high x low gaze accuracy and stable x volatile periods of gaze accuracy). Difference is significant (*p = 0.034) in the (B) volatile low accuracy phase (circled area). (C) During the same phase the number of trials in which the subjects took the advice, i.e., chose the card that is indicated by the highly misleading gaze, was correlated with AQ traits ($r = 0.52$, $95\%CI = 0.29/0.75$).

not capture these big differences in the perceptual parameters (not shown here). Although this was not a big limitation for the analysis presented here due to the focus on the social cue parameters alone, this issue needs to be addressed in future studies where one needs to identify processes that differ with respect to the cue type. One solution to this problem could be to simply increase the number of trials to obtain a better model fit and to design the experiment such that the volatile phases for different cues do not overlap.

The original model was compared to alternative models. The second model assumed that subjects make their decisions by ignoring the gaze cue completely. If $\zeta = e^{(-\infty)}$ is inserted in Eq. 5.4, one obtains the following weights: $w_{gaze} = 0$, $w_{card} = 1$, hence the combined belief for each trial becomes $b^{(t)} = \mu_{1,card}^{(t)}$ such that the choice behavior is based on the card probability alone. The third model was similar to the approach presented in Behrens et al. (2008), because it corresponds to a hierarchical Bayesian model for an optimal learner. To obtain this model, the free parameters were fixed by setting their prior variance to zero, and their prior means to the values of an optimal learning agent. As explained above, these values were informed from earlier studies and evaluated with simulations for plausibility. However, both alternative models had less evidence compared to the original model as indicated by Bayesian model selection. Also, balanced accuracies pointed out that the original model is more generalizable due to its higher predictive accuracy.

Parameter interdependence is frequently observed in hierarchical Bayesian models of cognition (Scheibehenne and Pachur, 2015). For model diagnostics, we considered the correlations among the parameters by plotting the posterior covariance. Only small correlations were observed between κ and ω. As a solution, a reparameterization can be applied by fixing one correlated parameter at a time, and retest for the predictive power of the parameter of interest. This approach was performed for the specificity of the ζ parameter to predict AQ traits by fixing κ or ω at a time and repeating regression analysis (not shown here). However, this was performed only for a sanity check because fixing each parameter leads to a different model and its evidence needs to be compared for any further conclusions. Another recommendation for the future evaluation of parameters would be to use a test-retest reliability to measure the stability of the parameters such that high correlations for each parameter from different measurement times mean a high stability.

We further investigated autistic trait-related differences in the extent to which healthy individuals integrate and make use of gaze cues. Here, the focus was on modeling perceptual as well as higher-order processing of both card and gaze cues and, in particular, their relationship to action selection, i.e. the extent to which individuals were actually biased by the social information provided on a trial-by-trial basis. Results of the computational analyses demonstrate striking evidence for AQ-related differences, such that individuals lower in AQ scores are influenced by the gaze cue more as indicated by a negative relationship between the response model parameter ζ and autistic traits. Importantly, these results show a positive relationship between ζ and the total individual scores obtained in the experiment, which indicates that reliance on the social information was actually what was helping subjects lower in AQ to obtain higher scores. Furthermore, the results indicate that individuals high in AQ had particular difficulties to use the social advice under volatile conditions.

By providing these new insights into AQ-related differences in social cognition, this study, is most relevant to current discussions concerned with mechanistic explanations of the autistic symptomatology: Predictive coding theories have reconstructed autism in terms of high-level attenuated precisions relative to sensory precision (Friston et al., 2013), which results in an enhanced weighting of prediction errors (Lawson et al., 2014) and a loss of the selective force when processing a context with multiple cues (Van de Cruys et al., 2014). As stated by Pellicano and Burr (2012), Bayesian models provide an important avenue, which can help to identify whether autistic trait-associated alterations lie in the reliance on prior knowledge or the optimal update of prior information during learning. In Bayesian formulation, this issue was addressed by assessing possible relationships of perceptual and response model parameters with AQ scores. However, no relationship between the perceptual model parameters and the AQ scores was found, which is suggestive of an intact inference machinery. On the other hand, the response model parameter ζ, which constitutes the weight on the precision of inferred gaze probability (see Eq. 5.4), reflected that participants who scored higher on the AQ questionnaire could not take advantage of the learned precision estimates of the social cue when mapping representations to beliefs. Taken together, these results suggest that the mechanisms of estimating the precision of social information do not differ, but that the application of new, updated priors depends upon the level of autistic traits. These findings appear consistent with a recent suggestion by Palmer et al. (2015), who propose that autism may not impair the ability to process social information per se, but rather lead to differences in how the relevant representations are weighted for action selection.

Another implication of this study is that - to the best of the author's knowledge - it is the first to utilize a hierarchical Bayesian model in the context of autistic traits. The modeling approach that is implemented here is a promising method for capturing individual differences in the learning and integration of social information. Given the heterogeneity of the population, this could be particularly useful for identifying subgroups that may map onto distinct mechanisms of impaired social interaction in autism. The "Observing the observer" approach has, indeed, been demonstrated to be useful for inference on hidden states and parameters that shape inter-individual differences in learning (Daunizeau et al., 2010b). Hierarchical Bayesian models of learning such as the HGF linked to action selection have been implemented in several different learning contexts (Diaconescu et al., 2014; Iglesias et al., 2013; Paliwal et al., 2014). The results indicate that Bayesian models may be particularly powerful in providing mechanistic explanations of social difficulties, which are particularly relevant to an understanding of psychiatric disorders (Schilbach et al., 2013; Schilbach, 2014, 2015). Advances in computational psychiatry (Montague et al., 2012; Friston et al., 2014; Stephan and Mathys, 2014; Wiecki et al., 2015) and studies such as this could, therefore, contribute to mechanistic formulations of psychopathology.

6 General Discussion

In this thesis, computational models of learning and decision making have been adapted and evaluated as a framework, which can be used for answering important questions in neuroscience and psychiatry. This chapter will discuss the general technical points about modeling and imaging methods that were presented throughout the thesis.

Reinforcement learning models
In the first study, participants performed a probabilistic learning experiment, which is also known as a bandit task in reinforcement learning, where they chose one option in each trial to maximize their total reward. Participants' performance in avoidance learning was associated with their genotype, whereas approach behavior i.e. performance in choosing the best action was not affected. The learning behavior was modeled with a Q-learning algorithm. During the model fitting, the free parameters α and β were restricted in meaningful ranges that were selected similar to earlier studies (Jocham et al., 2011; Frank et al., 2004; Frank and Hutchison, 2009). Log-likelihood was computed on a 2-D grid for each α, β combination. One should use built-in optimization functions (e.g. fminsearch or fmincon) with caution and use appropriate boundaries for free parameters, since the very small values of β would return infinity, hence, the softmax function can cause crashes (also see Daw (2011) about optimization of free parameters). In terms of exploration/exploitation behavior, most subjects maintained an exploratory behavior for the most conflicting stimulus pair EF, but they were more exploitative in the AB pair since it was easy to learn and their estimates converged quickly to the real values of the stimuli. Note that both free parameters are fitted to the complete dataset in one optimization procedure, i.e. it was assumed that a subject learns each pair with the same learning rate and decision temperature. Therefore, the parameter α and β did not capture possible differences in learning of different stimulus pairs during the training phase. For instance, in Fig. 3.6 the exploration in EF trials is higher for the subject displayed with red compared to the other subject, although the model estimates similar values for the temperature parameter β (0.23 and 0.27) due to the similar exploratory behavior in other stimulus pairs. Nevertheless, this is not a large limitation, since on average, learned Q values at the end of the training approached the real values (Fig. 3.7).

The strategy used by subjects for updating their beliefs about stimulus was assessed by comparing the single α model to an alternative model with separate learning rates for positive and negative feedback, α_+ and α_-. Since in general the more free parameters, the better the fitting would be, instead of simply comparing maximum likelihoods, BIC and exceedance probability were used as a measure of model comparison, which favored single α model. At the end of the experiment subjects' estimates of action values converged to the real reward frequency of the stimulus, confirming that RL model with single learning rate approximated the learning behavior adequately.

Model based fMRI analysis showed that different brain regions were involved in processing positive and negative feedback, and prediction errors. While striatum and mPFC engaged in processing positive feedback or rewards, negative feedback was processed mainly by ACC and insula, which is also known as error processing network (Bastin et al., 2016; Hester et al., 2009). This implies that these structures are involved when one makes an error (not in a quantitative meaning as we used in 'prediction error' term). These results support the functional segregation of brain systems for processing different types of reinforcers.

The BOLD signals in the midbrain, vStr and ACC were correlated with reward prediction errors. Neurogenetic analyses revealed that carrying alleles of *FTO* and *ANKK1* cause a decrease in the prediction error related activity in the midbrain. The association of the performance in avoidance learning and midbrain activity suggested a neurocomputational mechanism for the genetic modulation of behavior.

Functional imaging of midbrain

fMRI is an advanced and non-invasive modality, but it provides an indirect measure of neuronal activity. Further, due to the small sizes of VTA and SN structures, with its low spatial resolution one should be cautious with functional imaging of the midbrain. For example, the size of the VTA is approximately 60 mm^3 (D'Ardenne et al., 2008), which corresponds to 2 voxels for a voxel size of 3 mm in this study. The brain regions where the fluctuations in the measured BOLD signal correlated with the fluctuations in prediction error were reported. However, it raises the question whether the prediction error related activity reflects the activation of DAergic neurons. Within VTA/SN complex, there are A9, A10 subgroups and GABAergic inhibitory neurons and excitatory glutamatergic neurons (Molochnikov and Cohen, 2015). BOLD signal reflects the mean activation of a larger population of neurons in a certain region rather than a subgroup of neurons. Similarly, activation of inhibitory and excitatory neuronal populations in a region might cancel out the signal so that the computational processes cannot be captured by fMRI signal (O'Doherty et al., 2007). A functional dissociation within this complex can influence behavior and performance differentially (Alderson et al., 2008). Therefore, more precise results can be achieved with the use high-resolution fMRI combined with other modalities e.g. PET imaging (Düzel et al., 2009) for imaging the midbrain. Nevertheless, solutions exist for other issues regarding functional acquisition of VTA/SN signals such as using volume shims for field inhomogoneities or correcting for physiological noise from the recordings of cardiac and respiratory activity (Iglesias et al., 2013). Finally, other possible explanation for negative prediction error processing could be the existence of an additional system apart from dopamine, e.g. serotonin (Doya, 2002).

Dynamic causal modeling limitations

Despite its advantages, there are several limitations of DCM for fMRI to study learning related connectivity changes. First of all, the fMRI task had an event related design, where the sustained activation to study modulatory effects of experimental inputs is less compared to a blocked design (Ewbank and Henson, 2012). Second, the fact that the BOLD signal is indirect, hence a slow measure of neuronal activity, and that data points were collected with a repetition time (TR) of 2 seconds might result in underestimating the connectivity strengths (Camara et al., 2009). Another issue related with DCM regards the model specification. It cannot be used as an exploratory method. Instead, one needs to specify hypothesis driven regions in the model space. Furthermore, it is possible to do model partitioning based on different criteria. In this thesis, model space was partitioned with respect to the nonlinear interaction matrix D, but one could group the models regarding the input configuration instead, hence having 2 families, i.e. the first family with input to the midbrain, and second family with input to all regions, with 4 members (linear and nonlinear) in each family. The rationale of partitioning the model space with respect to the linearity was to capture possible gating roles of the midbrain. Therefore, it is recommended to consider the hypothesis during the model family partitioning.

Model comparison

Throughout the thesis, model comparisons were implemented using Bayesian statistics. Depending on the data and hypothesis, one can follow several model comparison steps. Fixed effects analysis is more suitable when studying basic physiological mechanisms such as vision (Penny et al., 2010), but random effects analysis assumes that different subjects can use different models in the same population. This is more likely the case in cognitive experiments e.g. in reward based learning, where the task related connectivity is likely to depend on individual biological factors such as DA levels or number of autoreceptors. Similarly, in spectrum disorders such as autism or schizophrenia, individuals might follow different strategies in learning and decision making. Therefore, throughout the thesis, random effects inference was preferred on both neuroimaging and behavioral data.

In addition to Bayesian model selection, balanced accuracy method was also used to evaluate performance of the models. However, this measure cannot replace model evidence due to the fact that it is not sensitive to the estimation of the observation model. Since any prediction below 0.5 will be considered as negative class by balanced accuracy, this measure will not be sensitive to the performance of model estimation. Therefore, using model evidence is superior to using balanced accuracy in such models.

Deciding the best fitting model depends also on the estimation technique. For reinforcement learning models, BIC and exceedeance probabilities were calculated by using maximum likelihoods. On the other hand, for the DCM and the HGF, model evidence or free energy was used, which incorporated the priors over the parameters $p(\theta|M)$ for model comparison. Since prior distributions are defined over model parameters in Bayesian estimation, it is necessary to integrate over these priors.

Hierarchical Bayesian models of learning

How an agent learns in a dynamic environment is another fundamental question of adaptive intelligence systems. In the face of uncertainty, human behavior was explained better by Bayesian updating than classical reinforcement learning algorithms (Payzan-LeNestour and Bossaerts, 2011). Bayesian algorithms have addressed the updates of an agent's beliefs as a joint probability distribution over the state variables (Behrens et al., 2007). Examples of Bayesian approaches to understanding how the brain can implement these computations are numerous, e.g. a hidden Markov model called Dynamic Belief Updating has shown Bayesian surprise signals in ACC (Ide et al., 2013), importance sampling methods were adapted to explain hierarchical message passing in visual cortex (Shi and Griffiths, 2009), or the representation of dynamic learning rate in the brain (Behrens et al., 2007).

In Chapter 5, a hierarchical Bayesian model was adapted to present a framework to study autistic traits related to differences in learning and decision making. In the experiment, gaze cues in the middle of the screen provided a social component for learning. Social agents - who unlike exclusively physical objects in the environment have a life of their own and do not follow simple rules most of the time - are arguably the most difficult to predict (L. Schilbach, personal communication, October, 2015). Consequently, the presence of other social agents increases uncertainty of a given situation dramatically. Also, it is in such uncertain situations that we have to rely most on our prior beliefs to make sense of a situation, which might be particularly hard for individuals on the autism spectrum.

As a methodological novelty, the Hierarchical Gaussian Filter (HGF) described in Mathys et al. (2011) was extended for multiple cues. Combining the HGF with the precision weighted response model made it possible to test perceptual hypothesis mechanistically. The original model pair was compared with two alternative and nested models. A model with no social component (Model 2) was included because participants were not given any explicit information about the nature and relevance of the gaze cue. Also, suboptimal performance in human subjects has been reported when dealing with perceptual estimation in uncertain environments (Landy et al., 2007). For these reasons, one had to make sure that subjects did not fully ignore the gaze (the reduced model), but take it into account during decision-making. Model comparison showed that this was the case, i.e. the full model that combines both sources (social and non-social) explained the behavior better. A third model with fixed perceptual parameters was applied and included for model comparison. This model corresponds to an optimal Bayesian learner and its inclusion in the model space was motivated by the fact that human learning behavior resembles an optimal Bayesian learner in similar settings (Behrens et al., 2007). Theoretically there are many other models that could be tested, such as models that have no third level in the hierarchy, or models that have no hierarchy at all. However, previous research has already shown that participants can process related features of the environment, such as volatility, in similar experimental settings.

For incorporating the influence of volatile structure of the environment on decision making, a precision weighted response model was used. Unlike softmax and e-greedy functions, it considers the volatility while modeling the degree of exploration. According to this (see Eq. 5.8), the more volatile the reward probabilities are, the more an agent will explore. For the study presented here, this means that a subject will explore more between taking the gaze advice and going against it in the highly volatile phases of the experiment. Similarly, if the reward associations of gaze are more stable, the subject will be more exploitive in his decisions. This behavior can be observed in the Fig. 5.4 where the probability of taking the advice for the simulated agents oscillates around 0.5 during highly volatile phases, i.e. trials 30 to 70. Therefore, response models incorporating the volatility feature of the environment are more appropriate for decision making in dynamic environments.

To further evaluate the influence of volatility on learning and decision making, differences in the number of points won during different phases of the experiment were reported. Similarly, the extent to which the subjects used the social gaze could be nonstationary and captured by modeling. However, due to the relatively small number of trials, an analogous analysis for the modeling parameters is beyond the scope of this study. There is a trade-off between the number of trials and the specification of the priors over the parameters of the model. The HGF has been tested across a variety of trial numbers and levels of noise (Mathys et al., 2014). However, ultimately, the choice of priors will determine whether the data generated (including the number of trials) are adequate to capture the learning prescribed by the model. In this case, the priors were relatively tight, as their selection was informed by previous implementations of the HGF in studies of advice-taking (Diaconescu et al., 2014) and associative learning (Iglesias et al., 2013). The current modeling approach presented here uses a static parameter to describe an overall measure of weighting the social information. However, in the future this can be addressed with a different modeling approach by using a nonstationary weighting parameter such as $\zeta(t)$ and with a paradigm including more trials for different phases.

Finally, possible concerns about sample size and specificity of the results were addressed by presenting bootstrapped correlations and a full model multivariate regression analysis, which includes all the available subject variables, namely age, gender, IQ, empathy quotient, and systemizing quotient. Crucially, this reanalysis demonstrates that the findings for the social weighting parameter are, in fact, highly specific to AQ scores. Moreover, it provides statistically robust evidence that this distinction can already be made based on this carefully selected sample of 36 subjects.

7 Conclusion

7.1 Summary and Outlook

The utility of reinforcement learning algorithms in neuroimaging and behavioral studies was evaluated. A standard Q-learning model was compared to an alternative model with separate learning rates for different types of feedback, i.e. reward and punishment. Simulations evaluated the behavior of agents with different learning rates and temperature parameters to understand the effect of different values of these parameters on the belief update, exploratory behavior, and convergence to the real action values. It was observed that different parameters result in unique trajectories, while the initial action values did not influence the speed of convergence. In the model with separate learning rates, the agent became too sensitive to one feedback type when the learning rate for that feedback was much larger. Both models were compared on a real dataset from a large number of subjects by using protected exceedance probability and Bayesian information criterion. This comparison favored the original model with a single learning rate, which was then used for further fMRI data that were collected while participants performed the learning experiment. Prediction error processing is long accepted to be mediated by dopaminergic neurons. Incorporating prediction errors that were derived by the learning model in the general linear model analysis allowed us to locate the activity related with prediction errors in addition to positive and negative feedback processing. Finally, these analyses could capture the differences in the dopaminergic brain responses and associated behavior. They showed how the dopamine dependent regulation of reward learning is affected by certain genes.

This study can further lead to development of drugs and improve dopamine-based treatments in disorders involving dysfunction of the dopamine system. Yet, it should be followed with pharmacological interventions with antagonists of different dopamine receptors and investigating changes in function and behavior. Therefore, multimodal imaging can play an important role. For instance, PET imaging can be used to understand pre- and postsynaptic function for different receptor subtypes and how it is affected by genotype. One should note that in future fMRI studies of midbrain structures, a minimum voxel size of 2 mm is recommended for an optimal spatial resolution.

The fourth chapter presented how reinforcement learning models can be combined with functional connectivity models for a comprehensive analysis of fMRI BOLD data. Dynamic causal modeling can be used for assessment of effective connectivity changes among the regions of a reward network during reinforcement learning. While bilinear state equations allowed us to model intrinsic connections, task-dependent modulation of connections, and the driving inputs, nonlinear terms made it possible to test gating mechanisms in this network. By using a family wise model comparison approach, the model space was partitioned into families, and model comparison was performed with the variational Bayesian approximations of the log model evidence. This procedure revealed that the family with no nonlinear effect explained the interactions best. The maximum a posteriori estimates of the connectivity strengths of the nested models in the winning family were averaged with Bayesian model averaging. Post-hoc statistical tests showed the significant connections during the reinforcement learning task. Finally, it was shown that the differences in connectivity strengths were related with genetic variation.

There are other causal connections that need to be investigated. For example, stimulus value coding and value comparison are important elements of reinforcement learning. Functional coupling of the regions involved in these processes can be analysed with similar models that were presented here. One can identify these regions by integrating Q-values in the general linear model for the fMRI data analysis. Further, since many studies indicate possible correlations between structure and function within dopaminergic pathways, this relationship needs to be addressed in the future e.g. with diffusion tractography data.

In the last study (Chapter 5), hierarchical Bayesian models were used for modeling a learning agent in a volatile environment. Hierarchical Gaussian Filter (HGF) approach was adopted to construct a biologically plausible learner and also to avoid complex integrals in common Bayesian learning algorithms. Learned values of cues were combined with the precision weighting model, which allowed us to incorporate individual estimates of volatility by means of a dynamic temperature parameter in the decision making or response model. Several simulations were performed to evaluate (i) the behavior of simulated agents for different values of the precision weighting parameter, (ii) parameter recovery, i.e whether the estimation procedure can capture differences in the parameters for both cues, (iii) and the influence of volatility on dynamic learning rates for both cues. A real dataset was collected from thirty-six subjects, and Bayesian model comparison was performed on this dataset to determine the best fitting model among the several other possible strategies. This procedure together with balanced accuracy measure showed that participants took into account both cues during decision making and it justified the inclusion of subject specific parameters in the perceptual model, opposite to an optimal Bayesian learner. Small correlations between coupling parameter (κ) and tonic part of the learning rate (ω) of the winning model were reported. Adopting the HGF and the precision weighted response model to learn multiple cues (here social and non-social) allowed us to quantify individual differences, which is important to understand psychiatric disorders. The final part of this study presented how this approach can be used to bring insight into autistic traits. They were associated with the extent to which subjects took into account the social cue (gaze direction), as indicated by correlations between precision weighting of social cue and the autistic traits. Autistic patients have everyday social impairments, but it is not known which sub-personal processes other than basic perceptual mechanisms play role in these impairments. The modeling approach presented here was able to demonstrate trait-related performance differences are not explained by an inability to process the social stimuli and its causes, but rather by the extent to which participants take into account social information during decision-making. Therefore, it is not a misinterpretation of a cue, but down prioritizing it due to the increasing volatility. To make such conclusions were possible by quantifying the behavior at the different levels of the hierarchy: For example, while parameter κ accounts for how much an individual integrates the predicted cue volatility to the predicted cue tendency, the parameter ζ describes to what extend the inferred cue probability is integrated to decision making. One can test different hypothesis by comparing these parameters, which would not be possible without the use of a computational model.

The modeling approach implemented here can be extended by replacing gaze weighting parameter with a dynamic version to incorporate how the changes in the environment dynamics affect behavior in different parts of the experiment. It will be important to follow up with an examination of a clinical sample in future research. Possibly, a sample spanning a larger autism quotient (AQ) range can be used to show the extension of AQ in a broader range. It can be recommended to use a control task with a different probability schedule to avoid potential confounds in the highly winning trials. Also, this behavioral study can be replicated with an fMRI experiment to see the neural representation of these social processes, which is crucial for clinical questions. Finally, this novel paradigm and parallel learning model scenario can be applied in other neurological and psychiatric patient groups where the processing of social information is altered such as schizophrenia or traumatic brain injury (TBI).

As a final note for the reader, the choice of model depends on the environment. We do not claim that either HGF or RL is superior for all clinical studies, but the choice mainly depends on the dynamics of the learning environment. The studies in this thesis can be considered as examples of when to apply which model.

7.2 Contributions

This thesis contributes to the evaluation of computational models of learning and decision making in clinical questions. One of the important contributions of this thesis is that the most commonly used reinforcement learning algorithms in neuroimaging of dopaminergic function were evaluated with simulations. Testing these models on a large real dataset allowed for conclusions to be made about the reliability and validity of using a particular computational model. Also, it was shown that protected exceedance probability measures can be adopted for accounting the variability in Bayesian information criterion in the model comparison step. Further, informing fMRI data analysis with model derived prediction errors provided a useful framework for capturing differences in dopaminergic brain function and associated behavior due to underlying genetics.

Another contribution of this thesis was presenting Dynamic Causal Models (DCMs) as a method for studying connectivity changes in a reinforcement learning network in the brain. This study showed that combining bilinear and nonlinear DCMs allows to test hypotheses related to synaptic plasticity changes in the regions that are involved in feedback and prediction error processing as well as the gating effects via the quadratic terms. From a methodological point, this is the first study to show that combining computational models of learning and dynamic models of brain networks can together capture differences in dopamine system due to genetic polymorphisms and the associated learning performance.

The final and major contribution of this thesis is adopting the Hierarchical Gaussian Filter (HGF) and the precision weighted response model for learning and combining the features of multiple cues in a volatile environment. This methodological novelty was assessed with simulations for parameter recovery or identifiability and the relation between volatility and dynamic learning rates during learning from different sources of information. Data modeling provided that hierarchical Bayesian models can capture and quantify subject specific differences in learning and the integration of different cues in a volatile environment, which are mainly caused by individual traits.

List of Algorithms

1 TD(0) . . . 7
2 Q-Learning . . . 8

List of Figures

1.1 Interaction of the organism with the environment. 3

2.1 Grid World 6
2.2 Reward Circuit 10
2.3 Midbrain DA neurons responses resemble PEs in TD learning. 11
2.4 Separate pathways for positive and negative feedback in basal ganglia. 13
2.5 Genetic dissociations of three genetic polymorphisms on learning rates. 14
2.6 Overview of the three level HGF. 19
2.7 Predictive Coding in the visual cortex. 21
2.8 Interaction of three brain regions manipulated by experimental input. 28
2.9 Bilinear and nonlinear DCMs. 29

3.1 Probabilistic Selection Task (PST). 36
3.2 Parametric regressors. 38
3.3 Simulated action values (Q) with the single learning rate model for different learners. 44
3.4 Simulations for the Q-learning model with two learning rates, α_+ and α_-. . . 45
3.5 Results of the behavioural post test. 46
3.6 Trajectories for action values. 47
3.7 Learned action values at the end of the learning phase. 47
3.8 Brain regions activated at Reward > Punishment contrast. 48
3.9 Brain regions activated at Punishment > Reward contrast. 48
3.10 Brain regions activated at positive prediction error contrast. 49
3.11 Differences of VTA/SN peak activation in response to positive prediction errors with regard to the interaction of *FTO* and *ANKK1* gene variants. 49
3.12 Association of prediction error representaion in midbrain and avoidance learning. 49

4.1 Driving input configurations. 54
4.2 Nonlinear influence of midbrain activation. 55
4.3 An example of nonlinear model. 56
4.4 Bayesian Model Comparison. 57
4.5 Family-wise Model Comparison. 58
4.6 Observed and fitted time series. 59
4.7 Influence of genotype on the brain RL network. 61
4.8 Correlation between connection strength from VTA/SN to NAcc and performance in avoidance learning. 61

5.1 Model Framework 64
5.2 Graphical depiction of two parallel learning systems that were assumed to influence the choice behavior. 65
5.3 The experimental design. 68

5.4 Simulations for an agent with same perceptual but different social cue weighting values. 73
5.5 Simulations for parameter identifiability. 74
5.6 Dynamic learning rates 75
5.7 Bayesian Model Comparison. 76
5.8 Learning trajectories from an example subject. 77
5.9 Correlations among the parameters. 78
5.10 Relationship of total scores, AQ and ζ. 79
5.11 The influence of structure of the environment on the behavior. 80

List of Tables

3.1 Descriptive (mean ± SEM) data of participants, gender, age, Wechsler Adult Intelligence Scale-Matrice Scale, BMI 34

4.1 Random effects analysis on averaged parameters 60

5.1 Descriptive (mean ± SEM) data of participants, gender, age, AQ, systemizing quotient (SQ), empathy quotient (EQ), IQ (verbal). 67
5.2 Priors on model parameters that varied across the models (HGF with volatility (Model 1), HGF reduced card with volatility (Model 2), normative HGF with decision noise (Model 3). 71
5.3 Confusion matrix 71
5.4 Mean and standard deviation of the maximum a posteriori (MAP) estimates of the winning model parameters. (ζ is in log-space). 76

Bibliography

Abler, B., Walter, H., Erk, S., Kammerer, H., and Spitzer, M. (2006). Prediction error as a linear function of reward probability is coded in human nucleus accumbens. *Neuroimage*, 31(2):790–5.

Alderson, H. L., Latimer, M. P., and Winn, P. (2008). A functional dissociation of the anterior and posterior pedunculopontine tegmental nucleus: excitotoxic lesions have differential effects on locomotion and the response to nicotine. *Brain Structure and Function*, 213(1-2):247–253.

Balodis, I. M., Grilo, C. M., Kober, H., Worhunsky, P. D., White, M. A., Stevens, M. C., Pearlson, G. D., and Potenza, M. N. (2014). A pilot study linking reduced fronto–striatal recruitment during reward processing to persistent bingeing following treatment for binge-eating disorder. *International Journal of Eating Disorders*, 47(4):376–384.

Baron-Cohen, S., Wheelwright, S., Skinner, R., Martin, J., and Clubley, E. (2001). The autism-spectrum quotient (aq): Evidence from asperger syndrome/high-functioning autism, malesand females, scientists and mathematicians. *Journal of autism and developmental disorders*, 31(1):5–17.

Bastin, J., Deman, P., David, O., Gueguen, M., Benis, D., Minotti, L., Hoffman, D., Combrisson, E., Kujala, J., Perrone-Bertolotti, M., et al. (2016). Direct recordings from human anterior insula reveal its leading role within the error-monitoring network. *Cerebral Cortex*, page bhv352.

Bayer, H. M. and Glimcher, P. W. (2005). Midbrain dopamine neurons encode a quantitative reward prediction error signal. *Neuron*, 47(1):129–41.

Beal, M. J. and Ghahramani, Z. (2003). Variational inference for bayesian mixture of factor analysers.

Bechara, A., Damasio, H., and Damasio, A. R. (2000). Emotion, decision making and the orbitofrontal cortex. *Cerebral cortex*, 10(3):295–307.

Beck, A., Steer, R., and Brown, G. (1996). Manual for beck depression inventory-ii (bdi-ii). Report, Psychological Corporation.

Behrens, T. E., Hunt, L. T., Woolrich, M. W., and Rushworth, M. F. (2008). Associative learning of social value. *Nature*, 456(7219):245–9.

Behrens, T. E., Woolrich, M. W., Walton, M. E., and Rushworth, M. F. (2007). Learning the value of information in an uncertain world. *Nat Neurosci*, 10(9):1214–21.

Bellman, R. (1957). A markovian decision process. Technical report, DTIC Document.

Berns, G. S., McClure, S. M., Pagnoni, G., and Montague, P. R. (2001). Predictability modulates human brain response to reward. *The Journal of Neuroscience*, 21(8):2793–2798.

Bertsekas, D. P., Bertsekas, D. P., Bertsekas, D. P., and Bertsekas, D. P. (1995). *Dynamic programming and optimal control*, volume 1. Athena Scientific Belmont, MA.

Bromberg-Martin, E. S., Matsumoto, M., and Hikosaka, O. (2010). Dopamine in motivational control: rewarding, aversive, and alerting. *Neuron*, 68(5):815–834.

Brozoski, T. J., Brown, R. M., Rosvold, H., and Goldman, P. S. (1979). Cognitive deficit caused by regional depletion of dopamine in prefrontal cortex of rhesus monkey. *Science*, 205:929–32.

Bunzeck, N. and Düzel, E. (2006). Absolute coding of stimulus novelty in the human substantia nigra/vta. *Neuron*, 51(3):369–379.

Bush, R. R. and Mosteller, F. (1951). A mathematical model for simple learning. *Psychological review*, 58(5):313.

Calabresi, P., Gubellini, P., Centonze, D., Picconi, B., Bernardi, G., Chergui, K., Svenningsson, P., Fienberg, A. A., and Greengard, P. (2000). Dopamine and camp-regulated phosphoprotein 32 kda controls both striatal long-term depression and long-term potentiation, opposing forms of synaptic plasticity. *The Journal of neuroscience*, 20(22):8443–8451.

Camara, E., Rodriguez-Fornells, A., and Münte, T. F. (2009). Functional connectivity of reward processing in the brain. *Frontiers in Human Neuroscience*, 2:19.

Camille, N., Coricelli, G., Sallet, J., Pradat-Diehl, P., Duhamel, J.-R., and Sirigu, A. (2004). The involvement of the orbitofrontal cortex in the experience of regret. *Science*, 304(5674):1167–1170.

Cazé, R. D. and van der Meer, M. A. (2013). Adaptive properties of differential learning rates for positive and negative outcomes. *Biological cybernetics*, 107(6):711–719.

Chase, H. W., Kumar, P., Eickhoff, S. B., and Dombrovski, A. Y. (2015a). Reinforcement learning models and their neural correlates: An activation likelihood estimation meta-analysis. *Cogn Affect Behav Neurosci*, 15(2):435–59.

Chase, H. W., Kumar, P., Eickhoff, S. B., and Dombrovski, A. Y. (2015b). Reinforcement learning models and their neural correlates: An activation likelihood estimation meta-analysis. *Cognitive, affective, & behavioral neuroscience*, 15(2):435–459.

Choudhry, Z., Sengupta, S. M., Grizenko, N., Thakur, G. A., Fortier, M.-E., Schmitz, N., and Joober, R. (2013). Association between obesity-related gene fto and adhd. *Obesity (Silver Spring)*, 21(12):E738–44.

Chowdhury, R., Guitart-Masip, M., Lambert, C., Dayan, P., Huys, Q., Duzel, E., and Dolan, R. J. (2013). Dopamine restores reward prediction errors in old age. *Nat Neurosci*, 16(5):648–53.

Chuang, Y.-F., Tanaka, T., Beason-Held, L. L., An, Y., Terracciano, A., Sutin, A. R., Kraut, M., Singleton, A. B., Resnick, S. M., and Thambisetty, M. (2015). Fto genotype and aging: pleiotropic longitudinal effects on adiposity, brain function, impulsivity and diet. *Mol Psychiatry*, 20(1):140–147.

Cohen, J. Y., Haesler, S., Vong, L., Lowell, B. B., and Uchida, N. (2012). Neuron-type-specific signals for reward and punishment in the ventral tegmental area. *Nature*, 482(7383):85–8.

Cools, A., van den Bercken, J. H., Horstink, M., Van Spaendonck, K., and Berger, H. (1984). Cognitive and motor shifting aptitude disorder in parkinson's disease. *Journal of Neurology, Neurosurgery & Psychiatry*, 47(5):443–453.

D'Ardenne, K., Eshel, N., Luka, J., Lenartowicz, A., Nystrom, L. E., and Cohen, J. D. (2012). Role of prefrontal cortex and the midbrain dopamine system in working memory updating. *Proceedings of the National Academy of Sciences*, 109(49):19900–19909.

D'Ardenne, K., McClure, S. M., Nystrom, L. E., and Cohen, J. D. (2008). Bold responses reflecting dopaminergic signals in the human ventral tegmental area. *Science*, 319(5867):1264–1267.

Daunizeau, J., Den Ouden, H. E., Pessiglione, M., Kiebel, S. J., Friston, K. J., and Stephan, K. E. (2010a). Observing the observer (ii): deciding when to decide. *PLoS one*, 5(12):e15555.

Daunizeau, J., Den Ouden, H. E., Pessiglione, M., Kiebel, S. J., Stephan, K. E., and Friston, K. J. (2010b). Observing the observer (i): meta-bayesian models of learning and decision-making. *PLoS One*, 5(12):e15554.

Dauvermann, M. R., Whalley, H. C., Romaniuk, L., Valton, V., Owens, D. G., Johnstone, E. C., Lawrie, S. M., and Moorhead, T. W. (2013). The application of nonlinear dynamic causal modelling for fmri in subjects at high genetic risk of schizophrenia. *Neuroimage*, 73:16–29.

Daw, N. D. (2011). Trial-by-trial data analysis using computational models. *Decision making, affect, and learning: Attention and performance XXIII*, 23:3–38.

Dayan, P. (1992). The convergence of td (λ) for general λ. *Machine learning*, 8(3-4):341–362.

Delgado, M. R., Nystrom, L. E., Fissell, C., Noll, D., and Fiez, J. A. (2000). Tracking the hemodynamic responses to reward and punishment in the striatum. *Journal of neurophysiology*, 84(6):3072–3077.

den Ouden, H. E., Daunizeau, J., Roiser, J., Friston, K. J., and Stephan, K. E. (2010). Striatal prediction error modulates cortical coupling. *The Journal of neuroscience*, 30(9):3210–3219.

den Ouden, H. E., Friston, K. J., Daw, N. D., McIntosh, A. R., and Stephan, K. E. (2009). A dual role for prediction error in associative learning. *Cerebral Cortex*, 19(5):1175–1185.

Diaconescu, A. O., Mathys, C., Weber, L. A., Daunizeau, J., Kasper, L., Lomakina, E. I., Fehr, E., and Stephan, K. E. (2014). Inferring on the intentions of others by hierarchical bayesian learning. *PLoS Comput Biol*, 10(9):e1003810.

Dina, C., Meyre, D., Gallina, S., Durand, E., Körner, A., Jacobson, P., Carlsson, L. M. S., Kiess, W., Vatin, V., Lecoeur, C., Delplanque, J., Vaillant, E., Pattou, F., Ruiz, J., Weill, J., Levy-Marchal, C., Horber, F., Potoczna, N., Hercberg, S., Le Stunff, C., Bougneres, P., Kovacs, P., Marre, M., Balkau, B., Cauchi, S., Chevre, J.-C., and Froguel, P. (2007). Variation in fto contributes to childhood obesity and severe adult obesity. *Nat Genet*, 39(6):724–726.

Doll, B. B., Bath, K. G., Daw, N. D., and Frank, M. J. (2016). Variability in dopamine genes dissociates model-based and model-free reinforcement learning. *The Journal of Neuroscience*, 36(4):1211–1222.

Doya, K. (2002). Metalearning and neuromodulation. *Neural Networks*, 15(4):495–506.

Düzel, E., Bunzeck, N., Guitart-Masip, M., Wittmann, B., Schott, B. H., and Tobler, P. N. (2009). Functional imaging of the human dopaminergic midbrain. *Trends Neurosci*, 32(6):321–328.

Egner, T., Monti, J. M., and Summerfield, C. (2010). Expectation and surprise determine neural population responses in the ventral visual stream. *J Neurosci*, 30(49):16601–8.

Epstein, L. H., Temple, J. L., Neaderhiser, B. J., Salis, R. J., Erbe, R. W., and Leddy, J. J. (2007). Food reinforcement, the dopamine d2 receptor genotype, and energy intake in obese and nonobese humans. *Behav Neurosci*, 121(5):877–86.

Ernst, M. O. and Banks, M. (2002). Humans integrate visual and haptic information in a statistically optimal fashion. *Nature*.

Ernst, M. O. and Bülthoff, H. H. (2004). Merging the senses into a robust percept. *Trends in cognitive sciences*, 8(4):162–169.

Ewbank, M. P. and Henson, R. N. (2012). Explaining away repetition effects via predictive coding. *Cognitive neuroscience*, 3(3-4):239–240.

Felsted, J. A., Ren, X., Chouinard-Decorte, F., and Small, D. M. (2010). Genetically determined differences in brain response to a primary food reward. *J Neurosci*, 30(7):2428–2432.

Frank, M. J. and Fossella, J. A. (2011). Neurogenetics and pharmacology of learning, motivation, and cognition. *Neuropsychopharmacology*, 36(1):133–152.

Frank, M. J. and Hutchison, K. (2009). Genetic contributions to avoidance-based decisions: striatal d2 receptor polymorphisms. *Neuroscience*, 164(1):131–40.

Frank, M. J., Moustafa, A. A., Haughey, H. M., Curran, T., and Hutchison, K. E. (2007). Genetic triple dissociation reveals multiple roles for dopamine in reinforcement learning. *Proc Natl Acad Sci U S A*, 104(41):16311–6.

Frank, M. J., Seeberger, L. C., and O'Reilly, R. C. (2004). By carrot or by stick: cognitive reinforcement learning in parkinsonism. *Science*, 306(5703):1940–1943.

Frayling, T. M., Timpson, N. J., Weedon, M. N., Zeggini, E., Freathy, R. M., Lindgren, C. M., Perry, J. R. B., Elliott, K. S., Lango, H., Rayner, N. W., Shields, B., Harries, L. W., Barrett, J. C., Ellard, S., Groves, C. J., Knight, B., Patch, A.-M., Ness, A. R., Ebrahim, S., Lawlor, D. A., Ring, S. M., Ben-Shlomo, Y., Jarvelin, M.-R., Sovio, U., Bennett, A. J., Melzer, D., Ferrucci, L., Loos, R. J. F., Barroso, I., Wareham, N. J., Karpe, F., Owen, K. R., Cardon, L. R., Walker, M., Hitman, G. A., Palmer, C. N. A., Doney, A. S. F., Morris, A. D., Smith, G. D., Hattersley, A. T., and McCarthy, M. I. (2007). A common variant in the fto gene is associated with body mass index and predisposes to childhood and adult obesity. *Science*, 316(5826):889–894.

Freeze, B. S., Kravitz, A. V., Hammack, N., Berke, J. D., and Kreitzer, A. C. (2013). Control of basal ganglia output by direct and indirect pathway projection neurons. *The Journal of neuroscience*, 33(47):18531–18539.

Friston, K. (2009). The free-energy principle: a rough guide to the brain? *Trends Cogn Sci*, 13(7):293–301.

Friston, K. (2016). The bayesian savant. *Biological Psychiatry*, 80(2):87–89.

Friston, K., Buechel, C., Fink, G., Morris, J., Rolls, E., and Dolan, R. (1997). Psychophysiological and modulatory interactions in neuroimaging. *Neuroimage*, 6(3):218–229.

Friston, K., Mattout, J., Trujillo-Barreto, N., Ashburner, J., and Penny, W. (2007). Variational free energy and the laplace approximation. *Neuroimage*, 34(1):220–234.

Friston, K. J. (1994). Functional and effective connectivity in neuroimaging: a synthesis. *Human brain mapping*, 2(1-2):56–78.

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

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

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

Friston, K. J., Lawson, R., and Frith, C. D. (2013). On hyperpriors and hypopriors: comment on pellicano and burr. *Trends Cogn. Sci*, 17(1):10–1016.

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

Friston, K. J. and Stephan, K. E. (2007). Free-energy and the brain. *Synthese*, 159(3):417–458.

Friston, K. J., Stephan, K. E., Montague, R., and Dolan, R. J. (2014). Computational psychiatry: the brain as a phantastic organ. *The Lancet Psychiatry*, 1(2):148–158.

Gershman, S. J. (2015). Do learning rates adapt to the distribution of rewards? *Psychon Bull Rev.*

Glaescher, J. (2009). Visualization of group inference data in functional neuroimaging. *Neuroinformatics*, 7(1):73–82.

Gradin, V. B., Kumar, P., Waiter, G., Ahearn, T., Stickle, C., Milders, M., Reid, I., Hall, J., and Steele, J. D. (2011). Expected value and prediction error abnormalities in depression and schizophrenia. *Brain*, 134(6):1751–1764.

Granger, C. W. (1969). Investigating causal relations by econometric models and cross-spectral methods. *Econometrica: Journal of the Econometric Society*, pages 424–438.

Gu, Y., Angelaki, D. E., and DeAngelis, G. C. (2008). Neural correlates of multisensory cue integration in macaque mstd. *Nature neuroscience*, 11(10):1201–1210.

Gu, Y., Deangelis, G. C., and Angelaki, D. E. (2012). Causal links between dorsal medial superior temporal area neurons and multisensory heading perception. *J Neurosci*, 32(7):2299–313.

Gutenkunst, R. N., Waterfall, J. J., Casey, F. P., Brown, K. S., Myers, C. R., and Sethna, J. P. (2007). Universally sloppy parameter sensitivities in systems biology models. *PLoS Comput Biol*, 3(10):1871–1878.

Haber, S. N. (2014). The place of dopamine in the cortico-basal ganglia circuit. *Neuroscience*, 282C:248–257.

Haber, S. N. and Knutson, B. (2010). The reward circuit: linking primate anatomy and human imaging. *Neuropsychopharmacology*, 35(1):4–26.

Hamilton, A. F. d. C. (2013). Reflecting on the mirror neuron system in autism: a systematic review of current theories. *Developmental cognitive neuroscience*, 3:91–105.

Hare, T. A., O'Doherty, J., Camerer, C. F., Schultz, W., and Rangel, A. (2008). Dissociating the role of the orbitofrontal cortex and the striatum in the computation of goal values and prediction errors. *The Journal of Neuroscience*, 28(22):5623–5630.

Hess, M. E. and Brüning, J. C. (2014). The fat mass and obesity-associated (fto) gene: Obesity and beyond? *Biochim Biophys Acta*.

Hess, M. E., Hess, S., Meyer, K. D., Verhagen, L. A. W., Koch, L., Brönneke, H. S., Dietrich, M. O., Jordan, S. D., Saletore, Y., Elemento, O., Belgardt, B. F., Franz, T., Horvath, T. L., Rüther, U., Jaffrey, S. R., Kloppenburg, P., and Brüning, J. C. (2013). The fat mass and obesity associated gene (fto) regulates activity of the dopaminergic midbrain circuitry. *Nat Neurosci*, 16(8):1042–1048.

Hester, R., Nestor, L., and Garavan, H. (2009). Impaired error awareness and anterior cingulate cortex hypoactivity in chronic cannabis users. *Neuropsychopharmacology*, 34(11):2450–2458.

Hikida, T., Kimura, K., Wada, N., Funabiki, K., and Nakanishi, S. (2010). Distinct roles of synaptic transmission in direct and indirect striatal pathways to reward and aversive behavior. *Neuron*, 66(6):896–907.

Honey, C. J., Kotter, R., Breakspear, M., and Sporns, O. (2007). Network structure of cerebral cortex shapes functional connectivity on multiple time scales. *Proc Natl Acad Sci U S A*, 104(24):10240–5.

Horga, G., Maia, T. V., Marsh, R., Hao, X., Xu, D., Duan, Y., Tau, G. Z., Graniello, B., Wang, Z., Kangarlu, A., et al. (2015). Changes in corticostriatal connectivity during reinforcement learning in humans. *Human brain mapping*, 36(2):793–803.

Hunt, L. T., Kolling, N., Soltani, A., Woolrich, M. W., Rushworth, M. F., and Behrens, T. E. (2012). Mechanisms underlying cortical activity during value-guided choice. *Nature neuroscience*, 15(3):470–476.

Ide, J. S., Shenoy, P., Angela, J. Y., and Chiang-shan, R. L. (2013). Bayesian prediction and evaluation in the anterior cingulate cortex. *The Journal of Neuroscience*, 33(5):2039–2047.

Iglesias, S., Mathys, C., Brodersen, K. H., Kasper, L., Piccirelli, M., den Ouden, H. E., and Stephan, K. E. (2013). Hierarchical prediction errors in midbrain and basal forebrain during sensory learning. *Neuron*, 80(2):519–530.

Jocham, G., Klein, T. A., and Ullsperger, M. (2011). Dopamine-mediated reinforcement learning signals in the striatum and ventromedial prefrontal cortex underlie value-based choices. *J Neurosci*, 31(5):1606–1613.

Joensson, E. G., Nöthen, M. M., Grünhage, F., Farde, L., Nakashima, Y., Propping, P., and Sedvall, G. C. (1999). Polymorphisms in the dopamine d2 receptor gene and their relationships to striatal dopamine receptor density of healthy volunteers. *Mol Psychiatry*, 4(3):290–296.

Karra, E., Daly, O. G., Choudhury, A. I., Yousseif, A., Millership, S., Neary, M. T., Scott, W. R., Chandarana, K., Manning, S., Hess, M. E., Iwakura, H., Akamizu, T., Millet, Q., Gelegen, C., Drew, M. E., Rahman, S., Emmanuel, J. J., Williams, S. C. R., Ruether, U. U., Brüning, J. C., Withers, D. J., Zelaya, F. O., and Batterham, R. L. (2013). A link between fto, ghrelin, and impaired brain food-cue responsivity. *J Clinic Invest*, 123(8):3539–3551.

Kenny, P. J. (2011). Reward mechanisms in obesity: new insights and future directions. *Neuron*, 69(4):664–679.

Kiviniemi, V., Kantola, J.-H., Jauhiainen, J., Hyvaerinen, A., and Tervonen, O. (2003). Independent component analysis of nondeterministic fmri signal sources. *NeuroImage*, 19(2):253–260.

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

Knill, D. C. (2007a). Learning bayesian priors for depth perception. *J Vis*, 7(8):13.

Knill, D. C. (2007b). Robust cue integration: a bayesian model and evidence from cue-conflict studies with stereoscopic and figure cues to slant. *J Vis*, 7(7):5 1–24.

Knill, D. C. and Saunders, J. A. (2003). Do humans optimally integrate stereo and texture information for judgments of surface slant? *Vision Research*, 43(24):2539–2558.

Körding, K. P., Beierholm, U., Ma, W. J., Quartz, S., Tenenbaum, J. B., and Shams, L. (2007). Causal inference in multisensory perception. *PLoS One*, 2(9):e943.

Korotkova, T. M., Ponomarenko, A. A., Brown, R. E., and Haas, H. L. (2004). Functional diversity of ventral midbrain dopamine and gabaergic neurons. *Mol. Neurobiol.*, 29(3):243–259.

Kravitz, A. V., Freeze, B. S., Parker, P. R., Kay, K., Thwin, M. T., Deisseroth, K., and Kreitzer, A. C. (2010). Regulation of parkinsonian motor behaviours by optogenetic control of basal ganglia circuitry. *Nature*, 466(7306):622–626.

Landy, M. S., Goutcher, R., Trommershauser, J., and Mamassian, P. (2007). Visual estimation under risk. *J Vis*, 7(6):4.

Lawson, R. P., Rees, G., and Friston, K. J. (2014). An aberrant precision account of autism. *Front Hum Neurosci*, 8.

Lee, M. D. and Wagenmakers, E.-J. (2014). *Bayesian cognitive modeling: A practical course.* Cambridge University Press.

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

Logothetis, N. K. and Wandell, B. A. (2004). Interpreting the bold signal. *Annu. Rev. Physiol.*, 66:735–769.

Lomakina, E. I., Paliwal, S., Diaconescu, A. O., Brodersen, K. H., Aponte, E. A., Buhmann, J. M., and Stephan, K. E. (2015). Inversion of hierarchical bayesian models using gaussian processes. *NeuroImage*, 118:133–145.

Ma, W. J., Beck, J. M., Latham, P. E., and Pouget, A. (2006). Bayesian inference with probabilistic population codes. *Nat Neurosci*, 9(11):1432–8.

Marreiros, A., Kiebel, S. J., and Friston, K. J. (2008). Dynamic causal modelling for fmri: a two-state model. *Neuroimage*, 39(1):269–278.

Mathys, C., Daunizeau, J., Friston, K. J., and Stephan, K. E. (2011). A bayesian foundation for individual learning under uncertainty. *Front Hum Neurosci*, 5:39.

Mathys, C. D., Lomakina, E. I., Daunizeau, J., Iglesias, S., Brodersen, K. H., Friston, K. J., and Stephan, K. E. (2014). Uncertainty in perception and the hierarchical gaussian filter. *Front Hum Neurosci*, 8.

McClure, S. M., Berns, G. S., and Montague, P. R. (2003). Temporal prediction errors in a passive learning task activate human striatum. *Neuron*, 38(2):339–346.

McIntosh, A. R. (1994). Structural equation modeling and its application to network analysis in functional brain imaging. *Hum Brain Mapp.*

Melloni, L., van Leeuwen, S., Alink, A., and Müller, N. G. (2012). Interaction between bottom-up saliency and top-down control: how saliency maps are created in the human brain. *Cerebral Cortex*, 22(12):2943–2952.

Melo, F. S. (2001). Convergence of q-learning: A simple proof. *Institute Of Systems and Robotics, Tech. Rep.*

Meredith, M. and Stein, B. (1983). Interactions among converging sensory inputs in the superior colliculus. *Science*, 221.

Messé, A., Rudrauf, D., Giron, A., and Marrelec, G. (2015). Predicting functional connectivity from structural connectivity via computational models using mri: An extensive comparison study. *NeuroImage*, 111:65–75.

Meyer-Lindenberg, A., Straub, R. E., Lipska, B. K., Verchinski, B. A., Goldberg, T., Callicott, J. H., Egan, M. F., Huffaker, S. S., Mattay, V. S., Kolachana, B., et al. (2007). Genetic evidence implicating darpp-32 in human frontostriatal structure, function, and cognition. *The Journal of clinical investigation*, 117(3):672–682.

Molochnikov, I. and Cohen, D. (2015). Hemispheric differences in the mesostriatal dopaminergic system. *Basal ganglia: physiological, behavioral, and computational studies.*

Montague, P. R., Dolan, R. J., Friston, K. J., and Dayan, P. (2012). Computational psychiatry. *Trends in cognitive sciences*, 16(1):72–80.

Montague, P. R., Hyman, S. E., and Cohen, J. D. (2004). Computational roles for dopamine in behavioural control. *Nature*, 431(7010):760–767. 10.1038/nature03015.

Morales, M. and Root, D. H. (2014). Glutamate neurons within the midbrain dopamine regions. *Neuroscience*, 282C:60–68.

Morgan, M. L., Deangelis, G. C., and Angelaki, D. E. (2008). Multisensory integration in macaque visual cortex depends on cue reliability. *Neuron*, 59(4):662–73.

Neville, M. J., Johnstone, E. C., and Walton, R. T. (2004). Identification and characterization of ankk1: a novel kinase gene closely linked to drd2 on chromosome band 11q23.1. *Hum Mutat*, 23(6):540–545.

Niv, Y. (2009). Reinforcement learning in the brain. *Journal of Mathematical Psychology*, 53(3):139–154.

Niv, Y., Edlund, J. A., Dayan, P., and O'Doherty, J. P. (2012). Neural prediction errors reveal a risk-sensitive reinforcement-learning process in the human brain. *J Neurosci*, 32(2):551–562.

Niv, Y., Joel, D., Meilijson, I., and Ruppin, E. (2002). Evolution of reinforcement learning in uncertain environments: A simple explanation for complex foraging behaviors. *Adaptive Behavior*, 10(1):5–24.

Noble, E. P., Gottschalk, L. A., Fallon, J. H., Ritchie, T. L., and Wu, J. C. (1997). D2 dopamine receptor polymorphism and brain regional glucose metabolism. *Am J Med Genet*, 74(2):162–166.

Noble, E. P., Noble, R. E., Ritchie, T., Syndulko, K., Bohlman, M. C., Noble, L. A., Zhang, Y., Sparkes, R. S., and Grandy, D. K. (1994). D2 dopamine receptor gene and obesity. *Int J Eat Disord*, 15(3):205–17.

O'Doherty, J. P., Dayan, P., Friston, K., Critchley, H., and Dolan, R. J. (2003). Temporal difference models and reward-related learning in the human brain. *Neuron*, 38(2):329–337.

O'Doherty, J. P., Hampton, A., and Kim, H. (2007). Model-based fmri and its application to reward learning and decision making. *Annals of the New York Academy of sciences*, 1104(1):35–53.

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

Ohshiro, T., Angelaki, D. E., and DeAngelis, G. C. (2011). A normalization model of multisensory integration. *Nat Neurosci*, 14(6):775–82.

Onge, J. R. S., Abhari, H., and Floresco, S. B. (2011). Dissociable contributions by prefrontal d1 and d2 receptors to risk-based decision making. *The Journal of Neuroscience*, 31(23):8625–8633.

Oruç, I., Maloney, L. T., and Landy, M. S. (2003). Weighted linear cue combination with possibly correlated error. *Vision Research*, 43(23):2451–2468.

Paliwal, S., Petzschner, F. H., Schmitz, A. K., Tittgemeyer, M., and Stephan, K. E. (2014). A model-based analysis of impulsivity using a slot-machine gambling paradigm. *Front. Hum. Neurosci*, 8(428):10–3389.

Palmer, C. J., Seth, A. K., and Hohwy, J. (2015). The felt presence of other minds: Predictive processing, counterfactual predictions, and mentalising in autism. *Consciousness and cognition*, 36:376–389.

Park, S. Q., Kahnt, T., Talmi, D., Rieskamp, J., Dolan, R. J., and Heekeren, H. R. (2012). Adaptive coding of reward prediction errors is gated by striatal coupling. *Proceedings of the National Academy of Sciences*, 109(11):4285–4289.

Payzan-LeNestour, E. and Bossaerts, P. (2011). Risk, unexpected uncertainty, and estimation uncertainty: Bayesian learning in unstable settings. *PLoS Comput Biol*, 7(1):e1001048.

Pellicano, E. and Burr, D. (2012). When the world becomes 'too real': a bayesian explanation of autistic perception. *Trends in cognitive sciences*, 16(10):504–510.

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

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

Pessiglione, M., Seymour, B., Flandin, G., Dolan, R. J., and Frith, C. D. (2006). Dopamine-dependent prediction errors underpin reward-seeking behaviour in humans. *Nature*, 442(7106):1042–1045.

Pignatelli, M. and Bonci, A. (2015). Role of dopamine neurons in reward and aversion: A synaptic plasticity perspective. *Neuron*, 86(5):1145–1157.

Pohjalainen, T., Rinne, J. O., Nagren, K., Lehikoinen, P., Anttila, K., Syvaelahti, E. K., and Hietala, J. (1998). The a1 allele of the human d2 dopamine receptor gene predicts low d2 receptor availability in healthy volunteers. *Mol Psychiatry*, 3(3):256–260.

Rao, R. P. and Ballard, D. H. (1999). Predictive coding in the visual cortex: a functional interpretation of some extra-classical receptive-field effects. *Nature neuroscience*, 2(1):79–87.

Rauss, K. and Pourtois, G. (2013). What is bottom-up and what is top-down in predictive coding? *Front Psychol*, 4:276.

Rescorla, R. A. and Wagner, A. R. (1972). A theory of pavlovian conditioning: Variations in the effectiveness of reinforcement and nonreinforcement. In *Classical conditioning: Current research and theory.*

Rigoux, L., Stephan, K. E., Friston, K. J., and Daunizeau, J. (2014). Bayesian model selection for group studies - revisited. *NeuroImage*, 84:971–985.

Robic, S., Sonie, S., Fonlupt, P., Henaff, M.-A., Touil, N., Coricelli, G., Mattout, J., and Schmitz, C. (2015). Decision-making in a changing world: A study in autism spectrum disorders. *Journal of autism and developmental disorders*, 45(6):1603–1613.

Robinson, E. B., Koenen, K. C., McCormick, M. C., Munir, K., Hallett, V., Happé, F., Plomin, R., and Ronald, A. (2011). Evidence that autistic traits show the same etiology in the general population and at the quantitative extremes (5%, 2.5%, and 1%). *Archives of General Psychiatry*, 68(11):1113–1121.

Rolls, E. T., McCabe, C., and Redoute, J. (2008). Expected value, reward outcome, and temporal difference error representations in a probabilistic decision task. *Cereb Cortex*, 18(3):652–63.

Rubinov, M. and Sporns, O. (2010). Complex network measures of brain connectivity: uses and interpretations. *Neuroimage*, 52(3):1059–1069.

Samejima, K., Ueda, Y., Doya, K., and Kimura, M. (2005). Representation of action-specific reward values in the striatum. *Science*, 310(5752):1337–1340.

Scheibehenne, B. and Pachur, T. (2015). Using bayesian hierarchical parameter estimation to assess the generalizability of cognitive models of choice. *Psychonomic bulletin & review*, 22(2):391–407.

Schilbach, L. (2014). On the relationship of online and offline social cognition. *Front. Hum. Neurosci*, 8(278):10–3389.

Schilbach, L. (2015). Eye to eye, face to face and brain to brain: novel approaches to study the behavioral dynamics and neural mechanisms of social interactions. *Current Opinion in Behavioral Sciences*, 3:130–135.

Schilbach, L., Timmermans, B., Reddy, V., Costall, A., Bente, G., Schlicht, T., and Vogeley, K. (2013). Toward a second-person neuroscience. *Behavioral and Brain Sciences*, 36(04):393–414.

Schonberg, T., Daw, N. D., Joel, D., and O'Doherty, J. P. (2007). Reinforcement learning signals in the human striatum distinguish learners from nonlearners during reward-based decision making. *J Neurosci*, 27(47):12860–7.

Schultz, W. (2010). Multiple functions of dopamine neurons. *F1000 biology reports*, 2.

Schultz, W., Apicella, P., and Ljungberg, T. (1993). Responses of monkey dopamine neurons to reward and conditioned stimuli during successive steps of learning a delayed response task. *The Journal of Neuroscience*, 13(3):900–913.

Schultz, W., Dayan, P., and Montague, P. R. (1997). A neural substrate of prediction and reward. *Science*, 275(5306):1593–1599.

Schultz, W. and Dickinson, A. (2000). Neuronal coding of prediction errors. *Annual review of neuroscience*, 23(1):473–500.

Schuyler, B., Ollinger, J. M., Oakes, T. R., Johnstone, T., and Davidson, R. J. (2010). Dynamic causal modeling applied to fmri data shows high reliability. *Neuroimage*, 49(1):603–611.

Schwarz, G. (1978). Estimating the dimension of a model. *The Annals of Statistics*, 6(2):461–464.

Shi, L. and Griffiths, T. L. (2009). Neural implementation of hierarchical bayesian inference by importance sampling. In *Advances in neural information processing systems*, pages 1669–1677.

Shi, Z. (2011). *Advanced artificial intelligence*, volume 1. World Scientific.

Shteingart, H., Neiman, T., and Loewenstein, Y. (2013). The role of first impression in operant learning. *Journal of Experimental Psychology: General*, 142(2):476.

Sinha, P., Kjelgaard, M. M., Gandhi, T. K., Tsourides, K., Cardinaux, A. L., Pantazis, D., Diamond, S. P., and Held, R. M. (2014). Autism as a disorder of prediction. *Proceedings of the National Academy of Sciences*, 111(42):15220–15225.

Sobczyk-Kopciol, A., Broda, G., Wojnar, M., Kurjata, P., Jakubczyk, A., Klimkiewicz, A., and Ploski, R. (2011). Inverse association of the obesity predisposing fto rs9939609 genotype with alcohol consumption and risk for alcohol dependence. *Addiction*, 106(4):739–748.

Steffensen, S. C., Svingos, A. L., Pickel, V. M., and Henriksen, S. J. (1998). Electrophysiological characterization of gabaergic neurons in the ventral tegmental area. *J Neurosci*, 18(19):8003–8015.

Stephan, K. E., Kasper, L., Harrison, L. M., Daunizeau, J., den Ouden, H. E. M., Breakspear, M., and Friston, K. J. (2008). Nonlinear dynamic causal models for fmri. *NeuroImage*, 42(2):649–662.

Stephan, K. E. and Mathys, C. (2014). Computational approaches to psychiatry. *Current opinion in neurobiology*, 25:85–92.

Stephan, K. E., Penny, W. D., Daunizeau, J., Moran, R. J., and Friston, K. J. (2009a). Bayesian model selection for group studies. *NeuroImage*, 46(4):1004–1017.

Stephan, K. E., Penny, W. D., Moran, R. J., den Ouden, H. E. M., Daunizeau, J., and Friston, K. J. (2010). Ten simple rules for dynamic causal modeling. *NeuroImage*, 49(4):3099–3109.

Stephan, K. E., Tittgemeyer, M., Knösche, T. R., Moran, R. J., and Friston, K. J. (2009b). Tractography-based priors for dynamic causal models. *Neuroimage*, 47(4):1628–1638.

Stice, E., Spoor, S., Bohon, C., and Small, D. M. (2008). Relation between obesity and blunted striatal response to food is moderated by taqia a1 allele. *Science*, 322(5900):449–452.

Sutton, R. S. (1988). Learning to predict by the methods of temporal differences. *Machine learning*, 3(1):9–44.

Sutton, R. S. (1992). Introduction: The challenge of reinforcement learning. In *Reinforcement Learning*, pages 1–3. Springer.

Sutton, R. S. and Barto, A. G. (1998). *Reinforcement Learning*. Adaptive Computation and Machine Learning. MIT press.

Tesauro, G. (1995). Temporal difference learning and td-gammon. *Communications of the ACM*, 38(3):58–68.

Tsai, H.-C., Zhang, F., Adamantidis, A., Stuber, G. D., Bonci, A., De Lecea, L., and Deisseroth, K. (2009). Phasic firing in dopaminergic neurons is sufficient for behavioral conditioning. *Science*, 324(5930):1080–1084.

Turner, R., Bihan, D. L., Moonen, C. T., Despres, D., and Frank, J. (1991). Echo-planar time course mri of cat brain oxygenation changes. *Magnetic Resonance in Medicine*, 22(1):159–166.

Van de Cruys, S., Evers, K., Van der Hallen, R., Van Eylen, L., Boets, B., de Wit, L., and Wagemans, J. (2014). Precise minds in uncertain worlds: predictive coding in autism. *Psychological Review*, 121(4):649.

van den Bos, W., Cohen, M. X., Kahnt, T., and Crone, E. A. (2012). Striatum–medial prefrontal cortex connectivity predicts developmental changes in reinforcement learning. *Cerebral Cortex*, 22(6):1247–1255.

Vilares, I. and Körding, K. (2011). Bayesian models: the structure of the world, uncertainty, behavior, and the brain. *Annals of the New York Academy of Sciences*, 1224(1):22–39.

Wacongne, C., Changeux, J. P., and Dehaene, S. (2012). A neuronal model of predictive coding accounting for the mismatch negativity. *J Neurosci*, 32(11):3665–78.

Watkins, C. C. H. and Dayan, P. (1992). Technical note: Q-learning. *Machine Learning*, 8(3-4):279–292.

Wiecki, T. V., Poland, J., and Frank, M. J. (2015). Model-based cognitive neuroscience approaches to computational psychiatry clustering and classification. *Clinical Psychological Science*, 3(3):378–399.

Woodbury-Smith, M. R., Robinson, J., Wheelwright, S., and Baron-Cohen, S. (2005). Screening adults for asperger syndrome using the aq: A preliminary study of its diagnostic validity in clinical practice. *Journal of autism and developmental disorders*, 35(3):331–335.

Worsley, K. J. and Friston, K. J. (1995). Analysis of fmri time-series revisited - again. *NeuroImage*, 2(3):173–181.

Yacubian, J., Glascher, J., Schroeder, K., Sommer, T., Braus, D. F., and Buchel, C. (2006). Dissociable systems for gain- and loss-related value predictions and errors of prediction in the human brain. *J Neurosci*, 26(37):9530–7.

Zaghloul, K. A., Blanco, J. A., Weidemann, C. T., McGill, K., Jaggi, J. L., Baltuch, G. H., and Kahana, M. J. (2009). Human substantia nigra neurons encode unexpected financial rewards. *Science*, 323(5920):1496–1499.

Zald, D. H., Boileau, I., El-Dearedy, W., Gunn, R., McGlone, F., Dichter, G. S., and Dagher, A. (2004). Dopamine transmission in the human striatum during monetary reward tasks. *J Neurosci*, 24(17):4105–12.

Erklärung

Ich versichere, dass ich die vorliegende Arbeit ohne unzulässige Hilfe Dritter und ohne Benutzung anderer als der angegebenen Hilfsmittel angefertigt habe. Die aus anderen Quellen direkt oder indirekt übernommenen Daten und Konzepte sind unter Angabe der Quelle gekennzeichnet.

Bei der Auswahl und Auswertung folgenden Materials haben mir die nachstehend aufgeführten Personen in der jeweils beschriebenen Weise unentgeltlich geholfen:

1. Frau Dr. A. Kuehn: funktionelle Magnetresonanztomographie Datenakquise, beschrieben im Kapitel 3.

2. Herr Dr. M. Hess: Genotypisierung, wie im Kapitel 3 beschrieben.

3. Frau Dr. A. O. Diaconescu: Anpassung der MATLAB Skripte zur Erstellung hierarchischer Gaus'schen Filter für mehrere Informationsquellen.

Weitere Personen waren an der inhaltlich-materiellen Erstellung der vorliegenden Arbeit nicht beteiligt. Insbesondere habe ich hierfür nicht die entgeltliche Hilfe von Vermittlungs- bzw. Beratungsdiensten (Promotionsberater oder anderer Personen) in Anspruch genommen. Niemand hat von mir unmittelbar oder mittelbar geldwerte Leistungen für Arbeiten erhalten, die im Zusammenhang mit dem Inhalt der vorgelegten Dissertation stehen.

Die Arbeit wurde bisher weder im In- noch im Ausland in gleicher oder ähnlicher Form einer Prüfungsbehörde vorgelegt.

Ich bin darauf hingewiesen worden, dass die Unrichtigkeit der vorstehenden Erklärung als Täuschungsversuch bewertet wird und gemäß § 7 Abs. 10 der Promotionsordnung den Abbruch des Promotionsverfahrens zur Folge hat.

Ilmenau, den 26. 09. 2016 Meltem Sevgi