Neural correlates of prediction errors during reward and punishment learning

Kamil Bonna

Supervisor: Prof. dr hab. Włodzisław Duch

Faculty of Physics, Astronomy and Informatics
Nicolaus Copernicus University in Toruń

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy

January 2022

Declaration

I hereby declare that except for a specific reference made to the work of others, the contents of this dissertation are original and have not been submitted in whole or in part for consideration for any other degree or qualification at this or any other university. This dissertation is my own work and and does not contain anything which would be the outcome of work done in collaboration with others, except for what is specified in the text and the Acknowledgements.

Kamil Bonna
January 2022

Acknowledgements

I want to thank my supervisor, prof. Włodzisław Duch, for his wise advice and support during my Ph.D. studies and research. I would like to express my deepest appreciation to my auxiliary supervisor and a longtime friend, dr Karolina Finc, for endless scientific discussions, loyalty to our ideals, and encouragement to pursue my goals. I would like to extend my deepest gratitude to my mentor and collaborator, Oliver Hulme, who created an excellent thinking environment and reignited my passion for science.

I also had the great pleasure of working with dr Jaromir Patyk, who always gave his technical support and advice during my research. I thank prof. Jerzy Łukaszewicz and dr Ewelina Wilska for sharing advanced research equipment for my experiment and giving me a chance to conduct my research at the Centre for Modern Interdisciplinary Technologies in Toruń.

My success would not have been possible without the support and the brilliant mind of my friend Bartosz Migała, who always upholds my skepticism. I also wish to thank my friends, Szymon Lesiński, Marcin Karpiński, Mateusz Grochowski, Bartłomiej Maciorowski, Bożena Pięta, Karolina Finc, Miriam Kosik, Michał Komorowski, Michał Kujawski, and Patryk Kujawski. I thank my family, especially my amazing mom Beata Bonna, who has supported me my entire life, my dad Marek Bonna, my grandparents Teresa Jąkała, Zdzisław Jąkała, and Maria Bonna. Finally, I want to thank my love and my partner, Joanna Wójcik, for always cheering me up, taking care of me, and for her mental support during difficult moments in my life.

My sincere appreciation also goes to all people who supported me during my studies and research and helped me on various occasions.

I am also grateful to institutions and organizations that funded my research and other scientific activities throughout my studies, including the National Science Centre (2020/36/T/HS6/00104), National Agency for Academic Exchange, Faculty of Physics, Astronomy and Informatics at Nicolaus Copernicus University (1034-F), Aleksander Jabłoński Foundation, and Centre for Modern Interdisciplinary Technologies at Nicolaus Copernicus University in Toruń.

Abstract

Learning from trial-and-error is facilitated by prediction errors – signals reflecting discrepancy between expected and experienced results of our decisions. Positive prediction errors promote approach behaviors, while negative prediction errors lead to avoidance. One of the most influential findings of modern neuroscience was the discovery of prediction error coding in dopaminergic neurons. Using functional magnetic resonance imaging (fMRI), we can identify brain regions broadcasting prediction errors in various learning scenarios. In this thesis, I provide a holistic description of prediction error correlates in the reward-seeking and punishment-avoiding contexts. To elicit and investigate prediction errors, I used a probabilistic reversal learning task and scanned a group of healthy subjects using fMRI. I merged three complementary perspectives – behavioral, localization, and network – to comprehensively characterize the brain's implementation of reinforcement learning. On the behavioral level, I found that learning speed depends only on the sign of the prediction error and not on the experimental context. In line with the dual system hypothesis, activation analysis localized two independent sets of brain regions signaling positive-going and negative-going prediction errors. Whole-brain network analysis revealed a multi-scale community structure with a separate striatal reward network emerging at a finer topological scale and a ventromedial prefrontal network emerging at a coarser topological scale. I also found that the integration between large-scale networks increased when switching from positive to negative prediction error processing. The pattern of large-scale network reconfiguration followed the predictions of the Global Workspace hypothesis.

Streszczenie

Uczenie się metodą prób i błędów jest możliwe dzięki błędom predykcji – sygnałom odzwierciedlającym różnicę pomiędzy oczekiwanym a rzeczywistym efektem naszych decyzji. Dodatnie błędy predykcji promują powtarzanie wykonanych czynności, podczas gdy ujemne błędy predykcji wytwarzają zachowania unikające. Jednym z najważniejszych sukcesów współczesnej neuronauki było odkrycie kodowania błędów predykcji w neuronach dopaminergicznych. Przy użyciu funkcjonalnego rezonansu magnetycznego (fMRI), możemy zidentyfikować obszary przetwarzające błędy predykcji w różnych sytuacjach decyzyjnych. W mojej pracy przedstawiam holistyczny opis korelatów błędów predykcji w kontekście kar i nagród. Do wywołania i analizy błędów predykcji przeprowadziłem badanie fMRI z udziałem zdrowych osób podczas wykonywania zadania wykorzystującego uczenie probabilistyczne. Połączyłem trzy uzupełniające się podejścia – behawioralne, lokalizacyjne i sieciowe – aby wyczerpująco scharakteryzować implementację uczenia ze wzmocnieniem w ludzkim mózgu. Analiza behawioralna pokazała, że tempo uczenia zależy jedynie od znaku błędu predykcji i jest niezmienne ze względu na kontekst nagrody i kary. Zgodnie z założeniami hipotezy dwóch systemów, analiza aktywacji pozwoliła mi zlokalizować dwa niezależne systemy w mózgu, które odpowiadają za przetwarzanie dodatnich i ujemnych błędów predykcji. Analiza sieci funkcjonalnych ujawniła wieloskalową strukturę modularną z osobną siecią nagrody w mniejszej skali topologicznej i małą siecią brzuszo-przyśrodkową kory przedczołowej w większej skali topologicznej. Integracja pomiędzy podsieciami funkcjonalnymi zwiększyła się podczas przetwarzania ujemnych błędów predykcji. Wzorzec rekonfiguracji podsieci okazał się zgodny z przewidywaniami teorii Globalnej Przestrzeni Roboczej.

Table of contents

List of figures

List of tables

Nomenclature

Acronyms

+PE Positive/Positive-going Prediction Error

−PE Negative/Negative-going Prediction Error

ACC Anterior Cingulate Cortex

ALE Activation Likelihood Estimation

BOLD Blood-Oxygenation-Level-Dependent

BSC Beta-Series Correlation

CD/CI Condition Dependent (Model Family)

CR Conditioned Response

CS Conditioned Stimulus

CSF Cerebrospinal Fluid

DMN Default Mode Network

FC Functional Connectivity

FD Framewise Displacement

FDR Fale Discovery Rate

FID Free Induction Decay

fMRI Functional Magnetic Resonance Imaging

FPN Fronto-Parietal Network

GLM General Linear Model

GWT Global Workspace Theory

HLM Hierarchical Latent-Mixture (Model)

HRF Hemodynamic Response Function

LSN Large-Scale Netowkr

MCMC Markov Chain Monte Carlo

MDP Markov Decision Process

MNI Montreal Neurological Institute and Hospital (Coordinate System)

MRI Magnetic Resonance Imaging

PD/PI Prediction-error Dependent/Independent (Model Family)

PE Prediction Error

PPI Psychophysiological Interaction

PRL Probabilistic Reversal Learning

RF Radiofrequency

ROI Region Of Interest

RPEH Reward Prediction Error Hypothesis of Dopamine

RT Reaction Time

RW Rescorla-Wagner (Model)

SE Spin Echo

TD Temporal-Difference

US Unconditioned Stimulus

vmPFC Ventromedial Prefrontal Cortex

VS Ventral Striatum

WM White Matter

Introduction

Human behavior is driven by rewards and punishments. On the most fundamental level, the goal of our behavior is to maximize rewards and minimize punishments. This goal is facilitated by our ability to tune our behavior by learning the associations between actions and consequences. Learning these associations can take different forms. One of the most essential is reinforcement learning, in which rewarding events cause behavioral reinforcement, whereas punishing events have an opposite effect. Machine learning research demonstrated that the critical part of reinforcement learning is the *prediction error signal* (Sutton and Barto, 2018). Prediction errors reflect the difference between the expected and experienced outcomes of our actions. Years of research showed that the brain implements prediction error coding as a modulation of the phasic activity of dopaminergic neurons (Colombo, 2014; Montague et al., 1996).

The prediction error sign distinguishes between better-than-expected and worse-than-expected outcomes and instructs an agent to reinforce or eliminate a particular action. From the computational perspective, positive and negative prediction errors arise from the exact underlying computation that considers reward and value functions. These functions usually map both rewarding and punishing events onto a common scale ranging from negative to positive. This observation suggests that a single system could be sufficient for signaling both types of prediction errors. However, the human brain is a physical system constrained by the laws of physiology. These constraints cast doubt upon the idea of a single neural circuit broadcasting full prediction error (Palminteri and Pessiglione, 2017). Researchers suggested that an opponent system carrying a negative portion of the signal should exist. In line with this assumption, multiple studies reported brain areas involved in negative prediction error processing outside the dopaminergic system (Fazeli and Büchel, 2018; Hauser et al., 2015; Yacubian et al., 2006). However, it is still unknown whether the opponent system can be observed on a functional network level.

Positive prediction errors usually reinforce actions that lead to rewarding outcomes. Yet, they can also promote behavior that avoids punishment in adverse environments

(Nieuwenhuis et al., 2005; Palminteri et al., 2015). An example of this type of prediction error is a relief from successful avoidance in anxiety disorder. Similarly, negative prediction errors can also be experienced when an anticipated reward is omitted in a reward-rich environment. For example, we might feel bad when our colleague gets a higher pay raise than us, despite the experience of getting raise being generally pleasant. These effects can be simply explained by the relative context-dependent encoding of value (Bavard et al., 2018; Rangel and Clithero, 2012). In this encoding type, an agent continuously adjusts the reference point to which experienced outcomes are compared. These considerations raise an interesting question of whether both learning systems signaling positive and negative prediction errors are invariant to the outcome valence.

In recent years, we observed an emergence of a new interdisciplinary field of *network neuroscience*. Research within this field demonstrated that functional brain networks constantly reorganize to meet the demands of various cognitive tasks (Bullmore and Sporns, 2009). Describing and understanding network dynamics during cognition and linking it with behavior is a fundamental challenge of cognitive neuroscience (Deco et al., 2015). Despite numerous studies on neural correlates of prediction errors, the question of network reconfiguration during prediction error processing is still open. Up to this day, only few studies investigated functional connectivity of specific brain areas like the ventral striatum or amygdala during reward and punishment processing. Moreover, these studies yielded inconsistent answers to whether regions encoding positive and negative prediction errors share similar or different connectivity profiles during prediction error processing.

This thesis comprehensively describes neural correlates of prediction errors on behavioral, activation, and connectivity levels. My goal is to fill the gaps in our understanding of the reinforcement learning mechanisms implemented in the brain. Throughout the thesis, I emphasize the network neuroscience perspective since it provides a complementary description of brain function during cognition and opens a new exciting area of scientific inquiry. I explain the scientific background of modern neuroscience of decision making, describing formal computational models of learning and advanced neuroimaging techniques. I also present the results of my fMRI study on probabilistic reversal learning in two opposite outcome environments – reward-seeking and punishment-avoiding.

In **Chapter 1**, I explain essential mathematical tool and algorithm of reinforcement learning – *Markov decision process* and *temporal-difference learning*. I explain the link between the formal mathematical model of learning and the dopaminergic system in

the mammalian brain. I also discuss theoretical problems associated with punishment-avoidance learning.

In **Chapter 2**, I describe the method of functional magnetic resonance imaging. I review existing findings on neural correlates of prediction errors in the human brain, separately discussing activity and connectivity studies. I accentuate results significant for the debate on punishment-avoidance learning.

In **Chapter 3**, I introduce the network theory – a mathematical framework for describing and investigating real-world networks. I explain the concept of modularity – the tendency of network nodes to organize into communities. I provide evidence that modern network neuroscience can expand our understanding of various cognitive functions.

In **Chapter 4**, I present the methods and results of my fMRI study. I discuss findings in light of three hypotheses – dual system, reference point, and Global Workspace.

I believe that this thesis provides a comprehensive picture of the brain's implementation of learning from trial-and-error. My contribution may expand our understanding of the relationship between reinforcement learning and brain networks. It may also be an important voice in the debate on punishment-avoidance learning and the brain's implementation of negative prediction errors. This knowledge is vital for a better understanding of addiction and learning impairments in many mental disorders. Moreover, it may inspire new developments in artificial learning research.

The study presented in this thesis was supported by:

- ETIUDA scholarship (2020/36/T/HS6/00104) for Ph.D. candidates, funded by the National Science Centre, Poland.
- PROM scholarship (edition 2019/2020), funded by the National Agency for Academic Exchange, Poland
- SUPPORTING GRANT for young scientists (1034-F), funded by the Faculty of Physics, Astronomy and Informatics, Nicolaus Copernicus University in Toruń, Poland

Chapter 1

Reinforcement learning

One of the most critical features of intelligence is the ability to learn. An intelligent agent can learn by interacting with its environment and observing the relationship between causes and effects. For example, a novice chess player may play random moves at first and start noticing basic tactical patterns, enabling him to outplay opponents as he gains more experience. This way, he can learn without an explicit teacher or guideline. This type of learning, often called *reinforcement learning* or *learning by trial and error*, is undeniably the most prevalent source of information about the surrounding environment. Studying basic learning principles allows us to understand how the brain interacts with the environment to acquire new skills and optimize behavior. Also, advancements in reinforcement learning allow artificial intelligence researchers to build more effective learning systems.

1.1 The behaviorist perspective on learning

The quest for understanding animal and human learning originated in behavioral psychology at the beginning of the XXth century. Behaviorists assumed that every behavior is learned from the environment as a result of stimulus-response associations. One of the founders of behaviorism, John B. Watson, stated that the purpose of psychology is "To predict, given the stimulus, what reaction will take place; or, given the reaction, state what the situation or stimulus is that has caused the reaction" (Watson, 1930).

In the 1890s, Russian physiologist Ivan Pavlov studied innate reflexes in dogs. Pavlov measured salivary reflex in dogs exposed to the neutral stimuli – a metronome sound – followed by the inborn triggering stimuli – food. He noticed that after several trials, dogs started to salivate to the metronome sound itself, indicating they were

able to learn the association between consistently paired stimuli (Pavlov, 1928). In behaviorism's terminology, dogs learned to elicit a conditioned response (CR) to the initially neutral stimulus that became conditioned stimulus (CS) after reliable pairing with natural triggering stimulus called unconditioned stimulus (US). This form of learning is called *classical* or *Pavlovian conditioning*. Classical conditioning is the mechanism of learning about predictive relationships between stimuli and anticipating events essential for the organism's survival.

Further investigation of the classical conditioning phenomenon led to discovering additional properties of this form of learning: *blocking* and *higher-order conditioning*. Blocking is the inability to learn new CR when a potential CS is presented together with another CS that has already been paired with US (Kamin, 1969). In other words, Pavlov dogs initially conditioned to salivate on the metronome sound would fail to learn the association between other neutral stimuli, e.g., a flash of light, and food, if the flash would be presented along with the sound. Learning the light would have been blocked by the initially learned association. Higher-order conditioning is the ability for CS to act as US conditioning other initially neutral stimuli. For example, Pavlov described an experiment where a dog was first conditioned to salivate to the sound of the metronome, and then another neutral stimuli – black square – was paired with the sound of the metronome but without following food reward. After few trials, the dog began to salivate upon seeing the black square despite the lack of inherently rewarding stimulus in all trials in which the black square occurred. This phenomenon is called second-order conditioning.

The discovery of various properties of classical conditioning posed a problem of theoretical explanation accounting for observed data. Search for the mathematical model of classical conditioning resulted in early behaviorist theories that later inspired the research in reinforcement learning. In 1972, Robert A. Rescorla and Allan R. Wagner created an influential mathematical model of classical conditioning (Rescorla and Wagner, 1972). *The Rescorla-Wagner model* (RW) assumed compound CS consisting of two components, A and X (e.g., the metronome sound and the black square). The model introduced the associative strength of each component stimulus – a single number representing how reliably that component can predict a rewarding event. According to the RW model, the animal adjusts the associative strength of component A (V_A) and X (V_A) according to:

$$\begin{aligned} \Delta V_A &= \alpha_A \beta (\lambda - V_{AX}) \\ \Delta V_X &= \alpha_X \beta (\lambda - V_{AX}) \,, \end{aligned} \tag{1.1}$$

where α_A and α_X are the salience parameters of components A and X, β is the rate parameter for the rewarding stimuli, λ is the maximum associative strength that the US can support, and $V_{AX} = V_A + V_X$ is the aggregate associative strength for the entire compound stimulus. The RW model introduced a novel assumption that animals learn only in situations in which they are surprised – i.e., the term $\lambda - V_{AX}$ is nonzero. The idea that learning relies on the difference between the reward that an organism experiences and the reward it expects based on past events, turned out to be key to understand basic principles of learning and decision-making (Bush and Mosteller, 1951).

This RW model provided an elegant explanation for blocking – after the initial phase of pairing stimulus A with the US, associative strength V_A reaches an asymptotic value of λ which results in $\lambda - V_{AX} = 0$ even after introducing stimulus X. This is equivalent to the perfect prediction of the future state, implying $\Delta V_X = 0$, reflecting blocking any potential changes in associative strength of stimulus X.

The RW model provided a quantitative explanation of some classical conditioning phenomena and motivated researchers to search for better mathematical models of learning. However, the RW model had its limitations. Being a trial-level model – a model ignoring any possible between-trial interactions – the RW model failed to explain higher-order conditioning. In the following section, I introduce the Markov decision process – the mathematical framework of the temporal-difference model of learning. The temporal-difference model can be viewed as an extension of the RW model with timing effects between stimuli.

1.2 Markov decision process

Studying reinforcement learning models requires one to formally define the problem that learning agents are trying to solve. In other words, to model learning and decision-making, we first need to provide a model of the interactions between the agent and his environment. This model should be simple enough to be manageable for precise mathematical analysis while still providing enough flexibility to cover many real-world learning setups. The Markov decision process (MDP) fulfills both of these criteria. MDP is a mathematically idealized model of sequential decision-making in which an agent's actions can influence immediate and delayed consequences (Sutton and Barto, 2018). This property introduces the temporal dependency between the events faced by the agent that needs to be taken into account while modeling animal and human learning.

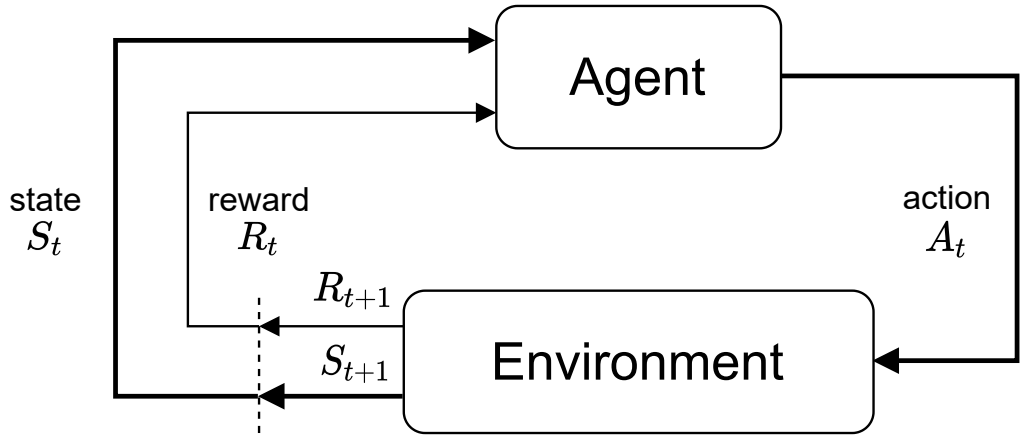

Fig. 1.1 System described by the Markov decision process. At each timestep, an agent is observing the state of the environment S_t and selects an action A_t which result in numerical reward R_{t+1} and the transition of the environment to the new state S_{t+1}.

The MDP is a chain of interactions between the agent and the environment (**Fig. 1.1**). The agent is a learner, i.e., a conditioned dog, a novice chess player deciding which move to play, or an automatic cleaning robot vacuuming the carpet when no one is home. The environment represents everything outside of the agent that is relevant to the modeled problem. The MDP consist of a sequence of discrete timesteps $t = 0, 1, 2, \ldots$. At each timestep, the agent observes the current state of the environment $S_t \in \mathcal{S}$ and chooses an action $A_t \in \mathcal{A}$. The action selected by the agent results in a numerical reward, $R_{t+1} \in \mathbb{R}$, and the transition to the next state S_{t+1}. Then the cycle of state, action, and reward repeats. Here I will only consider a finite MDP characterized by finite sets of states $\mathcal{S}$ and actions $\mathcal{A}$. The most important property of the MDP is, as the name implies, the Markov property:

$$\mathbb{P}\left[S_{t+1} \mid S_t\right] = \mathbb{P}\left[S_{t+1} \mid S_1, \ldots, S_t\right] , \tag{1.2}$$

in which $\mathbb{P}\left[S_{t+1} \mid S_t\right]$ is the probability of the environment transitioning into the new state $s = S_{t+1}$. The Markov property indicates that the process of agent-environment interactions is memoryless, i.e., the new state of the system is independent of the entire system history, and only the current state influences the transition probability. The Markov property is the simplification required to ensure the mathematical tractability of the decision process.

The full information about the MDP can be described by two functions: the *state-transition probability function*, and the *reward function*. The state-transition

probability function is defined as conditional probability that environment transitions to the state s' given that the agent chooses the action a in the state s:

$$p(s' \mid s, a) = \mathbb{P}\left[S_{t+1} = s' \mid S_t = s, A_t = a\right] . \tag{1.3}$$

The state-transition probability function completely describes how evolution of the state can be influenced by the agents actions. The reward function describes the expected immediate reward after choosing the action a in the state s:

$$r(s, a) = \mathbb{E}\left[R_{t+1} \mid S_t = s, A_t = a\right] . \tag{1.4}$$

The MDP framework is a highly flexible abstraction that allows researchers to model various decision-making and learning scenarios. For example, we might use it to describe a game of chess. In this case, the state would be equivalent to the current position of all pieces; the action would be the move selected by the player, and the reward signal would be one if the game is won or zero otherwise. It is important to note that this particular mapping of the real-world situation to abstract mathematical objects is not unique in any way. It might be useful for a computer scientist trying to develop a new chess engine but impractical for a psychologist trying to understand how psychological factors can impact a chess player's performance. For a psychologist, it might be appropriate to include psychological factors in the state representation to take into account the fact that the player's mental state can influence his decisions, hence affecting the next state. Regardless of the particular representation of the real-world problem, the main idea of the MDP stays the same – there are three discrete signals: state signal representing the situation the agent is facing, action signal representing the choices made by the agent, and the reward signal expressing the agent's goal.

1.3 Temporal-difference learning

A precise mathematical formulation of the sequential decision-making problem in the form of the MDP allows one to ask the following question: what is an effective and biologically plausible way to solve the MDP? We might first ask what properties should characterize a potential solution. First, it should take into account temporal dependencies between states and actions. In many real-world scenarios, agents trying to maximize long-term rewards must ignore the immediate consequences of their actions. For example, a chess player may sacrifice his most powerful piece to begin a devastating attack on the enemy king. This behavior is possible only if the agent

appreciates complex trajectories in the environment's state space. Second, the ideal model should offer a generalization of the previously successful model – the RW model in this case. The new model should explain some phenomena that previous models failed, i.e., higher-order conditioning, while still maintaining the predictive power that the previous model offered. Third, the model needs to be biologically plausible – it should be possible to map between the elements of the model and some biological process.

The first step to find the desired solution is to state the agent's goal precisely. In other words, we need to answer the question, What does it mean to solve the MDP – using rigorous mathematical terms? Solving the MDP can be broken down into two distinct components: prediction and control. Prediction involves evaluating the future rewards given some fixed strategy of an agent, whereas control focuses on tuning the agent's strategy to maximize the future rewards. To simplify, I will focus on the prediction component of the complete reinforcement learning problem. All classical conditioning experiments involve learning but not decisions so that they can be modeled as prediction processes. Formally, the prediction problem can be studied within the Markov reward process framework – the MDP without actions. The Markov reward process can arise if we combine the MDP with some fixed strategy of an agent, called *policy*, which is a probabilistic mapping from states to actions:

$$\pi(a \mid s) = \mathbb{P}\left[A_t = a \mid S_t = s\right] . \tag{1.5}$$

The inclusion of the policy makes the state-transition probability function and the reward function independent of actions that are now controlled by the probabilities $\pi(a \mid s)$. A learning agent's goal, assuming a fixed policy, is to estimate the expected value of future rewards:

$$v(s) = \mathbb{E}\left[R_{t+1} + \gamma R_{t+2} + \gamma^2 R_{t+2} + \ldots \mid S_t = s\right] , \tag{1.6}$$

where $v(s)$ is called the *value function*, and $\gamma \in [0, 1]$ is the $discount factor$. The discount factor should be included for several reasons. The first reason is purely mathematical – in cyclic MDPs, the sum of undiscounted rewards $R_{t+1} + R_{t+2} + \ldots$ can be infinite. Second, discounting essentially neglects the distant future, which can account for the inherent uncertainty related to long predictions. Third, it is well established that animals and humans show a strong preference for immediate rewards (Vanderveldt et al., 2016). The value function captures important information for the decision-maker – it quantifies how good a certain state is. If the value function

is accurate, an agent can effectively make the best decisions by greedily picking an action that moves the environment to the state with higher value. For example, in the game of chess, the value function can be used to evaluate a chess position – e.g., if the enemy king is a few moves away from being check-mated, the value function of the agent's position will be high (close to one if we assume that winning leads to a reward $R_T = 1$). Successful evaluation of the position enables an agent to play moves that increase the value function. A fundamental property of the value function is that it can be recursively expressed as an expectation of the immediate reward plus discounted value of the next state:

$$\begin{aligned} v(s) &= \mathbb{E}\left[R_{t+1} + \gamma R_{t+2} + \gamma^2 R_{t+3} + \dots \mid S_t = s\right] = \\ &= \mathbb{E}\left[R_{t+1} + \gamma(R_{t+2} + \gamma R_{t+3} + \dots) \mid S_t = s\right] = \\ &= \mathbb{E}\left[R_{t+1} + \gamma v(S_{t+1}) \mid S_t = s\right] = \\ &= r(s) + \gamma \mathbb{E}\left[v(S_{t+1}) \mid S_t = s\right] , \end{aligned} \tag{1.7}$$

where $r(s) = \mathbb{E}\left[R_{t+1} \mid S_t = s\right]$ is the reward function for the Markov reward process. This formula is known as the *Bellman equation*, and it is fundamental for reinforcement learning (Sammut, 2010). It allows reducing the possibly infinite sum $R_{t+1} + \gamma R_{t+2} + \gamma^2 R_{t+3} + \dots$ into the simple expression reflecting the relationship between the value function in two consecutive states. The Bellman equation expresses practically useful intuition – an agent can easily confirm that he learned the correct value function only if an estimated value of any given state equals the sum of the immediate reward plus the discounted value of the successor state. This intuition provides a basis for a learning rule for estimating value function from experience. Let's call an agent's estimate of the value function $V(s)$ to distinguish it from the true value function $v(s)$. The goal of learner is to find $V(s) = v(s) \quad \forall s \in \mathcal{S}$. The Bellman equation suggests that trial-wise learning can be achieved by estimating the discrepancy between right and left-hand side of (1.7) and using this value as a correction signal to nudge $V(s)$ towards $v(s)$ (Sutton, 1988). This correction signal is called the *temporal-difference* (TD) *prediction error* (PE) and it can be calculated as:

$$\delta_t = R_{t+1} + \gamma V(S_{t+1}) - V(S_t) , \tag{1.8}$$

where R_{t+1} is the experienced reward at time step $t + 1$. Note that the agent uses a piece of experience received from the environement in a form of sample reward R_{t+1} to update his own estimate of the value function. After observing the cycle of state,

reward and succesor state, agent can compute δ_t and use it to update the estimated value for the state S_t:

$$V(S_t) \leftarrow V(S_t) + \alpha\delta_t\,, \tag{1.9}$$

where $\alpha \in [0, 1]$ is the *learning rate* parameter controlling the rate of changes of $V(s)$. The sign of δ_t carries important information for the learner – it reflects whether the experienced reward was higher or lower than the anticipated reward. If the experience turned out to be better than expected, the value function was understated, and it needs to be increased; otherwise, the value function was overestimated, and it needs to be decreased.

The temporal-difference prediction error has a strikingly similar structure to the Rescorla-Wagner learning rule. We can change the notation used in the RW model to match the MDP framework and reexpress (1.1) as:

$$\delta_t = R_{t+1} - V(S_t), \tag{1.10}$$

where λ is now recognized as a reward signal R_{t+1}, associative strength V_{AX} is the value function in the state represented by stimuli A and X and the change in associative strength ΔV is now the update term for the value function. Now it is visible that the only difference between the RW error and the TD error is the addition of the term $\gamma V(S_{t+1})$ in the latter. This difference reflects the essence of both models: the RW model aims to maximize only the immediate reward R_{t+1} neglecting the possibility of higher delayed rewards resulting from temporal dependencies between states, whereas the TD model aims to maximize the sum of both immediate and delayed rewards. It turns out that the TD model offers an elegant explanation of high-order conditioning – a property of classical conditioning that the RW model failed to explain (Seymour et al., 2004).

In this section, I showed that the TD model offers a valuable solution to the MDP problem because it considers temporal dependencies between states, allowing an agent to maximize both immediate and delayed rewards. Moreover, the TD model can be viewed as a generalization of the RW model, providing a concise explanation for the high-order conditioning phenomenon. In the next section, I will describe a striking correspondence between the TD model and the dopaminergic system in the brain. This correspondence serves as the biological foundation of the TD model and its relevance to the explanation of animal and human learning.

1.4 Dopamine reward prediction error hypothesis

Dopamine neurons in the brain reside primarily in small nuclei in the midbrain and the brainstem from which they send widespread connections to the striatum, amygdala, hippocampu,s and frontal cortex (**Fig. 1.2A**). What is the function of these neurons? Years of research in this area converged on the two processes related to dopamine – movement control (Cenci, 2007) and appetitive behavior and reward processing (Wise, 1982). Dopamine neurons have been famously called the brain's *reward system* because of their reinforcing capabilities observed in the direct brain self-stimulation experiments in rats and various drug addictions (Wise, 1996). However, the reward system hypothesis was recently refined by observing an astounding connection between the phasic activity of the dopamine neurons and the temporal-difference prediction error.

In the 1980s and 1990s, neurobiologist Wolfram Schultz conducted a series of classical conditioning experiments with macaque monkeys. He recorded phasic responses of monkey's dopaminergic neurons related to the administration of the reward (droplets of the grape juice) in three types of trials: unexpected reward trials, expected reward trials, and expected reward trials with the omission of the reward (Schultz, 1986). In the expected reward trials, about 1s before reward administration, monkeys received a cue acting as a conditioned stimulus predicting reward. During the unexpected reward trials, dopaminergic neurons responded to the appearance of the reward, whereas during the expected reward trials, they responded shortly after the appearance of the cue. Moreover, in the expected reward trials with the omission of the reward, when monkeys noticed that the expected reward is missing, their dopaminergic neurons decreased firing rate below the baseline (**Fig. 1.2B**). This data suggested that dopaminergic neurons do not respond to the reward itself, but rather their response to the reward is modulated by the animal predictions about the reward. This puzzling pattern of dopaminergic neurons activity was initially explained in terms of "attentional and motivational processes underlying learning and cognitive behavior" (Schultz et al., 1993). However, this explanation was not satisfactory or convincing since it was not grounded in any significant learning or decision-making theory.

In the early 1990s, the dopaminergic neurons response pattern was recognized as the TD prediction error (Montague et al., 1993; Montague, 2007). To better understand the connection between TD prediction error and dopamine, it is worth looking at Schultz's experiment from the reinforcement learning perspective. Let us assume that the droplet of juice acts as a reward signal $R = 1$ and that the experiment is modeled as the Markov reward process without discounting, i.e., with $\gamma = 1$. The experiment can be

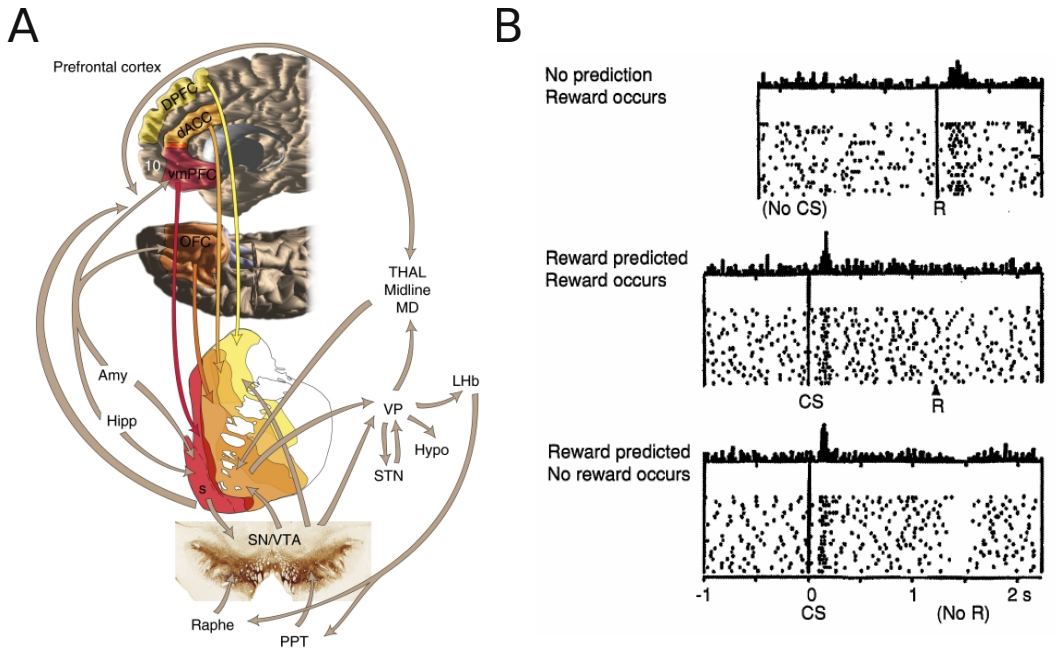

Fig. 1.2 Prediction error signaling in dopaminergic system. (**A**) The anatomy of the reward circuit illustrating the main structures and pathways. Amy - amygdala; dACC - dorsal anterior cingulate cortex; dPFC - dorsal prefrontal cortex; Hipp - hippocampus; LHb - lateral habenula; hypo - hypothalamus; OFC - orbital frontal cortex; PPT - pedunculopontine nucleus; S - shell, SNc - substantia nigra, pars compacta; STN - subthalamic nucleus.; Thal - thalamus; VP - ventral pallidum; VTA - ventral tegmental area; vmPFC - ventral medial prefrontal cortex. Figure from Haber and Knutson (2010). (**B**) Phasic activity of dopamine neurons during classical conditioning experiment. CS - conditioned stimulus; R - reward. Figure from Schultz et al. (1997).

characterized by two states: S_{base} when monkey waits for the next event, and S_{cue} when the cue is presented to the monkey. After the conditioning phase, the monkey learns that the cue predicts the reward, so the estimated value of the state with the cue is close to the value of the reward, $V(S_{\text{cue}}) \approx R = 1$. On the other hand, the state S_{base} is the neutral baseline with $V(S_{\text{base}}) \approx 0$. After receiving the unexpected reward, the TD prediction error can be calculated as $\delta = R + V(S_{\text{base}}) - V(S_{\text{base}}) \approx 1$. In the expected reward trial, when the cue is presented environment's state changes from S_{base} to S_{cue}. This change elicits the prediction error $\delta = V(S_{\text{cue}}) - V(S_{\text{base}}) \approx 1$. Then, the cue disappears, which is reflected by the environment transitioning back to the state S_{base}. If the reward is administered during that transition, the prediction error is close to zero, $\delta = R + V(S_{\text{base}}) - V(S_{\text{cue}}) \approx 0$, otherwise it is negative, $\delta = V(S_{\text{base}}) - V(S_{\text{cue}}) \approx -1$. It is now apparent that the phasic activitiy of monkey's dopaminergic neurons perfectly reflects the TD prediction error – the firing rate increases when $\delta > 0$, decreases when $\delta < 0$, or remains on the baseline when $\delta \approx 0$. The statement that phasic activity of dopaminergic neurons broadcast the prediction error learning signal was encapsulated with the *reward prediction error hypothesis* (RPEH) of dopamine.

Over thirty years of research provided converging evidence in favor of RPEH. Prediction error signal carried by the dopaminergic neurons was recognized in monkeys (Bayer and Glimcher, 2005; Nakahara et al., 2004; Satoh et al., 2003), rats (Oyama et al., 2010; Roesch et al., 2007) and humans (Zaghloul et al., 2009). Moreover, dopamine responses in the striatum turned out to be consistent with the TD model predictions in classical conditioning experiments with blocking (Waelti et al., 2001). Other experiments have shown that phasic dopamine activity is scaled by both reward magnitude and reward probability related to the specific cue (Fiorillo et al., 2003). This probability-scaling effect is also predicted by the TD model – cue that more reliably predicts future reward is associated with the state with a higher value, hence the signaled prediction error $\delta = V(S_{\text{cue}}) - V(S_{\text{base}})$ is higher.

The RPEH is considered as "one of the largest successes of computational neuroscience" (Colombo, 2014). RPEH provided a deep and elegant explanation of animal learning and inspired a new area of neuroscientific inquiry. Despite its undeniable success RPEH also has limitations like any scientific theory. For example, the RPEH was mainly verified in the setting with rewards and without punishments. However, it is well known that animals can effectively learn how to avoid punishments. This observation raises the question – *how the brain dopaminergic system implements punishment-avoidance learning*? In the following section, I will shed light on this issue by presenting recent research on punishment-avoidance learning.

1.5 Punishment-avoidance learning

According to behaviorism, punishment is any intervention that reduces the likelihood of repeated animal behavior. For behaviorists, punishment and reinforcement act like two opposite forces that either decrease or increase preceding behavior. Our conscious experience as human beings also shows that rewards and punishments relate to two unique categories of events and trigger distinct emotions and behaviors. However, the computational perspective on learning does not impose that rewards and punishments must be considered as separate categories. For example, in TD learning, it is perfectly reasonable to represent rewards and punishments as values from a single continuum ranging from negative to positive. Using that approach, one can simply model punishments as negative rewards $R < 0$. On the other hand, physiological constraints suggest that a single dopaminergic system may not be sufficient to represent negative prediction errors effectively (Bayer and Glimcher, 2005). Specifically, according to neurophysiological experiments, negative prediction errors are encoded as suppressed phasic activity of dopaminergic neurons (**Fig. 1.2B**). However, since these neurons' baseline activity (action potential frequency) ranges between 2-10Hz and phasic activity following the onset of the stimulus can reach up to 30Hz, the possible range of suppression is much narrower than the available range of activation. In other words, negative prediction errors cannot be precisely encoded because the firing rate is always above zero.

Several hypotheses have been proposed to reconcile these conflicting observations (**Fig. 1.3A**) (Palminteri and Pessiglione, 2017):

- **Single system hypothesis**. The first hypothesis postulates that only one dopaminergic learning system is responsible for broadcasting positive and negative PEs. According to this hypothesis, negative PEs are encoded as the duration of the suppressed baseline activity of dopaminergic neurons (Maia and Frank, 2011).

- **Gradient hypothesis**. The second hypothesis, similarly to the first hypothesis, states that only one dopaminergic learning system exists, but different parts of this system encode rewards and punishments as increased phasic activity of the neurons. Specifically, the assumption is that ventral parts of the striatum are signaling rewards, whereas dorsal parts are responsible for punishment-avoidance behavior (Seymour et al., 2007).

- **Serotoninergic opponent system hypothesis**. The third hypothesis assumes that the opponent system uses neuromodulator serotonin to signal negative prediction errors. The serotoninergic system is well known for its impact on decision-making and avoidance learning (Homberg, 2012).
- **Dual systems hypothesis**. The fourth hypothesis states that negative prediction errors are signaled by other cortical and subcortical structures not directly related to dopaminergic or serotoninergic systems. The precise organization of this negative prediction-error network is still a matter of debate, but recent meta-analysis has shown that aversive outcomes are reflected by increased activity of dorsomedial cingulate cortex, bilateral anterior insula, bilateral dorsolateral prefrontal cortex, thalamus, and amygdala (Fouragnan et al., 2018). Moreover, several animal studies using different experimental paradigms suggested that these regions are implicated in avoidance behaviors (Hayes et al., 2014).

The behavioral data analysis suggests the difference in processing positive and negative prediction errors supporting the dual systems hypothesis. Many studies have shown that the magnitude of the TD update in equation (1.9) is biased depending on the sign of the prediction error (Frank et al., 2007; Gershman, 2016; Gershman et al., 2009). The standard TD model does not offer the mechanism to treat positive and negative prediction errors differently. A common practice to incorporate the assumption of different neural substrates of reward-seeking and punishment-avoidance learning is introducing differential learning rates to the model (Frank et al., 2007). Differential learning rates for positive and negative prediction errors allow asymmetric value updates, explaining risk preference effects (Niv et al., 2012). For example, higher learning rates for positive prediction errors lead to risk-seeking behavior. However, it is unclear whether differential learning rates should distinguish between positive and negative prediction errors, rewards and punishments, or both.

Even though all four hypotheses are mutually exclusive, one can find evidence to all of them. According to Palminteri and Pessiglione (2017), this complicated picture can result from an improper separation between outcome valence and prediction error sign in most of the experiments. The distinction between rewards and punishments and between positive and negative prediction errors is orthogonal to one another because both types of prediction errors can arise in a purely rewarding or punishing context (**Fig. 1.3B**). In the reward-rich environment, a smaller reward or the omission of the reward will result in a negative prediction error. On the other hand, in the context of prevailing punishments, successful avoidance will result in a positive prediction error. The conflicting results on the opponent system may arise due to confounding outcome

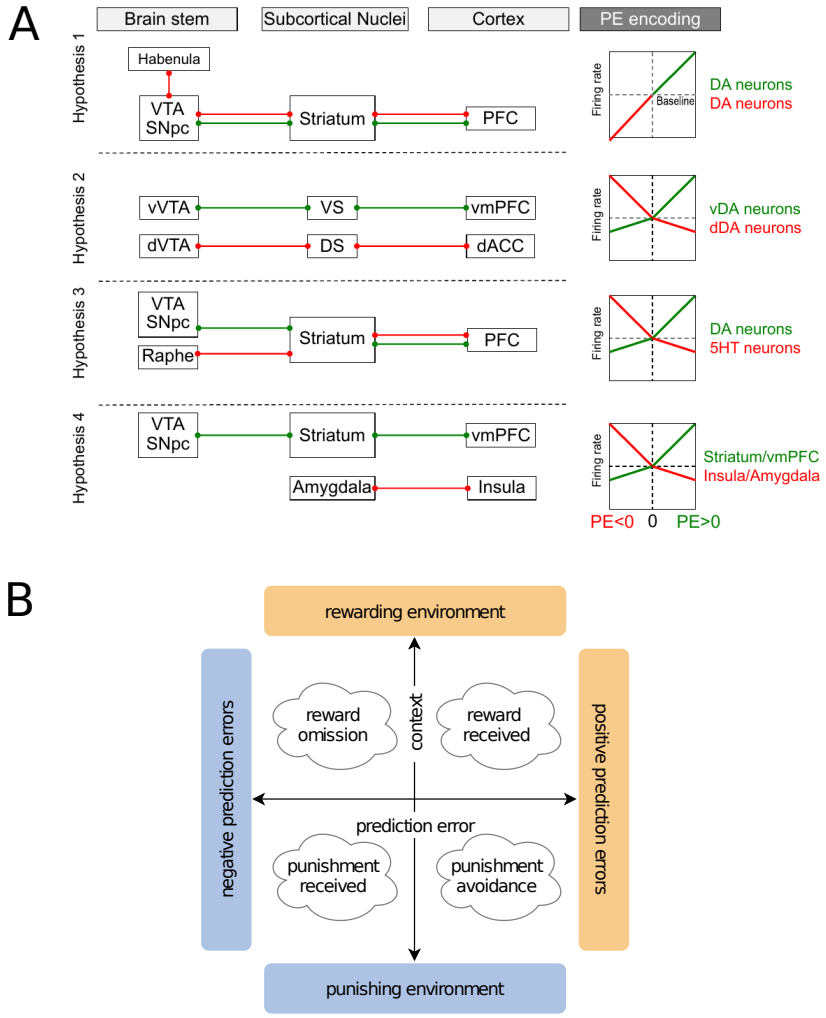

Fig. 1.3 Punishment-avoidance learning. (**A**) Main hypotheses on punishment-avoidance learning implementation in the brain. Red links are the connections within the punishment-avoidance system, whereas green links correspond to the reward-seeking circuit. 5HT - serotonin; DA - dopamine; dACC - dorsal anterior cingulate cortex; dDA - dorsal dopamine; DS - dorsal striatum; dVTA - dorsal VTA; PFC - prefrontal cortex; SNpc - substantia nigra pars compacta; vDA - ventral dopamine; vmPFC - ventromedial PFC; VS - ventral striatum; VTA - ventral tegmental area; vVTA - ventral VTA. Figure from Palminteri and Pessiglione (2017). (**B**) Classification of prediction errors along two orthogonal directions – prediction error sign and outcome valence. Both types of prediction errors can occur in a purely rewarding or punishing context.

valence changes with prediction error sign changes. Therefore, it is critical to design an experimental paradigm carefully to delineate between four types of prediction errors and distinguish between different kinds of changes.

It is still largely unknown along which experimental axis – outcome valence axis or prediction error sign axis – brain systems are organized. Several theories have been formulated to explain decision-making in rewarding and punishing contexts. A common theme among these theories revolves around a *reference effect* (Rigoli, 2019). *The reference effect hypothesis* states that the value of our decisions is not constructed in absolute terms but arises as the average stimulus value encountered in the past. It has critical implications for choice and learning processes. The reference effect is implicitly embedded in classical theories of choice like Expected Utility Theory or Prospect Theory (Camerer and Loewenstein, 2011; Kahneman and Tversky, 1979), and in the TD model of learning. If the reference effect determines the reward system's organization, the system should be sensitive to changes of prediction error signs but not outcome valence. Changing between reward-seeking and punishment-avoiding environments would be accompanied by the system's adjustments of the reference point, called *normalization*, allowing valence invariant processing of prediction errors.

The reference effect and normalization were recently demonstrated by behavioral studies on context-dependent choice (Khaw et al., 2017; Louie et al., 2013). Some studies provided evidence of the reference effect in the brain. For example, Rigoli et al. (2016) used fMRI to examine brain responses during a gambling task with two different reward distributions. They showed that value responses in the ventral tegmental area/substantia nigra and hippocampus were context-dependent. Moreover, this context-dependence effect increased with an increased contextual influence on choice. In another study, Park et al. (2012) investigated prediction errors elicited by Pavlovian cues for different lotteries and monetary payoffs. They found that the ventral striatum signaled normalized prediction errors consistent with the reference effect assumptions. This finding suggested that context-dependent normalization of prediction errors during learning is equally important as value normalization during choice (Rangel and Clithero, 2012).

In summary, punishment-avoidance learning is not fully understood because most research on reinforcement learning focused on positive reinforcement using the reward context. The essential question to understand how the brain implements punishment-avoidance behaviors is to delineate between outcome-valence changes and prediction error sign changes and describe how brain systems are organized in relation to this distinction. In the next chapter, I will review essential findings on the neural correlates

of prediction errors in the human brain, emphasizing the difference between different types of prediction errors and methodological approaches.

Chapter 2

Human brain imaging of prediction errors

The formulation of the reward prediction error hypothesis and converging evidence from electrophysiological studies supporting RPEH had an immense impact on the field of decision-making and learning. The astonishing connection between a relatively simple computational model of learning and neural signaling of a specific group of neurons encouraged neuroscientists to investigate this connection using various neuroimaging modalities. One of these modalities is functional resonance magnetic imaging, offering a noninvasive means of examining cortical and subcortical brain activity under changing experimental conditions. The non-invasiveness and the ability to study subcortical areas made fMRI a perfect candidate for a tool to investigate PE processing in humans. In this chapter, I will describe basic fMRI principles and methodology enabling the investigation of neural correlates of prediction errors. I will also review critical findings on human prediction error processing with an emphasis on punishment-avoidance learning.

2.1 Functional magnetic resonance imaging

Functional magnetic resonance imaging allows measuring brain activity by detecting blood flow fluctuations. The fMRI technique was developed in the early 1990s by Seiji Ogawa's group at Bell Laboratories (Ogawa et al., 1990). Since its invention, it has become a prevailing experimental technique in cognitive neuroscience. fMRI combines an image generation process from magnetic resonance imaging (MRI) with the knowledge of metabolic changes following brain activity.

The MRI technique relies on the *quantum magnetic resonance* process. Resonance can only occur within a powerful static magnetic field generated by *MRI scanner*. Modern scanners use superconducting magnets to produce static fields of usually 1.5, 3, or 7 Tesla. When a person's brain, composed of ∼75% of water (Mitchell et al., 1945), slides in the MRI scanner, hydrogen atom's nuclei, i.e., protons inside water molecules interact with the scanner's static magnetic field by aligning their spin in the same direction as the external field. Due to the interaction between magnetic momentum and external magnetic field, spins precess around the external field axis with *Larmor frequency*:

$$\omega_0 = \gamma B_0 \tag{2.1}$$

where γ is gyromagnetic ratio equal to $42.58\frac{\text{MHz}}{\text{T}}$ for hydrogen nucleus, and B_0 is strength of an external field. Furthermore, nuclei spin energies split into two states: low-energy state with spin parallel to the field and high-energy state with spin antiparallel to the field. This spectral splitting is called the *Zeeman effect*. In the equilibrium state, populations of both states are uneven, with slightly more nuclei occupying low-energy state (Landini et al., 2018).

All individual spins sum up to a net magnetization vector, $\vec{M}$. In the absence of an external magnetic field, the random orientations of individual spins cancel out, and $|\vec{M}| = 0$. However, when a static field is present, nonzero net magnetization arises due to the surplus of nuclei occupying a low-energy state. This magnetization is parallel to the external field. The component of the $\vec{M}$ parallel to the external field is called longitudinal magnetization $M_{\parallel}$. The other component of the net magnetization, perpendicular to the field axis, is called transversal magnetization $M_{\perp}$. In the equilibrium state, $M_{\perp} = 0$, because individual spins precess out of sync canceling each other on the plane perpendicular to the field axis.

After the brain is placed inside the scanner, a specialized set of radiofrequency coils emits an electromagnetic wave, called *RF pulse*, that matches the Larmor frequency of hydrogen nuclei. Emitted RF pulse resonates with nuclei affecting net magnetization vector $\vec{M}$ in two ways. First, a subpopulation of nuclei changes from the low-energy state to the high-energy state. In turn, more spins from a high-energy state are aligned antiparallel to the field, decreasing longitudinal magnetization $M_{\parallel}$. Second, individual spins start to precess in phase, which builds up transversal magnetization $M_{\perp}$. By adjusting the parameters of the RF pulse, one can influence net magnetization $\vec{M}$ in multiple ways. For example, the 90° RF pulse tilts net magnetization by 90° causing longitudinal magnetization to disappear, while the 180° RF pulse flips over both the transverse and longitudinal components of the net magnetization.

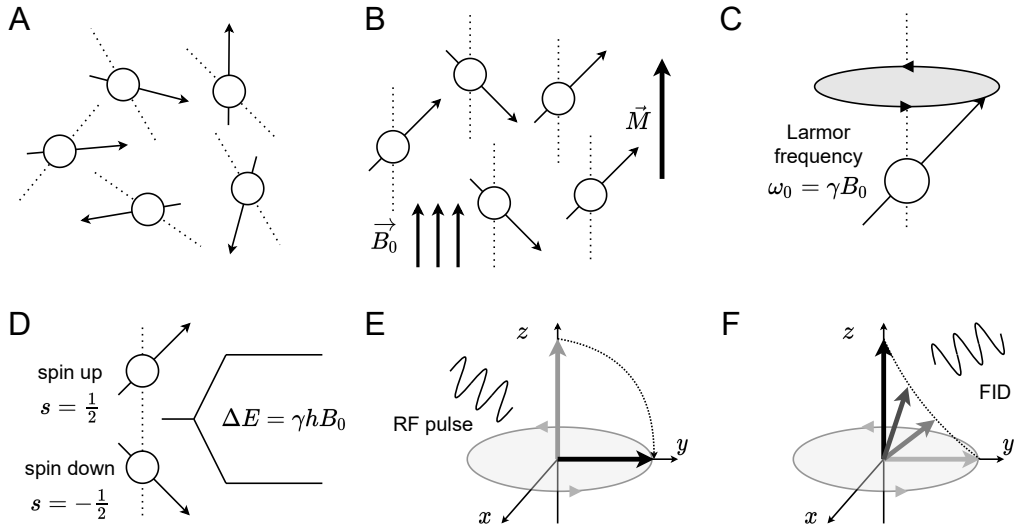

Fig. 2.1 MRI physics. Nuclear magnetic resonance process. (**A**) Random nuclear spin orientation without external magnetic field. (**B**) Nuclear spin alignment in magnetic field $\vec{B_0}$ leads to longitudinal net magnetization $\vec{M}$. (**C**) Larmor precession. (**D**) Zeeman effect for hydrogen nucleus. (**E**) 90° RF pulse flipping net magnetization vector. (**E**) Relaxation process generating free induction delay (FID) signal.

After a brief duration, the RF pulses are turned off, and nuclei begin returning to their equilibrium state in the process of *relaxation*. The time needed to go back to an equilibrium differs between both types of magnetization. For longitudinal relaxation, a time constant T1 determines the pace at which initial $M_{\parallel}$ is rebuilt. A similar time constant, T2, denotes the pace at which transversal magnetization decays due to spin precession decoherence. During these relaxation processes energy is released from the brain tissue in the form of electromagnetic waves. Receiver coils can then measure these waves as free induction decay (FID) signals. Different types of RF pulses are usually combined into *a sequence* to create different MRI contrasts. Each contrast is individually tuned to capture between-tissue differences in relaxation times T1 or T2.

In real systems, transversal magnetization decays at a quicker pace than T2 suggests. Local inhomogeneities of the magnetic field lead to more rapid dephasing of spins aligned by 90° RF pulse. This shorter transversal relaxation time, T2*, forms the basis of functional MRI. The relationship between T2 and T2* is given as:

$$\frac{1}{T_2^*} = \frac{1}{T_2} + \gamma \Delta B_0, \tag{2.2}$$

where ΔB_0 is the difference in local field strengths (Chavhan et al., 2009). Local magnetic field inhomogeneities have non-random component reflecting a map of the chemical environment within the brain. For example, they can be used to measure a concentration of oxygenated and deoxygenated hemoglobin. This is possible because both types of hemoglobin have different magnetic properties: oxygenated hemoglobin is diamagnetic, whereas deoxygenated hemoglobin is paramagnetic (Glover, 2011). Local magnetic field inhomogeneities can be measured using a *spin echo (SE) sequence*. In the SE sequence, the 90° RF pulse is followed by 180° RF pulses flipping net magnetization and causing spins dephased by T2* effects to rephrase. Rephased spins generate echoes that can be registered by receiver coils and used to recreate a T2* weighted brain image.

By measuring rates of concentration of oxygenated and deoxygenated hemoglobin reflected by T2* image, one can indirectly infer brain activity due to the hemodynamic response mechanism. The hemodynamic response is the rapid delivery of oxygenated blood to active neuronal populations. It is crucial for brain sustainability since neurons do not have their reservoir of oxygen and glucose but quickly deplete their energy when active. The hemodynamic response is modeled by the hemodynamic response function (HFR), which reflects expected blood-oxygenation-level-dependent changes (BOLD) following increased neuronal activity. Neuronal activity can be recovered using different modeling techniques from collected T2* images reflecting the BOLD signal.

In the following sections, I will introduce two complementary approaches to modeling BOLD signals. The first approach aims to discover brain activity correlated with latent cognitive processes. It is suitable for identifying brain regions signaling increasing and decreasing prediction errors in reward-seeking and punishment-avoiding environments. The second approach is a method for quantifying functional associations between brain regions during task execution. It allows the examination of task-modulated connectivity, and can be used to investigate brain network reconfiguration related to switching between positive and negative prediction errors and between reward-seeking and punishment-avoiding task conditions.

2.2 Prediction error-related activity

The discovery of prediction error coding in dopaminergic neurons using single-unit recording motivated a search for BOLD correlates of prediction errors. The non-invasiveness and flexibility of fMRI allowed researchers to ask more detailed questions about brain implementation of prediction errors. One of the most intriguing questions

concerned the debated issue of punishment-avoidance learning – **Does the brain employ one or two prediction-error signaling networks?** To provide the answer, researchers developed a *model-based fMRI* approach. It relies on correlating precomputed computational model variables against BOLD signal from a subject performing a cognitive task involving learning the value of the stimuli.

2.2.1 Model-based fMRI

The model-based fMRI approach extends a traditional activation analysis commonly employed to find brain regions involved in the cognitive process in question. However, while an activation analysis merely answers the question "where" a particular cognitive process takes place, the model-based analysis tackles both "where" and "how" questions (O'Doherty et al., 2007). Specifically, it tries to answer how a process is implemented, providing invaluable insights for improving theoretical models of cognition.

Before model-based fMRI, decision-making and learning neuroscience relied on either behavioral data or animal studies to discriminate between competing theories. Behavioral data was sufficient to measure latent variables in simple learning models like the Rescorla-Wagner model. In the RW model, an associative strength was simply estimated as the strength of conditioned response. However, as the models had become more complex and involved multiple latent processes interacting to produce behavior, the sole behavioral approach became less efficient. For example, in the temporal difference model, learning rates may be indirectly measured using learning curves, but prediction errors usually remain unobservable if only behavioral data is used (Rescorla, 2002). To observe what has been previously hidden, one can use the model-based approach by finding BOLD signals correlated with prediction errors. Furthermore, this approach provides an additional source of data – a compelling source of evidence for competing theories.

The backbone of the model-based fMRI method is a computational model providing mapping from task stimuli to the behavioral responses. In other words, a proper computational model can simulate an agent's behavior given the presented stimuli. Almost every model relies on some "internal" operations required to generate responses. These operations are an essential part of the model-based approach. In a typical value-based learning task, these operations include: calculating the value of competing stimuli, choosing the best option, computing prediction errors, and updating value estimates. Model-based fMRI aims to find brain regions signaling these internal operations.

The model-based fMRI analysis usually starts by selecting competing models during the experimental design phase. Then, the experiment is conducted, and models are

fitted to subjects' behavioral data using existing free parameters to minimize the error of their predictions. Once the best-fitting model parameters are found, models are compared, and the model with the highest explanatory power is selected for further analysis. It is important to remember that model complexity can trivially increase model fit and decrease model generalization (Schwarz, 1978). Hence, an essential consideration during model comparison should be the appropriate penalization of model complexity. Multiple solutions like Akaike information criterion (Akaike, 1998), or Bayesian information criterion were proposed as model selection metrics integrating model fit and complexity. The alternative to these standard approaches is Bayesian modeling offering a framework based on Bayes theorem unifying model fitting with model comparison adjusted for complexity (Lee, 2011). After the winning model is determined, internal model operations are extracted and regressed against fMRI data. Regression is usually performed using a general linear model (GLM) with additional model-based regressors. A standard second-level statistical modeling follows this procedure to show areas exhibiting significant correlation between BOLD signal and predicted timeseries reflecting internal model operations.

Just like all methodologies, model-based fMRI has some pitfalls. First and foremost, a significant correlation between a model variable and BOLD timeseries does not decisively prove that a region is implementing hypothesized computations. Other models with poorer fits to behavioral data may still better explain the BOLD signal. One way to overcome this issue is to investigate the explanatory power of competing models using both behavioral and fMRI data (O'Doherty et al., 2007). Second, a recent study has shown that the model-based fMRI approach can be insensitive to gross changes in free parameters like learning rates (Wilson and Niv, 2015). This finding suggests potential difficulties with precisely identifying computation implemented by brain areas and undermines the efficiency of using fMRI data to discriminate between competing models. However, another study contradicted these findings suggesting that sensitivity to free parameters is sufficient when considering different populations, i.e., mental disorder patients or healthy controls (Katahira and Toyama, 2021). Third, a model-based fMRI suffers from poor spatiotemporal resolution. For some cognitive processes, a sampling frequency of ~2s available for a typical fMRI sequence may not be sufficient to capture phenomena lasting hundreds of milliseconds. Similarly, a spatial resolution of ~3mm allows discovering only large neuronal populations carrying the same signal neglecting smaller structures or structures with more spatially diverse neural coding.

2.2.2 Positive and negative prediction errors in the brain

How does a human brain encode prediction errors? This question has been tackled by dozens of studies using the model-based fMRI approach. Researchers used a broad range of task designs, behavioral modeling approaches, and stimuli types. The most common task designs chosen to elicit prediction errors were: probabilistic reversal learning (PRL) task (Lin et al., 2012; Mattfeld et al., 2011; Meder et al., 2016; Nickchen et al., 2017; Schlagenhauf et al., 2014; Van den Bos et al., 2012), Markov decision task (Daw et al., 2011; Guo et al., 2016; Tanaka et al., 2016), multi-armed bandit task (Diuk et al., 2013; Schonberg et al., 2010), or simple Pavlovian conditioning task (O'Doherty et al., 2004; Seymour et al., 2007). Among behavioral models of learning, the three most influential were: the Rescorla-Wagner model (Gläscher et al., 2009; Kahnt et al., 2011; Li et al., 2006; O'Sullivan et al., 2011), the SARSA model (Gläscher et al., 2010; Gradin et al., 2011; Seger et al., 2010), and the temporal-difference model (Guo et al., 2016; Meder et al., 2016; O'Doherty et al., 2003; Van den Bos et al., 2012). To assess whether prediction error signaling is invariant to reward and punishment type, researchers used a variety of primary and secondary reinforcers: sweet juice and salty solutions (McClure et al., 2003; Metereau and Dreher, 2013; Valentin and O'Doherty, 2009), pictures of unhealthy food (Hare et al., 2008), warm and cold thermal stimuli (Rolls et al., 2008), smiley and angry faces (Katahira, 2015; Lin et al., 2012; Meder et al., 2016), money gains and losses (Gläscher et al., 2009; Guo et al., 2016; Ribas-Fernandes et al., 2011), and abstract symbols (Meder et al., 2016).

The most consistent finding of these studies demonstrates a *critical role of the ventral striatum* (VS), i.e., nucleus accumbens, in broadcasting the prediction error signal. Numerous studies have shown that the BOLD signal in VS correlates with modeled timecourse of prediction errors (Abler et al., 2006; Daw et al., 2011; Delgado et al., 2000; Guo et al., 2016; Hare et al., 2008; Katahira, 2015; Lin et al., 2012; Mattfeld et al., 2011; McClure et al., 2003; Meder et al., 2016; Metereau and Dreher, 2013; Ribas-Fernandes et al., 2011; Schlagenhauf et al., 2014; Seymour et al., 2007; Tanaka et al., 2016; Valentin and O'Doherty, 2009; Van den Bos et al., 2012; Watanabe et al., 2013). VS is one of the main parts of the basal ganglia and the receiving end of the dopaminergic pathway. It receives dense projections from dopaminergic neurons located in the ventral tegmental area, i.e., midbrain neurons originally associated with the prediction error signal in monkeys (Bayer and Glimcher, 2005; Nakahara et al., 2004; Satoh et al., 2003; Schultz, 1986). These projections carry PE signals originating in the midbrain to the VS, making it detectable by fMRI techniques.

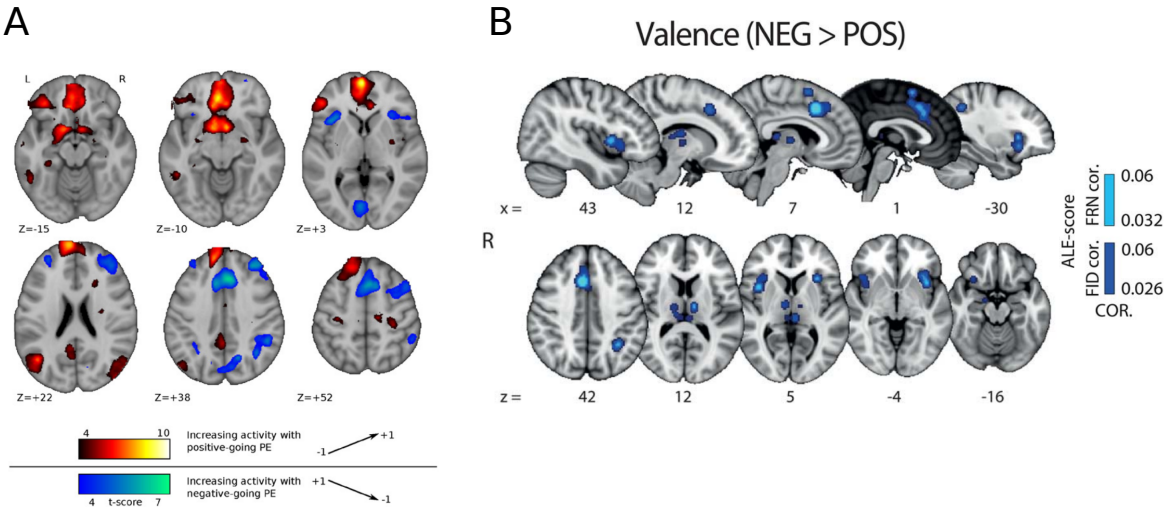

Fig. 2.2 Neural correlates of negative prediction errors. (**A**) Brain regions signaling positive (red) and negative (blue) prediction errors across reward-seeking and punishment-avoiding task conditions. Positive prediction error correlates were found in the dorsomedial prefrontal cortex, vmPFC, bilateral VS, posterior cingulate cortex, left inferior parietal cortex, and left orbitofrontal cortex. Negative prediction errors were signaled in dorsal ACC, right dorsal premotor cortex, right dorsolateral prefrontal cortex, bilateral insula, precuneus, and left dorsolateral prefrontal cortex. Figure from Meder et al. (2016) (**B**) Results from ALE meta-analysis along the PE sign component. In blue, brain areas processing negative prediction errors (activated by NEG>POS pattern). Figure from Fouragnan et al. (2018).

Several other cortical regions have been frequently implicated in signaling increasing prediction errors. For example, multiple studies found PE correlates in the ventromedial prefrontal cortex (vmPFC) (Baram et al., 2021; Daw et al., 2011; Meder et al., 2016; Van den Bos et al., 2012), an area commonly associated with signaling decision value or utility (Hare et al., 2009; Rangel et al., 2008). A large group of studies reported PE correlates in dorsal parts striatum, i.e., caudate or putamen (Delgado et al., 2000; Diuk et al., 2013; Mattfeld et al., 2011; McClure et al., 2003; Valentin and O'Doherty, 2009). Other studies also found BOLD correlates of prediction errors in anterior cingulate cortex (ACC) (Ide et al., 2013; Ribas-Fernandes et al., 2011), parahippocampal gyrus (Van den Bos et al., 2012), insula (Katahira, 2015), and medial temporal cortex (Delgado et al., 2000).

The flexibility of the model-based fMRI allowed researchers to investigate brain regions with response patterns resembling "inversed" prediction errors. These regions increase their activity with decreasing prediction errors, implementing punishment-avoidance learning by broadcasting the negative part of the PE signal. As described in the previous section, four main hypotheses regarding the negative prediction system have been proposed. According to the single system hypothesis, negative prediction errors are signaled by the same dopaminergic system but encoded as the duration of the suppressed neuronal activity (Maia and Frank, 2011). The single system hypothesis predicts that no regions should display increased activity with decreasing prediction errors. The gradient hypothesis suggests a functional separation of striatum into ventral and dorsal parts, reflecting the distinction between positive and negative prediction errors (Palminteri and Pessiglione, 2017). It predicts that caudate and putamen should exhibit increased BOLD signal with decreasing prediction errors. The third hypothesis suggests that the serotoninergic system processes the negative PE signal. Since neurotransmitter serotonin is located mainly within the brainstem raphe complex, this hypothesis is difficult to test using fMRI, which is not suitable for the brainstem imaging (Aghajanian and Liu, 2017; Beissner, 2015). The last dual system hypothesis postulates a distributed system of cortical and subcortical regions carrying the negative prediction error signal. Although the exact organization of this system is still a matter of debate, most researchers suggest the involvement of the insula, amygdala, and cingulate cortex as critical parts of network signaling negative PEs (Palminteri and Pessiglione, 2017; Yacubian et al., 2006).

Several studies investigated punishment-avoidance learning over the past twenty years. In one of the first studies directly investigating neural correlates of negative prediction errors, McClure et al. (2003) measured neural responses following omission of

the expected droplet of juice. Activity in left putamen correlated with negative PEs but also with positive PEs supporting the single system hypothesis. Two other studies used either the Pavlovian conditioning task (Seymour et al., 2007) or associative learning task (Mattfeld et al., 2011) to investigate reward and punishment learning. In both studies, researchers reported functional segregation within the striatum, with anterior parts signaling positive PEs and posterior parts signaling negative PEs. These findings directly corroborate the gradient hypothesis. However, subjects experienced rewards and punishments simultaneously in both experiments, confounding the PE sign with the outcome valence. A large body of evidence from model-based fMRI studies has also supported the dual system hypothesis. For example, Yacubian et al. (2006) reported negative prediction errors in the amygdala, suggesting that "the ventral striatum and the amygdala distinctively process the value of a prediction and subsequently compute a prediction error for gains and losses". In another study, Fazeli and Büchel (2018) used painful thermic stimuli and aversive pictures to investigate how predictions modulate processing of these unpleasant stimuli. Researchers reported aversive prediction error signaling in the anterior insula. Some studies reported a broad network of regions with BOLD signal increasing with decreasing prediction errors. Hauser et al. (2015) used probabilistic reversal learning task to investigate cognitive flexibility in adolescence. They have found negative prediction error processing in anterior insula, dorsomedial and dorsolateral prefrontal cortex, inferior parietal lobule, and precuneus.

Only one model-based fMRI study directly investigated the differences in PE processing between reward-seeking and punishment-avoiding conditions (Meder et al., 2016). Meder et al. (2016) used a factorial design with a probabilistic reversal learning task to disentangle effects related to prediction error sign from outcome valence effects. They reported a set of regions processing negative PE regardless of the outcome valence. These regions included: dorsal anterior cingulate cortex, right premotor cortex, bilateral dorsolateral prefrontal cortex, precuneus, bilateral anterior insula, and primary visual area (**Fig. 2.2A**). Interestingly, there was a significant difference between reward-seeking and punishment-avoidance learning. A stronger neural response to PE in reward-seeking compared with the punishment-avoidance condition was reported in: left inferior frontal gyrus, supplementary motor area, posterior cingulate cortex, left dorsomedial prefrontal cortex, left amygdala, bilateral secondary visual areas, right thalamus, and left middle temporal gyrus. A posthoc analysis showed that in most of these areas, the effect was driven by "strong positive contrast estimates in reward-seeking conditions and contrast estimates close to zero in punishment-avoidance conditions". That finding suggested that some brain regions were selectively involved

in processing positive PEs in the reward-seeking condition. This study supported the dual system hypothesis but challenged the view that the direction of prediction errors is the only important axis along which brain systems are organized. Nonetheless, an outcome-valence-invariant delineation between regions processing positive and negative PEs suggests that two distinct brain networks are responsible for reward-seeking and punishment-avoiding learning regardless of the type of used reinforcement.

A few meta-analyses focused on prediction error processing providing the most convincing evidence supporting the dual system hypothesis. Liu et al. (2011) conducted activation likelihood estimation (ALE) meta-analysis using 142 neuroimaging studies utilizing reward-related decision-making tasks. Although the authors focused on reward and punishment instead of positive and negative PEs, these quantities are often strongly correlated. Nucleus accumbens was reported as the region processing both rewarding and punishing outcomes. Exclusive processing of rewards was reported in the medial orbitofrontal cortex and posterior cingulate cortex. In contrast, punishments were processed within the anterior cingulate cortex, bilateral anterior insula, and lateral prefrontal cortex. Two years later, Garrison et al. (2013) published another ALE meta-analysis explicitly focused on differentiating between Pavlovian versus instrumental learning. The authors found distinct patterns of PE signaling for rewarding and aversive stimuli. Reward prediction errors were reported in the striatum, whereas punishment prediction errors were located within the insula and habenula. However, this meta-analysis ignored the sign of PE-related BOLD signal focusing solely on the types of reinforcers used as stimuli. The only ALE meta-analysis that explicitly modeled effects related to increasing and decreasing PEs was conducted by Fouragnan et al. (2018). In this study, two patterns of BOLD signaling were considered: NEG>POS, where neural responses were higher for negative compared with positive or null outcomes (negative PEs), and POS>NEG, where neural responses were higher for positive compared with negative or null outcomes (positive PEs). The authors found two distinct networks correlating with NEG>POS and POS>NEG patterns. Neural correlates of positive PEs were found in the VS, ventromedial prefrontal cortex, posterior cingulate cortex, ventrolateral orbitofrontal cortex, and dorsomedial prefrontal cortex. In contrast, negative PEs regions included: dorsomedial cingulate cortex, bilateral anterior insula, pallidum, middle frontal gyrus, inferior parietal lobule, middle temporal gyrus, amygdala, thalamus, habenula, dorsolateral prefrontal cortex, fusiform area, precentral cortex, and dorsomedial orbitofrontal cortex (**Fig. 2.2B**). Despite that these results seem to corroborate the gradient hypothesis because of the involvement

of the pallidum in negative PE processing, the amount and extensiveness of significant clusters for NEG>POS pattern firmly favors the dual system's hypothesis.

2.3 Prediction error-related connectivity

The identification of brain areas responsible for signaling prediction errors has been a primary focus of neuroscientists interested in learning and decision-making. As reviewed in the previous section, many studies have found a broad network of coactivated areas signaling positive and negative prediction errors. One critical question that a model-based fMRI cannot address arises from these studies – **How do prediction error areas interact with each other and the rest of the brain during learning?** Understanding both brain activations and interactions is crucial for understanding how the brain implements a specific cognitive process (Friston, 2011). One way to characterize interactions between brain systems is to estimate *functional connectivity*.

2.3.1 Functional connectivity

Functional connectivity (FC) is usually defined as a temporal correlation between activity patterns of spatially separated brain areas (Friston et al., 1993). Although any neuroimaging modality with temporal dimension is suitable for functional connectivity analysis, this approach has been most frequently used with fMRI data. In the context of fMRI, FC reflects a statistical dependency between the BOLD signals of different brain regions.

The idea behind functional connectivity assumes that if two brain areas are strongly coupled or connected, their activity should be correlated. It is important to acknowledge that FC does not show how brain areas influence each other. There are multiple possible sources of functional connectivity between a pair of brain regions. First, a direct structural connection may exist between regions, causing one region's activity to influence the other. It has been shown that the existence of a structural connection between areas usually correlates with a high degree of FC between them (Eickhoff et al., 2010). Second, regions may not share a connection but can be influenced by a third mediating region influencing them both. Third, regions may be involved in a more complicated network of regions consisting of loops which causes them to activate in unison. In some cases, FC can arise as a byproduct of some external event or structured noise. For example, activity in sensory areas induced by the occurrence of stimuli is usually cascaded into parietal regions responsible for the perceptual classification and

premotor cortex for response generation. This parallel processing may induce temporal correlations between structurally disconnected brain areas. It was also shown that physiological effects and motion artifacts could artificially inflate FC estimates (Birn, 2012). In spite of its limitations, FC offers a straightforward and effective approach to investigate brain interactions.

Most research on functional connectivity focused on measuring connectivity during resting-state, i.e., a state when no explicit task is performed. Resting-state functional connectivity can be simply calculated as a Pearson's correlation between BOLD signals of distinct brain areas. Investigation of resting-state interactions led to a famous discovery of *default mode network* and other *resting-state networks* (Raichle, 2015). Recently, there has been increasing recognition for the importance of studying FC during cognitive task execution (Di and Biswal, 2019). Unlike resting-state, cognitive tasks usually consist of short events or blocks distributed across the scanning duration. That feature of task-based fMRI makes simple correlation analysis impossible, posing a need for more sophisticated statistical methods to estimate task-based FC. The two most important approaches to task-based FC are *beta-series correlation* (BSC) (Rissman et al., 2004) and *psychophysiological interaction* (PPI) (Friston et al., 1997). Although both methods rely on different statistical assumptions, a recent study suggested that they should "in principle yield similar results" when comparing the differences between task conditions (Di et al., 2021).

Both BSC and PPI require a division of the brain into spatially separate units, corresponding to either single voxels or clusters of voxels called regions of interest (ROIs). Early task-based FC research implemented a *seed-voxel* approach focusing on connectivity between a single ROI, called seed, and the rest of the voxels in the brain. This limiting approach neglects a large-scale perspective on the brain, emphasizing that the brain is a complex network with many parts interacting to meet cognitive demands of the organism. An alternative to seed-voxel is a ROI-ROI approach focusing on pairwise interaction between dozens or even hundreds of regions (Fornito et al., 2012; Gerchen et al., 2014). The advantage of the ROI-ROI over the seed-voxel approach stems from the fact that ROI-ROI analysis does not require a priori assumptions about the role of particular brain regions and allowing to quantify complex interactions beyond a scale of a single region.

2.3.2 Beta-series correlation

Beta-series correlation is an efficient method for assessing context-dependent functional connectivity during task execution (Cisler et al., 2014). Rissman et al. (2004) proposed

BSC as a method intended for event-related designs. BSC relies on a simple idea that a pair of functionally connected areas should simultaneously activate or deactivate with response to similar events. In other words, BSC estimates the correlation of trial-by-trial BOLD activations for a pair of areas in a certain task condition.

In a typical BSC analysis, each trial is modeled as a separate regressor in the GLM. This procedure results in a set of *beta maps* representing the voxel-wise BOLD response specific to a particular trial. Beta maps are then averaged within the prespecified seed or ROIs to represent regional responses. A sequence of regional beta values for each trial is called *beta series*. The last step of BSC consists of splitting the beta series according to investigated task modulation and calculating condition-specific correlations between beta series. For example, to investigate brain interactions during prediction error processing, each outcome event of a learning task may be modeled as a separate trial. Then, trials can be divided according to the outcome valence. Positive outcome trials (win or loss-avoidance) induce positive prediction error signaling, whereas negative outcome trials (loss or win-omission) produce negative prediction errors. BSC allows estimating FC for positive and negative prediction errors separately, enabling to investigate differences between the two.

A recent study suggested that BSC is more robust than PPI in detecting FC in event-related designs, especially with many trials and short event durations and inter-stimulus-intervals (Cisler et al., 2014). However, another study using a large data sample did not support this claim (Di and Biswal, 2019). Another report investigated the differences and similarities between BSC and PPI. Authors concluded that "when context-sensitive changes in effective connectivity are present, [...] BSC can reflect similar connectivity differences as measured by PPI" (Di et al., 2021).

2.3.3 Connectivity during prediction error processing

How do prediction error networks interact with each other and the rest of the brain during learning? The answer to this question would provide a more complete understanding of the brain's implementation of reinforcement learning. However, only a handful of studies investigated functional connectivity dynamics associated with prediction error processing.

In one of the first studies on PE-related connectivity, Kahnt et al. (2009) investigated FC of ventral and dorsal striatum during the probabilistic reversal learning task. The PPI analysis showed increased connectivity between VS seed and ventral/anterior midbrain, right hippocampus, pons, and cerebellum during positive prediction error processing. Conversely, the dorsal striatum strengthened its coupling with the dor-

sal/posterior midbrain, right hippocampus, thalamus, and cerebellum. The authors suggested that striatal-midbrain connectivity may implement a value updating process in the striatum.

A similar study conducted by Camara et al. (2009) used the seed-voxel BSC approach to investigate connectivity between the ventral striatum and the rest of the brain during monetary gains and losses. The study has reported a broad network of regions connected with VS regardless of the outcome valence comprised of the amygdala, hippocampus, insula, and orbitofrontal cortex. Researchers have found valence-dependent differences in connectivity between VS and orbitofrontal cortex and between VS and amygdala, with both connections more pronounced for losses.

In another study, Van den Bos et al. (2012) used a probabilistic reversal learning task and a seed-voxel PPI approach to examine developmental changes in learning from trial-and-error. The FC analysis revealed that connectivity between VS and medial prefrontal cortex increased with age. Furthermore, a stronger association between these regions correlated with the learning rate for negative prediction errors demonstrating behavioral relevance of the interaction between the striatum and prefrontal cortex. Some researchers focused on more ecological tasks to investigate neural interactions related to reinforcement learning. For example, Horga et al. (2015) used a virtual maze task with hidden rewards and seed-voxel BSC analysis to examine how gradual learning modulate striatal connectivity. They have found progressive connectivity enhancement between the sensorimotor cortex and posterior putamen as subjects learned the task. Moreover, these connections differentiated participants who learned the task from those who failed. Although the study did not directly address the difference between positive and negative PEs, successful learning is often related to a gradual decrease in PEs, hence the study findings can shed light on the question of PE-related FC.

Two more recent studies investigated how FC between PE-signaling regions is altered in major depressive disorder (MDD) (Kumar et al., 2018) and internet gaming disorder (IDG) (Lei et al., 2020). Both studies used reward-based learning tasks to evoke positive and negative PEs and the PPI approach to assess context-dependent FC. Kumar et al. (2018) found impaired reward prediction error signaling in MDD patients and reduced functional connectivity between the midbrain ventral tegmental area and striatum during PE processing. Researchers also tested the group difference in connectivity between striatum and habenula, i.e., area commonly associated with negative prediction errors, but found no significant differences. On the other hand, Lei et al. (2020) found increased connectivity between the right caudate, right putamen,

bilateral dorsolateral prefrontal cortex, and right dorsal anterior cingulate cortex in the IDG patients compared to healthy controls.

Several other studies also examined functional connectivity during reinforcement learning but did not directly address prediction-error-related changes. Specifically, some studies investigated connectivity during the choice phase (Cohen et al., 2005; Erdeniz and Done, 2019; Fouragnan et al., 2015), used effective connectivity with constrained hypotheses offered by dynamic causal modeling approach (den Ouden et al., 2010; Den Ouden et al., 2009), or investigated differences between the goal-directed and habitual decisions (de Wit et al., 2012).

It is still unclear how regions signaling prediction errors are connected and how these connections are modulated by outcome valence and prediction error signs. Previous studies provided conflicting or, at best inconclusive results regarding functional connectivity changes following PE processing. Some studies reported changes in connectivity between the ventral striatum and orbitofrontal regions (Camara et al., 2009; de Wit et al., 2012; Kumar et al., 2018). In contrast, others showed altered connectivity between the VS and midbrain (Kahnt et al., 2009) or between the putamen and sensorimotor cortex (Horga et al., 2015). The observed diversity of findings is likely a consequence of diversity in task designs, methodological approaches to estimate functional connectivity, and tested hypotheses. In addition, none of the presented studies used an experimental design to differentiate between outcome valence (reward vs. punishments) and prediction error signs (positive vs. negative). Until these two experimental axes remain entangled, the question of which one is accountable for observed changes will remain unanswered.

To the best of my knowledge, all previous studies of PE-related connectivity used a seed-voxel approach with seeds usually located within striatum or an effective connectivity approach with only few regions involved in PE processing. Drawing conclusions solely based on these studies would neglect that the brain is an immensely complex network of interactions between many cortical and subcortical areas. Furthermore, it is well established that the whole brain network consists of multiple stable subnetworks observed during rest and task execution (Gratton et al., 2018). Understanding how these networks interact during learning is critical for understanding how reinforcement learning is implemented in the brain. In the next chapter, I will introduce core ideas of network neuroscience – the study of brain networks across temporal and spatial scales. I will also show how researchers had successfully used network neuroscience to explain cognitive processes, emphasizing why this approach is critical for understanding reinforcement learning.

Chapter 3

Network neuroscience

The human brain is an exceptionally complex network comprised of almost a hundred billion neurons and more than a hundred trillion synapses (Azevedo et al., 2009). The emergence of graph theory and network neuroscience allowed measuring and modeling brain networks using data from different neuroimaging modalities. Graph theory provided a mathematical framework to abstract a brain as a system comprised of elements and interactions among them. Network neuroscience has used that framework to identify connectivity patterns reflecting the structural and functional brain organization. These patterns revealed that the brain network has a modular and small-world structure with highly connected hub regions (Bullmore and Sporns, 2009; Bullmore and Bassett, 2011). In this chapter, I will introduce the assumptions and core concepts of the network theory. I will then review the essential findings on functional brain networks, highlighting research on brain network dynamics during cognitive processing.

3.1 Network theory

Network theory has its roots in graph theory – a branch of discrete mathematics studying graphs. A graph is a simple mathematical structure modeling pairwise associations between objects. The foundations of graph theory were established in 1736 by Leonhard Euler, who solved the famous Königsberg bridge problem. The Königsberg bridge problem concerned finding a closed path through the prussian city of Königsberg that traversed each of seven bridges only once. Euler proved the negative result for the problem by recognizing the significance of *topological properties* – features invariant to any continuous deformation of the geometrical object, which led him to formulate the definition of a graph. Since its inception, graph theory has provided solutions to multiple puzzling mathematical problems, like map coloring, traveling

salesman, or finding a shortest path in the graph. The advancement of technology in the XXI century enabled scientists from various fields to collect large datasets, which provided rich information about interacting elements of studied systems. The need to discover meaning within these datasets inspired a network theory – a study of real-world networks using graph theory mathematics. Network theory has been applied in many disciplines, e.g., computer science (Riaz and Ali, 2011), sociology (Barnes, 1969), and biology (Mason and Verwoerd, 2007). Recently, neuroscientists applied a network theory to study functional connectivity in the human brain (Goldenberg and Galván, 2015).

3.1.1 Weighted undirected graph

A graph is a mathematical representation of a structure comprised of a set of objects related to each other. The most basic graph is an *undirected graph*, defined as a pair $G = (\mathcal{N}, \mathcal{E})$, where $\mathcal{N} \equiv \{n_1, n_2, \ldots, n_N\}$ is a set of nodes (or vertices), and $\mathcal{E}$ is a set of edges (or connections). The number of nodes and edges is usually denoted as $|\mathcal{N}| = N$ and $|\mathcal{E}| = K$. Graph nodes are labeled by the integer index i, which corresponds to their order in a set $\mathcal{N}$. In the undirected graph, an edge denoted as e_{ij} represents a bidirectional association between nodes n_i and n_j. The nodes connected by an edge are referred to as *adjacent*. A set of nodes $\mathcal{N}$ is assumed to be finite, which implies that $\mathcal{E}$ is also finite.

A *directed graph* is a generalization of an undirected graph in which edges have orientation pointing from one node to another. In a directed graph, the order of indices indexing an edge is important, $e_{ij} \neq e_{ji}$, because e_{ij} denotes an edge pointing from n_i to n_j while e_{ji} denotes an edge pointing from n_j to n_i. In the case of brain interactions, a directed graph is suitable to represent effective connectivity which shows the causal influence of one brain region on another. In contrast, functional connectivity typically reflects directionless statistical dependency between signals. Hence an undirected graph is a more appropriate model for functional connectivity.

A graph of size N can have at most $\binom{N}{2} = \frac{N(N-1)}{2}$ edges. Graph with an edge between all pairs of nodes is referred to as *fully connected*. Graphs that are almost fully connected, i.e., $K = \mathcal{O}(N^2)$ are called *dense*, whereas graphs with little connections, i.e., $K \ll N^2$ are called *sparse*.

Both directed and undirected graphs described above were also *unweighted* (or *binary*). In unweighted graphs, edges simply exist or not, and none of their properties is relevant. In contrast, in *a weighted graph*, each edge is associated with a numerical value called weight. Edge weight can correspond to any connection property associated

with the studied system. For example, in a functional brain network, a weight of functional connection may represent a correlation coefficient between the BOLD signal of two brain regions. Formally, a weighted graph is a triple, $G^W = (\mathcal{N}, \mathcal{E}, \mathcal{W})$, which additionally include a set of weights (or values), $w_{ij} \in \mathcal{W}$, associated with each edge. The one-to-one correspondence between edges and weights implies that $|\mathcal{W}| = |\mathcal{E}| = K$. Generally, weights are positive real numbers $w_{ij} \in \mathbb{R}_+$. In some applications, this constraint is relaxed to include negative weights $w_{ij} \in \mathbb{R}$. This is especially important in network neuroscience, where negative weights represent an associations between brain regions with the anticorrelated neural signals.

Both weighted and binary graphs can be directed or undirected, resulting in four distinct graph type combinations. Each graph, regardless of its type, has *an adjacency matrix* representation. The adjacency matrix $\mathbf{A}$ is a $N \times N$ square matrix containing complete information about the graph. An element of the adjacency matrix, A_{ij} represents an edge between nodes n_i and n_j, or lack thereof. If an edge is absent, $A_{ij} = 0$. If an edge is present, $A_{ij} = 1$ in a binary graph, or $A_{ij} = w_{ij}$ in a weighted graph. Additionally, the adjacency matrix can encode information about edge directions for directed graphs. In undirected graphs, $A_{ij} = A_{ji}$ which implies that the adjacency matrix is symmetrical. In directed graphs, A_{ij} and A_{ji} represent both possible edge directions pointing from n_i to n_j or from n_j to n_i. The adjacency matrix provides a unified way of representing different graphs and simplifies the computation of many graph properties.

One of the most fundamental graph properties is the degree distribution. It provides information on how connections are distributed across network nodes. The degree distribution reflects the probability distribution over node degree in binary networks or node strength in weighted networks. The *degree* (or *strength*) of a node n_i is simply the number of edges connected to this node (or sum of its connection weights). Both node degree and strength can be calculated using adjacency matrix as $k_i = \sum_j A_{ij}$.

3.1.2 Modularity

Many real-world networks reveal community structure. For example, a group of friends in a social network can be considered a network *module* (or community) – a subnetwork with densely interconnected nodes and sparse connections with the rest of the network. In a functional brain network, modules correspond to sets of brain areas with highly correlated activity. Understanding network community structure is crucial for understanding network organization and dynamics. Modules are beneficial for the network dynamics, fostering efficient information flow between nodes. Different groups

of nodes can be independently assigned to different functions in modular systems, promoting network efficiency. It was shown that large modular networks are usually more efficient than non-modular networks (Tosh and McNally, 2015).

Several methods have been developed to identify the network's community structure: distance-based modules, Infomap algorithm, block models, independent component analysis, and modularity maximization (Sporns and Betzel, 2016). Modularity maximization is by far the most widely used approach to investigate brain network communities. In this approach, a network is divided into a set of nonoverlapping partitions to maximize quality function Q. Function Q is referred to as *modularity* and reflects the quality of the community division. Modularity reflects a simple heuristic – a network is considered modular if the *within community connection density is higher than expected by chance*. The expected probability of a connection between pairs of nodes can be estimated using a null graph model with random topology but preserved degree distribution. In the weighted graph, probability of a connection is replaced by expected connection weight. According to the null model, a probability that two nodes with degrees $k_i = \sum_j A_{ij}$ and $k_j = \sum_i A_{ij}$ will share an edge is given as:

$$P_{ij} = \frac{k_i k_j}{2m}\,, \tag{3.1}$$

where $2m = \sum_{ij} a_{ij}$ is the total number of network edges in a binary graph or sum of edge weights in a weighted graph. The intuition behind equation (3.1) is simple – the probability of random connection between nodes is proportional to the product of node degree, i.e., network nodes with many connections with the rest of the network have a high chance of being randomly connected. This null model enables to express the modularity heuristic as:

$$Q = \sum_{ij}(A_{ij} - P_{ij})\delta(\sigma_i, \sigma_j)\,, \tag{3.2}$$

where σ_i indexes the community to which node i is assigned, and δ is the Kronecker delta function.

The process of finding optimal community structure requires changing a candidate partition, hoping to find the one that maximizes quality function Q. The brute force approach cannot be used in this case because even for relatively small networks, the space of all possible partitions is enormous. Multiple heuristic-based algorithms have been developed to allow approximating optimal community structure in real-world networks. These algorithms include spectral decomposition (Newman and Girvan, 2004), simulated annealing (Guimera and Amaral, 2005) and greedy Louvain method

(Blondel et al., 2008). The Louvain method is widely used in network neuroscience and promises to run in time $\mathcal{O}(N \log N)$. It relies on an iterative process in which small communities are found by optimization of local modularity and then grouped into a single node. The Louvain algorithm is based on a random process, so its output can vary from run to run. This implies that when applied to real data, the algorithm should be run multiple times to produce a representative set of high-quality partitions.

3.1.3 Small-worldness, hubs, and scale-free networks

A *small-world graph* is a graph with a high degree of clustering and low distance between nodes. Specifically, high clustering reflects nodes' tendency to form clusters, which means that two nodes sharing the same neighbour have a high probability of being connected. At the same time, in a small-world graph, an average number of edges required to traverse to move between random pair of nodes is roughly $\log N$, where N is the number of nodes in the network. The small-world phenomenon was initially discovered in social networks (Milgram, 1967) and formalized by Watts and Strogatz (1998). According to the Watts–Strogatz model, the small-world graph is the intermediate step between the random network with low clustering and a short distance between nodes and the regular lattice network with high clustering and long distance between nodes. A degree of network small-worldness can be calculated as the standardized ratio of the clustering coefficient and path length (Humphries and Gurney, 2008). The small-world phenomenon has been found in almost all real-world networks, i.e., communication, genetic, social, and brain (Bullmore and Sporns, 2009). The small-world organization of a network promotes two types of information processing: segregated local processing within clusters of nodes and integrated whole-network processing facilitated by short path (Bullmore and Bassett, 2011).

Hubs are nodes with high node centrality. The *node centrality* is any measure that ranks node position within a graph. It can be based on many properties like information flow, neighbors, or influence on other nodes. The most straightforward measure of centrality is a node degree. Hubs are critical nodes allowing efficient communication within a graph (Freeman, 1977). Studies showed that damaging hubs might disrupt a graph's capability to process information efficiently (Fornito et al., 2010). The existence of hubs is directly related to a scale-free property (Barabási, 2009).

A *scale-free network* is a network with a power-law degree distribution. In other words, the probability of finding a node with a degree k can be approximated as $k^{-\eta}$, where η is a scale parameter usually ranging between 2 and 3 (Choromański et al., 2013). The power-law degree distribution with a "heavy-tail" implies that the network

has no specific scale and contains nodes with extremely high degrees (hubs). Barabási and Albert (1999) showed that scale-free networks emerge in a process in which new network nodes are preferentially connected to already existing high-degree nodes. This model of network growth is called *preferential attachement*. Many real-world networks demonstrate a truncated power-law degree distribution associated with fewer hubs than in ideal scale-free networks. This is a consequence of cost constraints of real-world networks embedded in a physical space (Amara et al., 2011).

3.2 Functional brain networks

The growing interest in investigating functional connectivity during rest and tasks using the ROI-ROI approach led to the birth of a new branch of network neuroscience focusing on understanding functional brain networks. Functional brain networks can be modeled as weighted undirected graphs with negative weights. In these networks, brain regions are nodes, and interregional functional connectivity estimates are edge weights. Modeling brain interactions as networks enabled researchers to utilize the mathematical tools of network science to ask critical questions about brain dynamics and organization. The network approach provided a better understanding of brain network properties and led to the discovery of resting-state networks. Recently there has been growing interest in studying functional brain networks during various cognitive processes. Combining the functional brain network approach with the study on human cognition can shed new light on the implementation of these processes in the brain.

3.2.1 Resting-state networks

In one of the first resting-state experiments in fMRI, Biswal et al. (1995) discovered that BOLD signals in the left and right motor cortex were highly synchronized. Biswal's discovery motivated researchers to explore spontaneous synchronization between other brain areas. Several studies found correlated signals between regions of the primary visual network, auditory network, or cognitive networks (Cordes et al., 2001; Damoiseaux et al., 2008, 2006; Van Den Heuvel et al., 2008; Xiong et al., 1999). These studies demonstrated that the brain is constantly active and that its activity forms highly correlated spatial patterns (Greicius et al., 2009). These patterns comprised of brain areas with synchronized activity at rest were termed *large-scale networks* (LSNs). It turned out that the organization of LSNs closely resembles activity patterns elicited by various cognitive tasks (Smith et al., 2009; Thomas Yeo et al., 2011). The ubiquity and

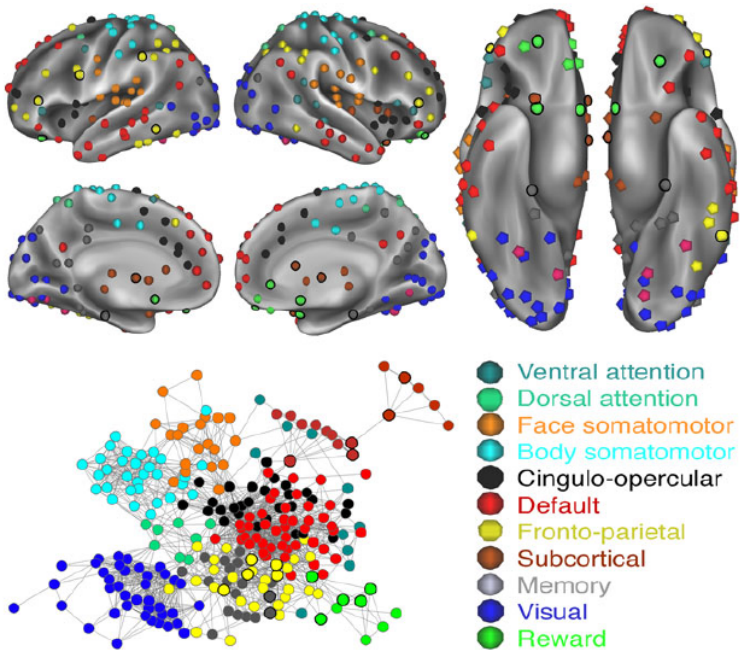

Fig. 3.1 Resting-state networks. Brain division into eleven large-scale networks during resting-state for a large cohort of 828 subjects. Regional network membership represented as a color code for different brain areas was detected using community detection algorithm. Reward-related regions formed preferentially coupled reward system with high stability (green). Bottom panel shows a spring-embedded layout of the network shown on the top panel. Figure from Huckins et al. (2019).

importance of LSNs were emphasized by the studies demonstrating LSNs in humans under sedation and during sleep (Mhuircheartaigh et al., 2010; Tagliazucchi and Laufs, 2014) and in other mammalian species like rats and monkeys (Lu et al., 2012; Mantini et al., 2011).

The exact number and organization of LSNs is still a matter of debate, but most researchers delineate from ten to fifteen LSNs (De Luca et al., 2006; Huckins et al., 2019; Power et al., 2011). Commonly reported LSNs include: default mode network, fronto-parietal network, cingulo-opercular network, ventral and dorsal attention networks, and somatosensory networks (**Fig. 3.1**). The issue of LSNs topography is challenging because of the hierarchical organization of functional brain networks (Doucet et al., 2011). For example, a recent study on functional network organization in individuals reported consistently finding nine subnetworks within the default-mode network (Gordon et al., 2020). Understanding the role of LSNs is vital for our understanding of the human brain and cognition. Below I describe the topography and functions of major LSNs.

Default mode network

The default mode network (DMN) is the largest LSN comprised of the medial prefrontal cortex, posterior cingulate cortex, lateral parietal cortex, and lateral temporal cortex (Buckner et al., 2008). The DMN topography can be recovered from seed-voxel analysis using the posterior cingulate cortex, one of the most important hubs in a functional brain network, as a seed region (Greicius et al., 2003).

To this day, the exact role of the DMN remains elusive. Interestingly, the areas of the DMN have been found to consistently deactivate during the execution of cognitive tasks (Fox et al., 2005). The DMN is most active when subjects are resting (Raichle et al., 2001). Some studies suggested that the DMN is responsible for self-generated cognitive processes like moral judgment or autobiographical memory (Andrews-Hanna, 2012). These properties of DMN led some researchers to coin the terms *task-negative* and *taks-positive* networks for LSNs comprised of regions deactivating or activating during cognitive tasks. Some researchers suggested that the dichotomy of task-negative DMN and other task-positive networks represent antagonism resulting in the interference between self-generated thoughts and task execution (Cocchi et al., 2013). Despite its claimed "task-negativity," recent studies demonstrated that the DMN might play an important role in task-related cognitive processes like working memory (Finc et al., 2017), cognitive training (Finc et al., 2020), and reinforcement learning (Dohmatob et al., 2020).

Fronto-parietal network

The fronto-parietal network (FPN) is a task-positive network comprised of the prefrontal cortex, inferior parietal lobule, middle temporal gyrus, dorsomedial prefrontal cortex, and anterior cingulate cortex. The regions of the FPN activate when a task requires adaptive control or its cognitive demands increase. The FPN is thought to be responsible for goal-oriented control processes and working memory. Some researchers hypothesized that the FPN inhibits intrusive thoughts and distractions generated by the DMN and switches attention in line with current cognitive demands (Vincent et al., 2008). Others suggested that the FPN can act as a mediator between the DMN and other task-positive networks to support goal-directed control (Spreng et al., 2010).

Cingulo-opercular network

The cingulo-opercular network (CON) is another task-positive network composed of the operculum, anterior cingulate cortex, anterior insula, and thalamus (Sadaghiani

and D'Esposito, 2015). In contrast to the FPN, the CON is responsible for maintaining goal-directed control, task monitoring, and alertness (Coste and Kleinschmidt, 2016).

Dorsal and ventral attention networks

The dorsal attention network (DAT) is a task-positive network encompassing areas adjacent to the intraparietal sulcus and frontal eye fields. The activity of DAT regions is anticorrelated with the DMN, and it constitutes the most consistent negative correlation in the functional brain network (Fox et al., 2005). The DAT is active during tasks requiring spatial attention and visual working memory (Vossel et al., 2014). It was suggested that the DAN plays a significant role in a top-down attention process (Corbetta and Shulman, 2002).

The counterpart of the DAT responsible for bottom-up attention is the ventral attention network (VAT). Two main parts of the VAT are the ventral frontal cortex and temporoparietal junction (Vossel et al., 2014). The regions of VAT are activated during unexpected or stimulus-driven attention. The VAT displays a high degree of asymmetry between left and right-lateralized areas constituting this network. Specifically, VAT regions within the left hemisphere overlap with Broca and Wernicke areas; thus, they are strongly activated during language processing tasks.

Reward network

The reward network was recently recognized as a LSN by Huckins et al. (2019) in a study examining a large dataset of resting-state fMRI. Researchers wanted to establish whether regions involved in reward processing form a separate LSN. They identified a system of interconnected reward-related areas, stable across a wide range of connectivity thresholds. This reward network consisted of the VS, lateral and medial orbitofrontal cortex, and vmPFC. It was classified as a subordinate system along with other LSNs like DAT and VAT. Interestingly, the reward network was the second most stable subordinate network.

3.2.2 Functional brain networks during cognition

Functional brain networks are primarily dominated by stable group and individual features (Gratton et al., 2018). Despite that stability, changes between distinct task states correlate with predictable transitions between network's functional states (Cole et al., 2014). Investigating these transitions can give us critical insight into understanding how functional networks facilitate cognitive processes.

Since the inception of network neuroscience, the most scientific effort has been put into investigating connectivity patterns during resting-state. However, in the past decade, there has been a growing interest in exploring these patterns during human cognition. According to the recent systematic review, the central areas of application of network neuroscience to human cognition include human intelligence, cognitive load, working memory performance, and behavioral performance in natural environments (Farahani et al., 2019). A number of studies also investigated brain network reconfiguration during motor and value learning (Bassett and Mattar, 2017; Gerraty et al., 2018; Mattar et al., 2018).

Human intelligence is a subtle and complex function of human cognition. According to the Parieto-Frontal Integration Theory (P-FIT), intelligence is promoted by the long-range interactions between brain areas residing within frontal and parietal cortices (Jung and Haier, 2007). These brain areas comprise task-positive fronto-parietal and attentional networks. Many studies have supported P-FIT by demonstrating intriguing relationships between general intellectual abilities and properties of functional brain networks (Hilger et al., 2017; Langer et al., 2012; Van Den Heuvel et al., 2009). For example, higher intelligence was associated with increased functional integration between frontal and parietal regions, high centrality of hub regions within the salience network, and the short overall distance between network nodes. It was also shown, that the intelligence quotient is positively correlated with nodal centrality within the attention network and negatively correlated with nodal centrality of DMN regions (Wu et al., 2013).

Several studies investigated functional network reconfiguration during increasing cognitive load. Theoretical neuroscientists formulated the Global Workspace Theory (GWT), which states that the brain processes low-effort tasks within specialized modules (Baars, 2002; Dehaene et al., 1998). In contrast, high-effort complex tasks require the formation of an integrated workspace characterized by increased integration between distinct brain modules (Bullmore and Sporns, 2012). Many studies empirically supported GWT assumptions. Shine et al. (2016) demonstrated that the brain network fluctuates between integrated and segregated states with increased involvement of integrated states during high cognitive demands. In another study, Braun et al. (2015) showed that increased network reconfiguration within frontal networks correlated with working memory performance. Vatansever et al. (2015) showed that network modularity decreased with increasing demands of the n-back working memory task. Moreover, the higher magnitude of modularity decrease was associated with better behavioral performance. Two further studies supported and expanded on Vatansever et al. (2015)

findings by showing that modularity breakdown associated with increased cognitive demands is related to decreased segregation of DMN and increased integration between DMN and other LSNs (**Fig. 3.2**) (Finc et al., 2020, 2017). The effect of cognitive load on modularity was also observed during the semantic decision-making task (DeSalvo et al., 2014). Compared with resting-state, intra-modular connections strengthened, while intra-modular connections weakened during choice.

Gradual reorganization of functional brain networks has been observed during various types of learning. For example, motor sequence training has been related to increased segregation between visual and motor networks (Bassett et al., 2015). Moreover, better learning performance was correlated with increased autonomy of hubs in frontal and cingulate cortices and elevated network flexibility (Bassett et al., 2011, 2015). Another study reported increased segregation of the DMN and increased integration of task-positive LSNs during 6-week working memory training (Finc et al., 2020). Mattar et al. (2018) investigated functional network reconfiguration during 4-day learning of the values of novel stimuli. They found that connections between visual, frontal, and cingulate networks became stronger with learning progression. Studies of functional network reconfiguration following learning suggest that task automation and efficiency are related to a more segregated network organization with increased autonomy of task-relevant systems.

To this date, only a handful of studies explored functional brain network reconfiguration associated with reinforcement learning. Gerraty et al. (2018) used a modified version of the probabilistic learning task to investigate gradual changes in network coordination associated with value learning. Dynamic connectivity analysis revealed increased integration between the striatum and distributed brain regions located within visual, orbitofrontal, and ventromedial prefrontal cortices. Additionally, the flexibility of the striatal network was correlated with the learning rate and precision fitted to subjects' responses. In another study, Sadler et al. (2020) investigated the dynamics of a reward-related network during a taste-motivated reinforcement learning task. In this task, the administration of sweet taste was related to the positive prediction error processing, whereas bitter taste elicited negative prediction errors. Researchers reported prediction-error-related effects on community structure. During positive PE processing, the ventromedial prefrontal cortex was coupled with bilateral precuneus, whereas during negative PE processing, it formed a separate module with the bilateral pre/postcentral gyrus and bilateral dorsal striatum. On the other hand, ventrolateral prefrontal cortices separated from the rest of the reward network during positive PE processing and integrated with the ventromedial prefrontal cortex and dorsal striatum

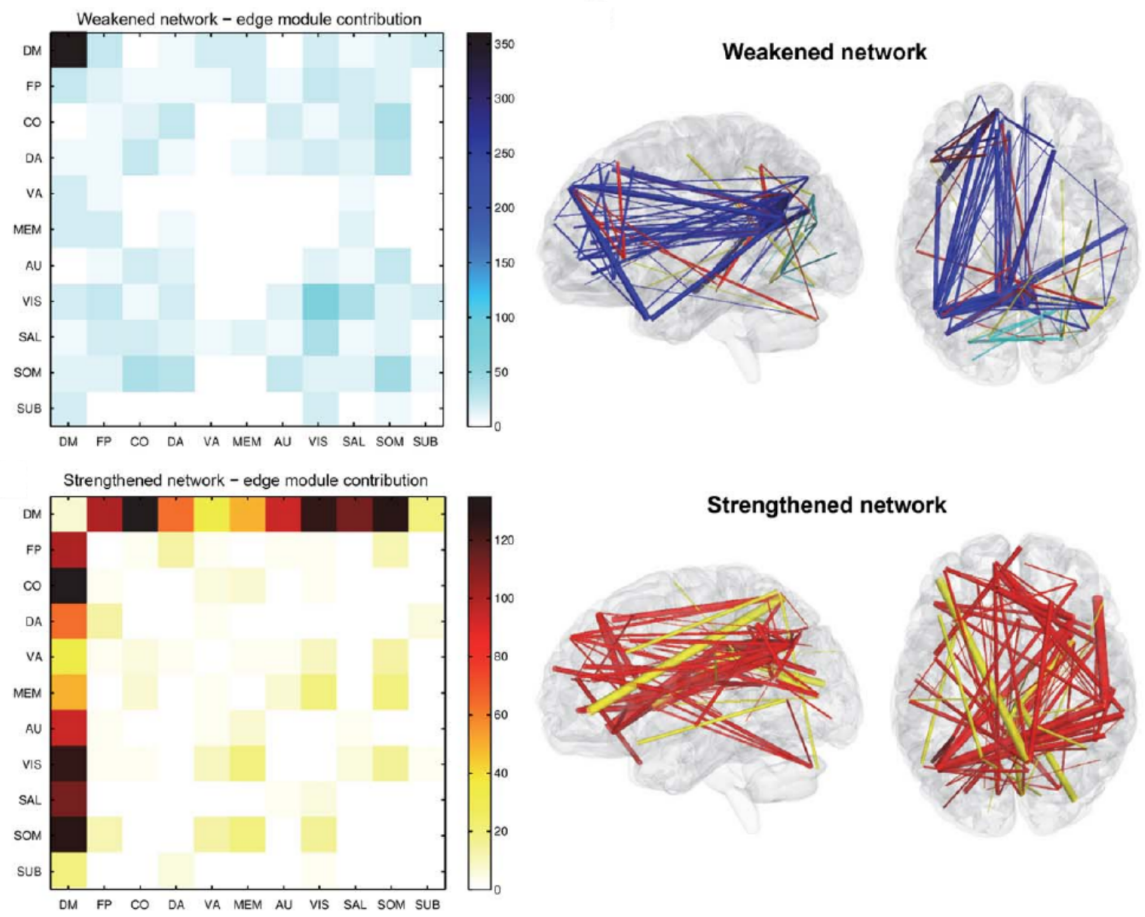

Fig. 3.2 Network reorganization following increasing cognitive load. Functional connections that significantly increased (bottom panel; strengthened network) or decreased (top panel; weakened network) their connection strength with increasing difficulty of working memory n-back task. The weakened network is comprised mostly of intra-DMN edges, whereas the majority of strengthened network connections link DMN with other LSNs. This effect is in line with similar studies on cognitive load showing increased integration and decreased segregation of large-scale networks. DM - default mode; FP - fronto-parietal; CO - cingulo-opercular; DA/VA - dorsal/ventral attention; MEM - memory; AU - auditory; VIS - visual; SAL - salience; SOM - somatomotor; SUB - subcortical. Figure from Finc et al. (2017).

during negative PE processing. One limitation of this study was that it did not investigate other large-scale networks and their interactions with the reward network. Moreover, reward-related regions were selected a priori and might not have precisely reflected the underlying reward system.

The network neuroscience approach of investigating functional brain network reconfiguration during human cognition provided vital insight into our understanding of intelligence, cognitive load, and gradual learning. It helped to extend neuroscientific theories of various cognitive processes by providing supporting evidence and new ideas. However, there have been only few attempts to study functional network dynamics during reinforcement learning. Therefore, the question of large-scale network reconfiguration during prediction error processing remains unresolved. To address this gap, I designed and conducted an fMRI study focused on examining how large-scale networks interact and how these interactions change when subjects switch between the processing of positive and negative prediction errors and between reward-seeking and punishment-avoiding environments. In the following chapter, I will describe this study.

Chapter 4

Prediction error processing during reward and punishment learning

4.1 Introduction

In the previous chapters, I showed that the prediction error signal reflecting the difference between expected and experienced outcomes is the critical element of reinforcement learning. Despite that positive and negative prediction errors arise from the same underlying computation, Palminteri et al. (2010) suggested that the brain uses separate systems for processing both types of prediction errors. A recent meta-analysis identified two spatially distinct learning systems for processing positive and negative prediction errors (Fouragnan et al., 2018). However, it is unclear whether the functional brain network resembles a similar distinction between positive and negative PE systems.

Although positive prediction errors are usually related to rewarding outcomes, they can also signal relief from avoiding punishment when the agent perceives the environment as generally adverse (Nieuwenhuis et al., 2005; Palminteri et al., 2015). Similarly, negative prediction errors can arise when the reward is anticipated but not provided in a rewarding environment. These reference effects assume that the brain uses the reference point to which experienced outcomes are compared (Bavard et al., 2018; Rangel and Clithero, 2012). This consideration raises an interesting question of whether both learning systems are invariant to the outcome valence. Using outcome valence as an explicit experimental factor, studies have found that the ventral reward system signaled positive prediction errors irrespectively of the outcome dimension (Meder et al., 2016; Palminteri et al., 2015). On the other hand, regions of the amygdala, inferior frontal gyrus, and dorsomedial prefrontal cortex signaled positive PEs in reward but not punishment context (Meder et al., 2016). Due to these inconclusive findings,

outcome invariance of learning systems is still a debated issue. Moreover, it is unknown whether the reference effect is also reflected by the outcome invariance of the functional brain network.

Despite a substantial number of studies focused on neural underpinnings of positive and negative prediction errors, the current view lacks an understanding of the neural interactions associated with prediction error processing. Only a few studies investigated connectivity profiles of the ventral striatum or amygdala during reward and punishment processing. For example, a similar network of brain regions encompassing the amygdala, orbitofrontal cortex, and insula was coupled with VS during monetary gambling irrespectively of the outcome valence. However, the orbitofrontal cortex had stronger connections with VS during punishing trials (Camara et al., 2009). Other studies found increased functional connectivity between VS and vmPFC (Van den Bos et al., 2012) and between the amygdala and VS, midbrain, cingulate cortex, thalamus, orbitofrontal cortex, and dorsolateral prefrontal cortex during reward processing (Cohen et al., 2005). Overall, these findings yield inconsistent answers to whether regions encoding prediction errors share similar or different connectivity profiles during positive and negative prediction error processing.

To this date, there have been only few attempts to investigate large-scale brain networks during reinforcement learning. However, most studies investigated gradual network changes associated with value learning neglecting possible rapid changes related to switching between positive and negative prediction errors (Gerraty et al., 2018; Mattar et al., 2018). Only one study directly investigated network reorganization associated with prediction error switching. In this study, Sadler et al. (2020) used a taste-motivated learning task to examine the dynamics of the reward network. The community structure of a functional network was affected by the prediction error sign – ventromedial and ventrolateral prefrontal cortices changed their module assignment during sweet taste eliciting positive PEs compared with bitter taste eliciting negative PEs. However, this study investigated only the reward network ignoring other large-scale networks. Therefore, there is still no explanation for how prediction error processing shapes whole-brain network dynamics.

To answer these outlined research questions, I designed an fMRI study using a probabilistic reversal learning paradigm with separate reward-seeking and punishment-avoiding conditions. I analyzed collected data on three levels: behavioral, neural activity, and neural connectivity. On the behavioral level, I employed a Bayesian model with four competing submodels to assess which computational model of reinforcement learning best explains the subjects' decisions. I used parameters of the best fitting model

to estimate experienced prediction errors. These prediction errors were then subjected to the model-based fMRI analysis and beta-series correlation analysis, allowing the investigation of PE-related changes in neural activity and neural interactions. I was particularly interested in examining the whole-brain functional network during prediction error processing. Specifically, I wanted to describe how this network organizes itself into modules and how this organization shifts when subjects switch between the processing of positive and negative outcomes. These three complementary approaches allowed me to construct a thorough characterization of prediction error processing in the human brain.

4.2 Hypotheses

I stated three central hypotheses using available theoretical work and empirical findings on punishment-avoidance learning and large-scale network reconfiguration during cognition. I hypothesized that (1) two separate systems are responsible for positive and negative PE processing, (2) agents rescale their prediction errors according to the reference effect, and (3) large-scale networks increase their integration during negative prediction error processing. I tested the first two hypotheses using all three data analyses: behavioral, activation, and connectivity. The third hypothesis was tested only on the connectivity level since it described specific network effects unrelated to the behavioral or activation levels.

Dual systems hypothesis

I hypothesized that a separate set of brain regions outside the dopaminergic system signals negative prediction errors. This hypothesis is based on multiple activation studies and meta-analyses showing that negative prediction errors are signaled in the dorsomedial cingulate cortex, anterior insula, dorsolateral prefrontal cortex, and amygdala (Fazeli and Büchel, 2018; Fouragnan et al., 2018; Hauser et al., 2015; Meder et al., 2016; Yacubian et al., 2006). On the activation level, I hypothesized that these regions would exhibit increasing BOLD response with decreasing prediction errors. On the connectivity level, I hypothesized that prediction error processing regions would form a separate community with two sub-communities corresponding to the dopaminergic system signaling positive prediction errors and opponent cortico-insular system signaling negative prediction errors.

According to the dual systems hypothesis, independent brain systems implement learning from positive and negative prediction errors. This independence may be

behaviorally reflected by the differential speed of learning from positive and negative prediction errors. From the behavioral modeling perspective, this would result in independent learning rates for temporal-difference value updates following positive and negative outcomes. I hypothesized that a family of models with separate learning rates for positive and negative prediction errors would outperform other model families in explaining subjects' decisions.

Reference effect hypothesis

I hypothesized that the primary axis along which the brain's learning systems are organized is the prediction error sign axis, not the outcome valence axis. This is one formulation of the reference effect hypothesis, which states that decision values are actively constructed based on average values present in the environment. The reference hypothesis was supported by both behavioral (Khaw et al., 2017; Louie et al., 2013) and neuroimaging studies using fMRI (Park et al., 2012; Rigoli, 2019).

The average stimulus values differ between reward-seeking and punishment-avoiding conditions in the probabilistic learning task with explicitly controlled outcome valence. The reference hypothesis postulates the reference effect, which predicts dynamically adjusted reference point reflecting generally positive outcomes during reward-seeking and generally negative outcomes during punishment-avoiding. This mechanism is responsible for signaling negative prediction errors in the reward-seeking condition (omission of reward) and positive prediction errors in the punishment-avoiding condition (avoidance of punishment). The reference effect enables utilizing both positive and negative prediction error systems regardless of the distribution of rewards and punishments.

According to the reference effect hypothesis, prediction error systems should be invariant to the outcome valence. On the behavioral level, this hypothesis would predict that subject's learning rates for positive and negative prediction errors are invariant to the task condition. I hypothesized that a behavioral model with separate learning rates for positive and negative prediction errors and identical learning rates for reward-seeking and punishment-avoiding conditions would outperform other competing reinforcement learning models. On both activity and connectivity levels, outcome valence invariance suggests that neural correlates of prediction errors should be identical during reward-seeking and punishment-avoiding. Specifically, statistical maps for the PE effect should reveal significant areas of the reward system, whereas statistical maps for the condition effect should not contain any significant changes. However, I expected that the study might not confirm this strict version of the hypothesis since

evidence suggests subtle valence-related differences in PE processing (Meder et al., 2016). Therefore, I relaxed the original statement and hypothesized that PE-related changes in activity and connectivity would be far more pronounced than outcome-valence-related changes. Similarly, on the connectivity level, I hypothesized that both global and local network reconfiguration occurs when switching between positive and negative prediction errors and not when changing between reward-seeking and punishment-avoiding environments.

Global Workspace hypothesis

The Global Workspace hypothesis states that low-effort tasks are processed within specialized modules, whereas high-effort tasks require the formation of an integrated workspace characterized by a high level of integration between specialized modules (Baars, 2002). What are low-effort and high-effort tasks in the context of probabilistic reversal learning? I hypothesized that negative prediction errors processing requires higher cognitive effort than positive prediction errors processing. In the reversal learning setup, positive prediction errors usually confirm the subject's internal judgment about a more beneficial option and are typically followed by choice repetition. On the other hand, negative prediction errors lead to a conflict – the subject has to judge whether the source of the error lies in the environment stochasticity or reflects a shift of reward contingencies. This conflict requires the increased engagement of cognitive resources.

Several studies demonstrated that increased cognitive effort was related to decreased network modularity, increased integration, and reduced segregation of cognitive systems (Braun et al., 2015; Finc et al., 2017; Shine et al., 2016; Vatansever et al., 2015). Finc et al. (2017) showed that the load-related modularity breakdown result from decreased segregation of DMN and its increased integration with other large-scale networks, especially task-positive ones. I hypothesized that the same effect would be observed when switching from positive to negative prediction errors. Specifically, I expected decreased segregation of the DMN and the reward network and increased integration between these networks and the task-positive networks. I also hypothesized that whole-brain network modularity should decrease along with decreasing prediction errors.

4.3 Experimental design

4.3.1 Probabilistic reveral learning task

A probabilistic reversal learning (PRL) task was selected to examine neural correlates of prediction errors (Cools et al., 2002). The PRL task is a powerful paradigm for investigating updating and relearning the stimulus-reward association through reinforcement. Here, participants performed the PRL task in two conditions (1) reward-seeking and (2) punishment-avoiding. In each of these variants, winning and losing are opposed to neutral outcomes, enabling disentangling between often confused dimensions: outcome valence and prediction error sign (Palminteri and Pessiglione, 2017). This quality is crucial to investigate how these dimensions independently affect decision behavior and neural correlates of prediction errors.

Both task conditions were performed inside the fMRI scanner. Participants were instructed to repeatedly choose between yellow and blue boxes to collect as many points as possible in the reward-seeking condition or lose as few points as possible in the punishment-avoiding condition (**Fig. 4.1A**). One of the boxes had the probability of being correct (rewarding or non-punishing depending on the condition) $p = 0.8$ and the other one $p = 0.2$. These probabilities were unknown to the subjects and had to be learned from experience. The reward/punishment contingency changed four times throughout each task condition (**Fig. 4.1C**). Each box had associated reward/punishment magnitude, randomly selected at the beginning of each trial. Magnitudes for both boxes were integers summing up to 50, with the difference between them not exceeding 40. They were represented as white numbers on the boxes, indicating possible gain in the reward-seeking condition or loss in the punishment-avoiding condition. Successful performance in the PRL task requires the decision-maker to correctly estimate correct choice probabilities from experience and integrate them with reward/punishment magnitudes to choose an option with a higher expected value.

Each task condition was associated with the separate fMRI run and consisted of 110 trials. Each trial began with the decision phase indicated by the question mark appearing within the fixation circle (**Fig. 4.1B**). During the decision phase a subject had 1.5 s to choose one of the boxes by pressing a button on the response grip with either left or right thumb. The decision phase was followed by a variable inter-stimulus-interval (ISI; 3-7 s, jittered), after which an outcome was presented for 1.5 s. During the outcome phase fixation circle was colored accordingly to the rewarded or punished box, and the number within the circle represented the number of gained

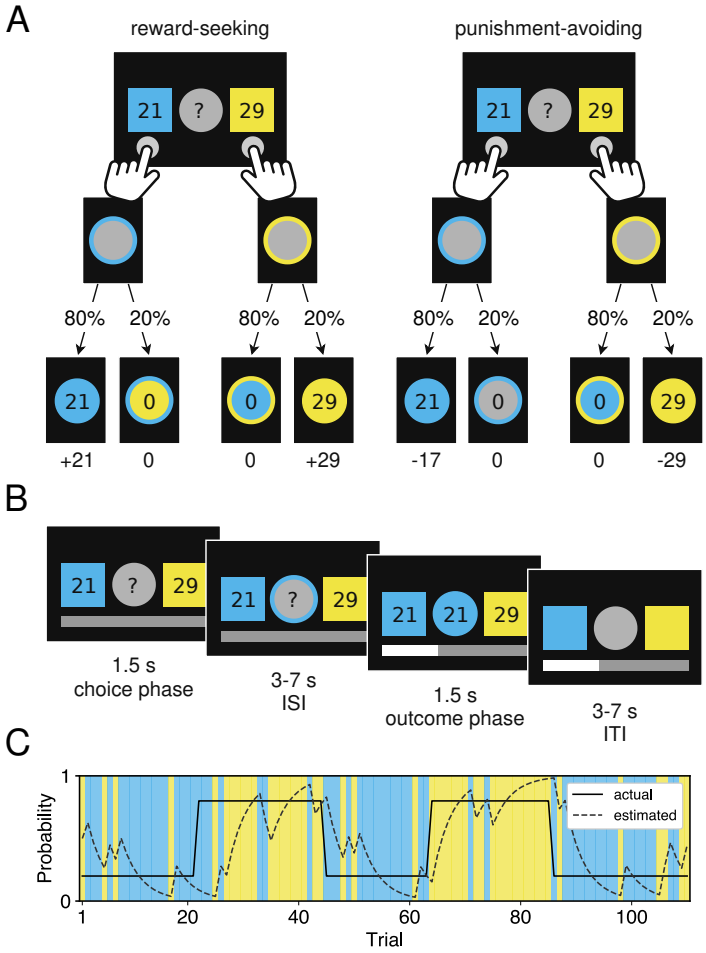

Fig. 4.1 Probabilistic reversal learning task. (**A**) Task structure for reward-seeking and punishment-avoiding conditions of the PRL task represented as a decision tree. In the punishment-avoiding condition, numbers on the boxes represented punishment magnitudes, the fixation circle changed its color according to the punished box, and the number of lost points was displayed in the middle. (**B**) Example trial of probabilistic reversal learning task in reward-seeking condition. Each trial began with a choice phase where the subject had 1.5s to choose between blue and yellow boxes. Subjects had to consider both reward magnitudes visible on the boxes and reward probabilities estimated from previous trials. After variable inter-stimulus-interval (ISI), an outcome phase began giving the subject choice feedback. During the outcome phase, the fixation circle changed its color according to the rewarded box, and the number of rewarded points was displayed in the middle. The outcome phase was followed by a variable inter-trial-interval, after which a new trial began. (**C**) Each task condition consisted of 110 trials. Reward contingency changed four times throughout the task. The solid line represents true reward probability for the yellow box, whereas the dashed line represents prediction of a standard TD model.

or lost points. The outcome phase was followed by a variable inter-trial-interval (ITI; 3-7 s, jittered).

The gray account bar on the bottom o the screen represented points that a subject gathered in the reward-seeking condition or the remaining points in the punishment-avoiding condition. In the reward-seeking condition, subjects were informed that if they fill half of the bar or the entire bar, they will receive 10 PLN ($\approx$ 2.5 USD) or 20 PLN ($\approx$ 5 USD), respectively. Similarly, in the punishment-avoiding condition, they were informed that they would receive 20 PLN if left with more than half of the bar, 10 PLN if left with less than half of the bar, and no money if they lose all of their points. To maintain a constant motivation throughout the task, incentives thresholds were set such that all participants acquired 10 PLN from either task.

Heterogeneity in the prior expectations regarding the task structure may lead to heterogeneity in behavior even in simple tasks leading to inaccurate behavioral modeling (Shteingart and Loewenstein, 2014). Therefore, participants were explicitly instructed that one of two boxes (without telling which) will be more frequently rewarded in the reward-seeking condition or punished in the punishment-avoiding condition and that this contingency may reverse several times throughout the task. Before the MRI scan, subjects practiced both task conditions on the lab computer. During the first phase of the practice, participants were provided with feedback indicating which box is more frequently correct to ensure that they grasp the correct model of the task environment.

PsychoPy software (v. 1.90.1; www.psychopy.org; Peirce (2007)) was used for task presentation on the MRI compatible NNL goggles (NordicNeuroLab, Bergen, Norway). Behavioral responses were collected using MRI-compatible NNL response grips (NordicNeuroLab, Bergen, Norway), which were held in both hands. Each condition lasted approximately 24 min. The order of task conditions and the colors for the left and right boxes (yellow and blue) were counterbalanced across subjects.

4.3.2 Subjects

Thirty-two healthy volunteers (14 female; mean age: 20.9$\pm$2.24; age range: 18-28) were recruited from the local community through social networks and word-of-mouth. All participants were right-handed, had a normal or corrected-to-normal vision, and did not suffer from neurological or psychiatric disorders at the time of examination or in the past. Informed consent was obtained in writing from each participant, and ethical approval for the study was obtained from the Ethics Committee of the Nicolaus Copernicus University Ludwik Rydygier Collegium Medicum in Bydgoszcz, Poland, in

accordance with the Declaration of Helsinki. One participant was excluded from the neuroimaging part of the analysis due to reversed placement of the response grips.

4.3.3 Data acquisition

Brain imaging data were collected using a GE Discovery MR750 3 Tesla fMRI scanner (General Electric Healthcare) with a standard 8-channel head coil. Anatomical images were obtained using a three-dimensional high resolution T1-weighted (T1w) gradient-echo (FSPGR BRAVO) sequence (TR = 8.2 s, TE = 3.2 ms, FOV = 256 mm, flip angle = 12 degrees, matrix size = 256 × 256, voxel size = 1 × 1 × 1 mm, 206 axial oblique slices). Functional images were obtained using a T2*-weighted gradient-echo, echo-planar imaging (EPI) sequence sensitive to BOLD contrast (TR = 2,000 ms, TE = 30 ms, FOV = 192 mm, flip angle = 90 degrees, matrix size = 64 ×64, voxel size 3 × 3 × 3 mm, 0.5 mm gap). Two runs of probabilistic reversal learning in the reward-seeking and punishment-avoiding conditions were acquired (24 min 20 s; 730 volumes each). Forty-two axial oblique, interleaved slices were scanned for each functional run, and five dummy scans (10 s) were collected at the beginning of each run to stabilize magnetization.

4.3.4 Data preprocessing

The raw DICOM data were converted to NifTI format, structured according to the Brain Imaging Data Structure (BIDS) standard, and validated using BIDS Validator (https://bids-standard.github.io/bids-validator/) (Gorgolewski et al., 2016; Yarkoni et al., 2019). The preprocessing was performed using fMRIPrep version 1.4.1 (Esteban et al., 2019; Gorgolewski et al., 2011).

The T1-weighted images were corrected for intensity non-uniformity (INU) with N4BiasFieldCorrection (Tustison et al., 2010), distributed with ANTs 2.2.0 (Avants et al., 2008), and used further used as reference T1-weighed image. The T1w reference was then skull-stripped with a Nipype implementation of the `antsBrainExtraction.sh` workflow (from ANTs), using OASIS30ANTs as the target template. Brain tissue segmentation of cerebrospinal fluid (CSF), white matter (WM), and gray matter was performed on the brain-extracted T1w using FAST (FSL 5.0.9, Zhang et al. (2000)). Brain surfaces were reconstructed using recon-all (FreeSurfer 6.0.1, Dale et al. (1999)), and the brain mask estimated previously was refined with a custom variation of the method to reconcile ANTs-derived and FreeSurfer-derived segmentation of the cortical gray matter of Mindboggle (Klein et al., 2017). Volume-based spatial normalization

to the standard space (MNI152NLin2009cAsym) was performed through nonlinear registration with antsRegistration (ANTs 2.2.0), using brain-extracted versions of both T1w reference and the T1w template. The ICBM 152 Nonlinear Asymmetrical template was selected for spatial normalization (version 2009c; TemplateFlow ID: MNI152NLin2009cAsym; Yoon et al. (2009)).

For each of the two fMRI runs per subject, the following preprocessing was performed. First, a reference volume and its skull-stripped version were generated using a custom methodology of fMRIPrep. The blood-oxygen-level dependent reference was then co-registered to the T1w reference using bbregister (FreeSurfer), which implements boundary-based registration (Greve and Fischl, 2009). Co-registration was configured with nine degrees of freedom to account for distortions remaining in the BOLD reference. Head-motion parameters for the BOLD reference were estimated before any spatiotemporal filtering using mcflirt (FSL 5.0.9, (Jenkinson, 2003)). fMRI time series were slice-time corrected using 3dTshift from AFNI 20160207 (Cox and Hyde, 1997).

Then, time series were resampled to surfaces on the fsaverage5 space, their original native space, and standard MNI152NLin2009cAsym space by applying a single composite transform to correct head-motion and susceptibility distortions. These resampled time series are referred to as preprocessed time series. First, a reference volume and its skull-stripped version were generated using a custom methodology of fMRIPrep. Several confounding time series were calculated based on the preprocessed time series: framewise displacement (FD), spatial standard deviation of successive difference images (DVARS), and three region-wise global signals. FD and DVARS were estimated for each functional run, using their implementations in Nipype. The three global signals were extracted from the CSF, WM, and whole-brain masks. Additionally, head-motion estimates calculated during the correction step were stored as confounds time series. All head-motion and global signal confounds were expanded with the inclusion of temporal derivatives and quadratic terms (Satterthwaite et al., 2013). Frames that exceeded a threshold of 0.5 mm FD or 1.5 standardized DVARS were marked as motion outliers.

4.4 Behavioral modeling

4.4.1 Model space

Reinforcement learning models can quantitatively account for learning by trial-and-error (Montague, 1999). Precisely, models based on the idea of the temporal-difference

learning postulate that stimulus values are updated proportionally to the prediction error weighted by adjustable learning rate (Wagner and Rescorla, 1972). Because of the anticorrelated design of correct choice probabilities during the PRL task, simultaneous stimulus value updates for the chosen and non-chosen options were assumed (O'Doherty et al., 2007). Following previous studies, models with separate learning rates for positive and negative PEs were considered to account for possible risk-sensitivity effects. These models would be referred to as prediction-error dependent (PD) models (Niv et al., 2012; Reiter et al., 2016; Van den Bos et al., 2012). To capture possible valence effects, e.g., loss aversion, models with separate learning rates for reward-seeking and punishment-avoiding conditions were also included. These models would be referred to as condition-dependent models (CD). To fully explore model space, all possible combinations of models were considered, which resulted in four different models:

1. PICI model with single learning rate independent of the sign of prediction error and outcome valence,
2. PICD model with two separate learning rates for reward-seeking and punishment-avoiding conditions,
3. PDCI model with two separate learning rates for positive and negative PEs,
4. PDCD model with four separate learning rates for prediction error signs and task conditions.

All models assumed that prediction error at each trial t is computed as the difference between experienced outcome, r_t, and the expected probability that chosen option will be correct (rewarding / not-punishing), p_t^c:

$$\delta_t = r_t - p_t^c. \tag{4.1}$$

Outcomes could be either $r_t = 1$ when a stimulus is rewarded / not punished or $r_t = 0$ when stimulus is punished / not rewarded at trial t. Note that randomly drawn reward magnitudes did not follow any structured rules. Hence, participants learned and estimated only correct choice probabilities and not action values. In each trial, probability estimates for both chosen, p_t^c, and unchosen stimulus, p_t^u, were simultaneously updated according to a standard Rescorla-Wagner rule:

$$\begin{cases} p_t^c = p_{t-1}^c + \alpha_t \delta_t \\ p_t^u = p_{t-1}^u - \alpha_t \delta_t \end{cases}, \tag{4.2}$$

where $\alpha_t \in [0, 1]$ is the learning rate used in trial t. Note that learning rate values close to zero result in minor updates and slow learning, whereas values close to one result in probability matching behavior. In the simplest PICI model, a single learning rate, α_t^{PICI}, was used to update probability estimates regardless of PE sign and task condition. The rest of the models assumed separate learning rates depending on task condition, prediction error sign, or both:

$$\begin{aligned}
\alpha_t^{\text{PICD}} &= \begin{cases} \alpha^{\text{RS}} \text{ if condition is reward-seeking} \\ \alpha^{\text{PA}} \text{ if condition is punishment-avoiding} \end{cases}, \\
\alpha_t^{\text{PDCI}} &= \begin{cases} \alpha^{+} \text{ if } \delta_t > 0 \\ \alpha^{-} \text{ if } \delta_t < 0 \end{cases}, \\
\alpha_t^{\text{PDCD}} &= \begin{cases} \alpha^{RS+} \text{ if } \delta_t > 0 \text{ and condition is reward-seeking} \\ \alpha^{RS-} \text{ if } \delta_t < 0 \text{ and condition is reward-seeking} \\ \alpha^{PA+} \text{ if } \delta_t > 0 \text{ and condition is punishment-avoiding} \\ \alpha^{PA-} \text{ if } \delta_t < 0 \text{ and condition is punishment-avoiding} \end{cases}.
\end{aligned} \tag{4.3}$$

Action selection was modeled based on reward/punishment magnitudes and continuously updated probability estimates. Utility of both options was assumed as Pascalian expected value, i.e., a product of the expected probability that the box is rewarded/punished, ρ_t, and reward/punishment magnitude for that box, x_t:

$$v_t^{\text{left/right}} = \rho_t^{\text{left/right}} x_t^{\text{left/right}}. \tag{4.4}$$

Note that in the case of punishment-avoiding condition expected probability that stimulus leads to punishmnt equals one minus the expected probability that it is a correct choice:

$$\rho_t = \begin{cases} p_t & \text{if condition is reward-seeking} \\ 1 - p_t & \text{if condition is punishment-avoiding} \end{cases}. \tag{4.5}$$

Finally, the choice probability was derived by coupling expected values with the softmax policy rule (Luce, 1957):

$$\begin{cases} P_t^{\text{left}} = \exp^{-1}(\beta(v_t^{\text{left}} - v_t^{\text{right}})) \\ P_t^{\text{right}} = 1 - P_t^{\text{left}} \end{cases}, \tag{4.6}$$

where precision (or "inverse-temperature") parameter $\beta \in [0, \infty)$ reflects choice stochasticity by controlling sensitivity of the choice probability to differences in expected value between the two stimuli. Choice probability values served as likelihood functions generating choices in the Bayesian modeling framework.

4.4.2 Bayesian modeling

Bayesian modeling provides a strict and flexible way of relating formal cognitive models to behavioral data (Lee and Wagenmakers, 2014). Bayesian modeling aims to estimate a set of parameters, θ, which can represent hidden parameters of behavioral models or variables used for model comparison. In Bayesian statistics, parameters are represented as probability distributions instead of single numbers, which preserves the information about their uncertainty. The general principle of Bayesian analysis is to use collected data, X, to update the *prior* beliefs about parameters to *posterior* beliefs represented as new probability distributions over parameters in question. The "updating process" is captured in a Bayes formula:

$$p(\theta \mid X) = \frac{p(X \mid \theta)}{p(X)} p(\theta) \tag{4.7}$$

where $p(\theta \mid X)$ is a posterior distribution of θ given the data X, $p(X \mid \theta)$ is a probability of observing data X given prior beliefs about θ, $p(\theta)$ is a prior probability distribution, and $p(X)$ is a normalizing constant. Posterior distribution $p(\theta \mid X)$ represents updated belief about parameters and takes into account both prior beliefs and collected data X.

Here, a Bayesian hierarchical latent-mixture (HLM) model was used to perform model selection and parameter estimation. This type of Bayesian model contains parameters representing behavioral parameters like learning rates or precision and parameters reflecting the belief about which competing model generates the observed data. The hierarchical structure of the model assumes behavioral parameters of individuals coming from group-level distributions. The latent-mixture structure assumes that observed behavior can arise as a combination of computations from different cognitive processes. The HLM model was estimated with PICI, PICD, PDCI, and PDCD submodels as competing submodels for model comparison (**Fig. 4.2**). Then a hierarchical model with a winning submodel as a generative model for behavioral responses was estimated for parameter recovery purposes.

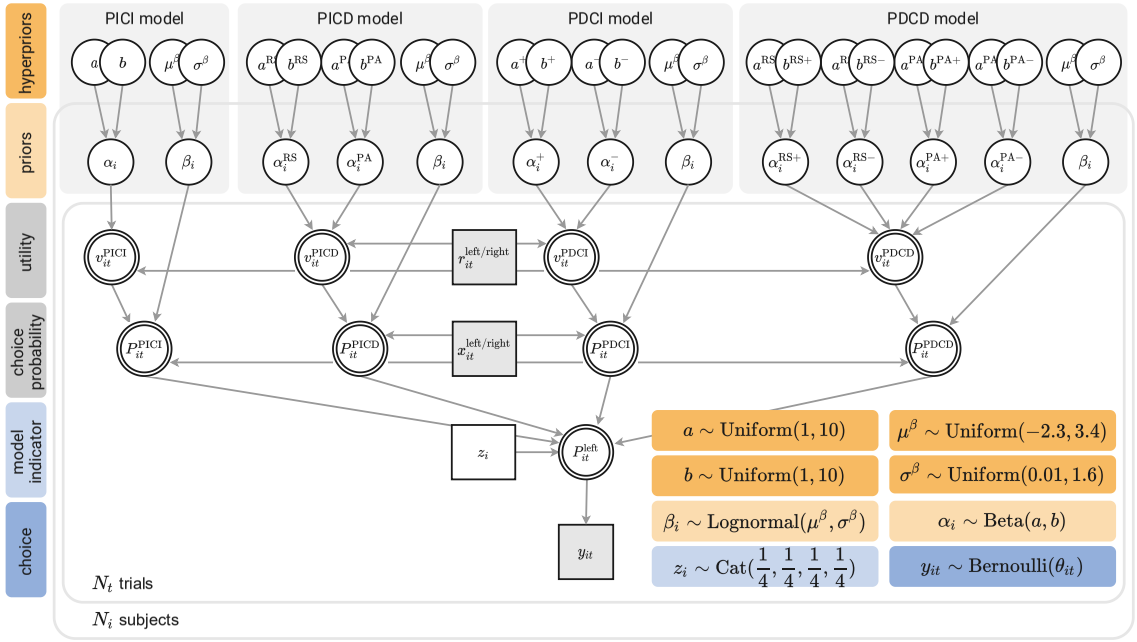

Fig. 4.2 Hierarchical latent-mixture model. Choice probability for the trial t and subject i, is modeled as a latent mixture of choice probabilities calculated for four different reinforcement learning submodels: PICI with single learning rate for each subject, PICD with two separate learning rates for reward-seeking and punishment-avoiding condition, PDCI with two separate learning rates for positive and negative prediction errors and PDCD with four separate learning rates for reward-seeking and punishment-avoiding conditions and positive and negative prediction errors. Probabilities for each submodel are approximated by a subject-dependent model indicator variable z_i. Circular nodes represent continuous variables, square nodes represent discrete variables, unshaded nodes represent unobserved variables, shaded nodes represent observed variables, single border nodes represent stochastic variables, double border nodes represent deterministic variables. Boxes on the left-hand side describe the role of each variable in the model, boxes on the bottom-right present the details of the prior and hyperprior distributions.

Weakly informative hyperpriors were set to model the distribution of the learning rates and precision across individuals. Following previous studies, learning rates were assumed to come from a beta distribution (Gershman, 2016). Beta distribution is defined for the continuous interval $\alpha \in [0, 1]$ covering entire range of possible learning rate values:

$$f(a, b, \alpha) = c_{a,b}\alpha^{a-1}(1-\alpha)^{b-1} \tag{4.8}$$

where a and b are two non-negative shape parameters. These were set to come from the uniform hyperprior distribution $a \sim \text{Uniform}(1, 10)$ and $b \sim \text{Uniform}(1, 10)$. Note that both shape parameters were assumed greater than 1, since for $a < 1$ and $b < 1$ beta distribution takes a behaviorally implausible U-shape. Precision parameter β was independently included for each submodel to provide additional flexibility in model parameter space and avoid potential dependence between competing models. Generative distribution for precision parameter β was represented as a lognormal distribution. Hyperpriors for lognormal parameters distributions were set to $\mu \sim \text{Uniform}(-2.3, 3.4)$ for lognormal group mean and $\sigma \sim \text{Uniform}(0.01, 1.6)$ for lognormal standard deviation. These weakly informative hyperpriors were used in previous studies and derived from the constrain to confine group mean for the precision parameter within 0.1 to ~ 30 interval (Meder et al., 2019; Nilsson et al., 2011). The latent-mixture part of the model assumed a subject-specific model indicator variable to represent a latent mixture of four competing submodels. This assumption allowed to model each subject response pattern independently by using a different mixture of competing submodels. Model indicator variable, z_i, was modeled with uniformly distributed flat prior representing lack of expectations toward which behavioral model best explains subjects response patterns.

4.4.3 Markov Chain Monte Carlo

Typically, the posterior distribution $p(\theta \mid X)$ can only be calculated analytically for a limited set of simple Bayesian models. For more complex models, the analytical approach becomes ineffective and different solutions are needed. An efficient, computer-driven sampling method known as *Markov Chain Monte Carlo* (MCMC) has been developed to overcome this issue (Gamerman and Lopes, 2006; Kass et al., 1998). The MCMC method samples posterior probability distribution $p(\theta \mid X)$ by generating Markov chains. A sufficiently large number of samples in a chain allows accurate reproduction of $p(\theta \mid X)$. Here, the MCMC sampling was performed using JAGS

software (v4.3.0; http://mcmc-jags.sourceforge.net/) called from MATLAB R2017a via `matjags.m` script (v1.3.3; https://github.com/msteyvers/matjags). A sampling procedure with 2000 burn-in samples, four chains, and 15,000 samples per chain was used to estimate the hierarchical latent-mixture model.

To ensure that chain samples are unaffected by starting values and come from a stationary distribution, it is common to measure sampling convergence. A popular diagnostic measure – potential scale reduction, $\hat{R}$, was calculated for each model variable, and corresponding chain traceplots were visually inspected to ensure that the procedure converged. $\hat{R}$ combines information on the variation within and between chains to determine whether all chains reflect the same stationary target distribution. Convergence was declared for $\hat{R}$ values less than 1.1.

For a model selection, a posterior model probability was calculated for each subject using posterior samples of the model indicator variable. PI and PD models were independently compared by marginalizing model indicator variables over subjects and calculating Bayes factors for PICI+PICD (PI model family) models versus PDCI+PDCD (PD model family) models. Bayes factor is a measure of relative evidence of two competing hypotheses calculated as:

$$BF_{12} = \frac{p(\theta \mid \mathcal{H}_1)}{p(\theta \mid \mathcal{H}_0)}, \tag{4.9}$$

where $\mathcal{H}_1$ and $\mathcal{H}_2$ are two competing hypotheses. Here, these hypotheses refer to PI and PD model families generating observed behavioral data, and θ represents the model indicator variable. An analogous procedure was repeated to compare between CI and CD models. A standard interpretation of Bayes factors was used to report levels of evidence (van Doorn et al., 2021).

The single best submodel was selected by calculating estimated model frequencies and protected exceedance probabilities (Rigoux et al., 2014) for the model frequency being highest among other submodels. These calculations were performed with the Variational Bayesian Analysis toolbox (v1.9.2; https://github.com/MBB-team/VBA-toolbox), which implements Bayesian model selection for group studies. Model parameters for the winning model were fitted by estimating a reduced hierarchical latent-mixture model. A full model was reduced to the single hierarchical Bayesian model by removing the model indicator variable and all branches representing competing submodels. Posterior samples for learning rates and precision were used to derive point estimates for these parameters. These point estimates were used in subsequent model-based fMRI analyses.

4.4.4 Behavioral performance

Successful performance in the PRL task requires subjects to take into account changing reward or punishment probabilities and magnitudes associated with each choice. As in any version of the PRL task, probabilities had to be learned from experience by trial-and-error.

An individual subject's performance accuracy was calculated to establish if subjects succeeded in learning probabilistic associations. Accuracy was defined as the proportion of choices that led to rewarding outcomes in reward-seeking condition or non-punishing outcomes in punishment-avoiding condition. Since the study examined healthy subjects without valence-dependent learning deficits, I hypothesized similar performance levels for both task conditions.

The accuracy was significantly higher than a chance level for both reward-seeking ($acc_{\mathrm{RS}} = 62.81\%$; one-sided t-test $t(31) = 13.00$; $p < 0.0001$) and punishment-avoiding condition ($acc_{\mathrm{PA}} = 62.13\%$; one-sided t-test $t(31) = 11.38$; $p < 0.0001$). Subjects performed equally well in both experimental contexts – no significant difference in performance between task conditions was found (two-sample t-test; $p = 0.62$).

I was also interested in whether reward and punishment magnitudes influenced the subject's choices. A significant relationship between reward/punishment magnitude and choice proportion suggests that subjects correctly understood the task structure, considering magnitudes as an essential factor for successful performance. The probability of choosing the right box was estimated for each possible difference in reward/punishment between the right and left boxes. This probability was defined as the proportion of right box choices pooled over subjects and task conditions. A significant correlation between the two variables was found ($r = 0.95$; $p < 0.0001$), indicating that subjects used magnitude information to guide their choices (**Fig. A.1A**).

Next, I wanted to test whether task conditions affected probability matching behavior, which manifests as a loose-shift strategy in a probabilistic learning scenario (Gaissmaier and Schooler, 2008). The number of choice reversals was calculated for each subject and task condition. Subject frequently reversed their choices – mean number of reversals was equal to 30.41 ± 11.93 in reward-seeking and 30.25 ± 11.98 in punishment-avoiding condition. Moreover, the number of choice reversals did not significantly differ between the two task conditions (two-sample t-test; $p = 0.94$).

Finally, I wanted to quantify possible differences in choice duration, i.e., reaction time (RT), between reward-seeking and punishment-avoiding conditions and positive and negative prediction errors. Choice times following gain/no-lose trials were assigned to the +PE group, whereas these following lose/no-gain trials reflected −PEs. A

two-way repeated measures ANOVA (rmANOVA) was performed on RT values with task condition and PE sign as factors. There was a significant interaction effect between task condition and PE sign ($F(31) = 14.06$; $p < 0.001$; **Fig. A.1B**) indicating that choice durations were modulated by both experimental factors. In general, choice durations were longer for the punishment-avoiding compared to reward-seeking conditions as indicated by the significant task condition effect ($\mathrm{RT_{RS}} = 687.4 \pm 126.3$ms; $\mathrm{RT_{PA}} = 651.7 \pm 110.9$ms; $F(31) = 4.27$; $p < 0.05$). Post hoc within-condition analysis showed that choice times following negative prediction errors were longer in punishment-avoiding condition (paired t-test; $t(31) = 3.04$; $p < 0.01$) but not in reward-seeking condition (paired t-test; $p = 0.26$).

4.4.5 Model selection

The hierarchical latent-mixture model with four competing submodels was evaluated to select the winning submodel. A sampling of the probability distribution of the model indicator variable z_i resulted in posterior distribution for each subject reflecting posterior probability for each competing submodel.

A majority of the subjects (27/32; 84.4%) had most of the probability mass located over either the PDCI model (19/32; 59.4%) or PDCD model (8/32; 25%) (**Fig. 4.3A**). Models with separate learning rates for reward-seeking and punishment-avoiding conditions were favored in five subjects (1/32; 3.1%; PICD and 4/32; 12.5% PICI). Posterior distribution of the model indicator variable was marginalized over subjects and model classes to investigate whether subject response patterns are better explained by models with: (1) prediction error dependent (PD) versus prediction error independent (PI) learning rates, and (2) condition dependent (CD) versus condition independent (CI) learning rates. These comparisons reflected two orthogonal experimental axes: prediction error sign and task condition. Moderate evidence was found in favor of the hypothesis that behavioral responses are better explained by the models with separate learning rates for positive and negative prediction errors ($\mathrm{BF_{PD-PI}} = 3.19$; **Fig. 4.3B**). On the other hand, there was a weak evidence against the hypothesis assuming separate learning rates for reward-seeking and punishment-avoiding conditions ($\mathrm{BF_{CD-CI}} = 0.67$; **Fig. 4.3B**). Measures of the likelihood for each submodel being more frequent than all other submodels of the HLM model was computed as protected exceedance probabilities. This procedure determined a single winning submodel for further fMRI analysis. The PDCI model was the most likely model across participants with probability close to 1 (protected exceedance probability; $p = 0.9995$; **Fig. 4.3D**).

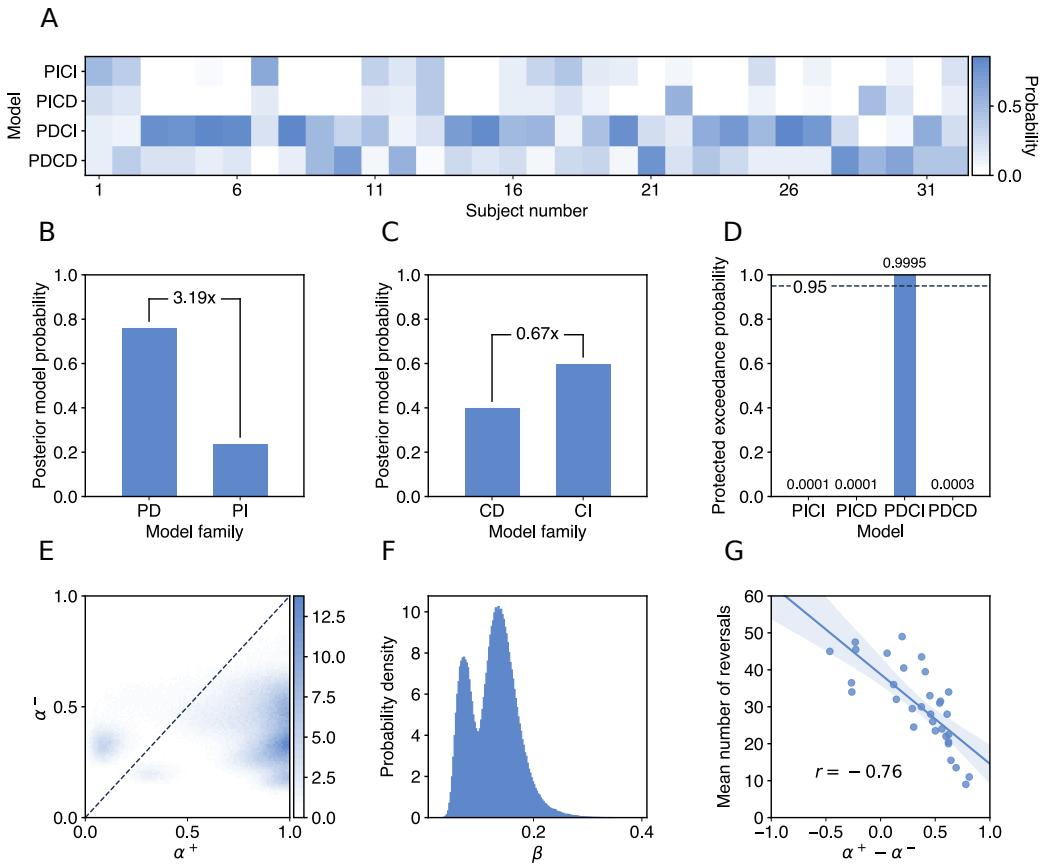

Fig. 4.3 Model selection and parameter recovery results. (**A**) Posterior distribution of the model indicator variable z_i. Most of the subject had the probability mass over either PDCI (19/32) or PDCD (8/32) models. (**B** and **C**) Posterior distribution of the model indicator variable z_i marginalized over subjects and model classes. Model families: prediction error dependent (PD; PDCI and PDCD models); prediction error independent (PI; PICI and PICD models); condition dependent (CD; PICD and PDCD models); condition independent (CI; PICI and PDCI models). Numbers over the bars represent Bayes factors for the hypothesis that behavioral responses are better explained by the models with separate learning rates for (**B**) positive and negative prediction errors or (**C**) reward-seeking and punishment-avoiding conditions. (**C**) Protected exceedance probabilities for each submodel being more frequent than all other submodels. (**E**) Posterior probability distribution of learning rates for positive, α^+, and negative prediction errors, α^-, for the winning PDCI model. The hierarchical model with PDCI submodel as a generative model for subject responses was evaluated independently of the main HLM model. (**F**) Posterior probability distribution of the precision parameter for the winning PDCI model. (**G**) Relationship between the difference in positive and negative learning rates estimated from the PDCI model and a mean number of reversals indicating probability matching behavior.

4.4.6 Parameter recovery

The winning PDCI submodel was independently evaluated as a generative model for subjects' behavioral responses. A hierarchical model comprised of the branch of the full HLM model without the model indicator variable node was created and sampled using the MCMC sampling technique. Posterior distributions of learning rates and precision nodes enabled estimating individual values of these latent cognitive variables.

First, group-level distributions of behavioral parameters were investigated by marginalizing individual distributions over subjects. Learning rates for positive prediction error were higher than learning rates for negative prediction error ($\alpha^+ = 0.737 \pm 0.275$; $\alpha^- = 0.415 \pm 0.149$; **Fig. 4.3E**). A Bayesian hypothesis testing was used to determine whether learning rates for positive prediction errors are higher than for negative prediction errors. Moderate evidence was found in favor of this hypothesis (BF = 4.78). However, after restricting this hypothesis to the 19 subjects with most of the probability mass located over the PDCI model, very strong supporting evidence (BF = 37.90) was found. These results suggest a greater influence of positive than negative prediction errors on value estimates, especially in subjects whose behavior is best explained by the PDCI model.

4.4.7 Relationship between model parameters and behavioral performance

To better understand a behavioral significance of the observed gap between learning rates for positive and negative prediction errors, I investigated a possible relationship between the difference in learning rates and the mean number of reversals. A significant negative correlation between the two variables was found ($r = -0.76$; $p < 0.0001$), indicating that the higher difference between the learning rate for positive and negative prediction error was related to the lower reversal tendency (**Fig. 4.3G**). These results suggest that probability matching behavior is reflected by (1) a high learning rate for positive prediction errors because the subjects are more likely to repeat their choice after the previous correct choice and (2) relatively low learning rate for negative prediction errors because the subjects tend to switch their choice after a previous incorrect choice.

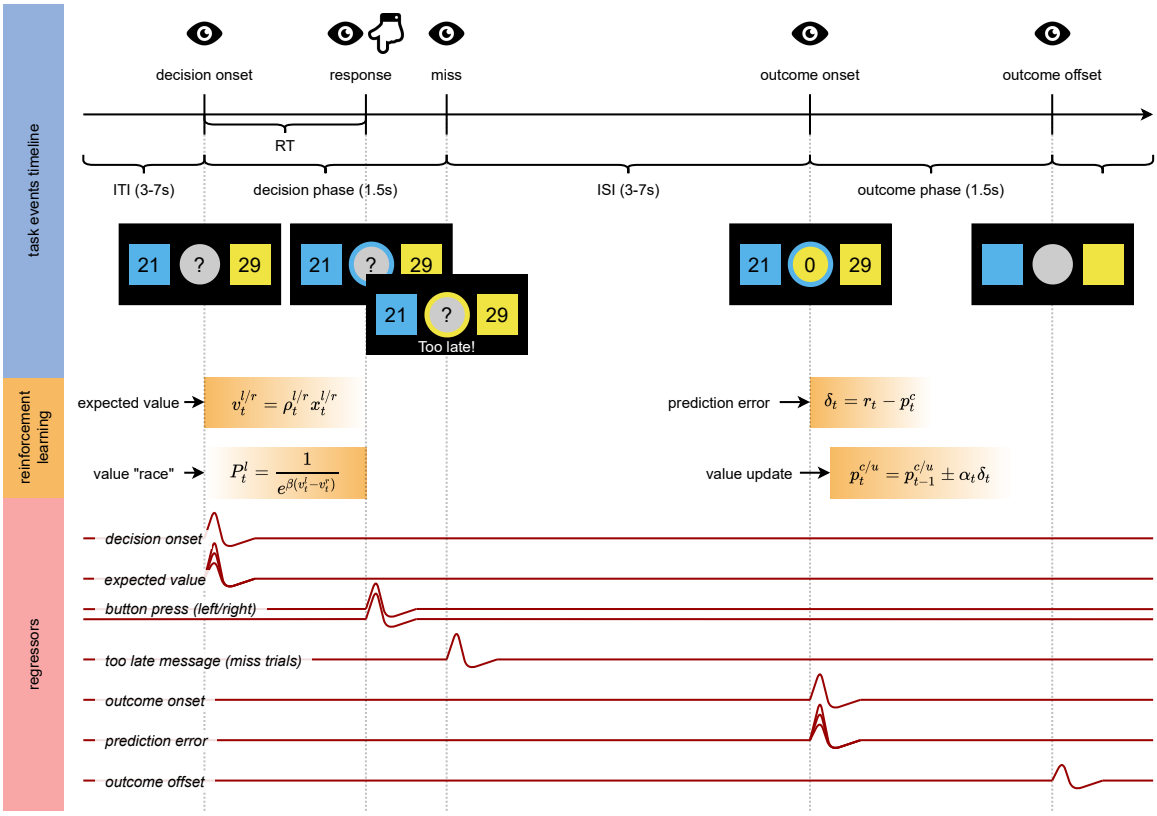

Fig. 4.4 Event model for single trial. Task events, latent computational variables, and GLM regressors are shown on a timeline for single experimental trial. Top panel shows a trial timeline with corresponding visual changes (eye icon) and motor responses (hand icon). Middle panel shows latent reinforcement learning variables like expected value and prediction errors time-locked to task events (see section 4.4.1). Bottom panel presents modeled BOLD responses used in first-level GLM.

4.5 Model-based fMRI

4.5.1 First-level GLM

The single-subject effect of processing increasing prediction errors was modeled with a subject-level general linear model. Two separate GLMs were created for reward-seeking and punishment-avoiding task conditions. Each GLM was composed of regressors modeling task events and estimated latent decision parameters and nuisance regressors accounting for noise components related to head motion and scanner drift. Task regressors of interest included regressor modeling onset of the outcome phase and its parametric modulation with the estimated trial-wise prediction error (PE regressor). Task regressors of no interest modeled other experimental events: onset of the decision and its parametric modulation with the expected probability that chosen option will be

correct, left and right button press, trials with a missing response, offset of the outcome phase (**Fig. 4.4**). Nuisance regressors included 24 head motion parameters to remove residual movement artifacts and cosine functions derived from the cosine drift model to remove low-frequency artifacts. All task events were modeled as impulse functions with zero duration and convolved with a canonical double-gamma hemodynamic response function. First-level contrast was defined as a one-sample t-test for the PE regressor effect against the baseline. Resulting statistical parametric maps were used in the second level of analysis.

4.5.2 Second-level GLM

The second level of the analysis was performed as random-effects analysis separately modeling each task condition and within-subject effects. The significance of the two effects of interest across the group was tested: (1) the combined effect of increasing and decreasing prediction error across both task conditions and (2) difference in response to increasing and decreasing prediction error between reward-seeking and punishment-avoiding conditions. A threshold of $p < 0.0001$ with false discovery rate (FDR) correction was used for the former effect, accounting for its high statistical power compared with the latter effect. This stringent thresholding was imposed to reduce the number of significant clusters and improve the readability of the statistical maps. A threshold of $p < 0.001$ with FDR correction was used for the latter effect. An additional threshold for cluster size of 20 connected voxels was used in either case.

Differences in response to PE between reward-seeking and punishment-avoiding conditions cannot be unambiguously interpreted since statistical testing of this effect relies on testing the difference between slopes of the regression between voxel activity and the prediction error. For example, significant response for increasing PEs in reward-seeking compared to the punishment-avoiding condition may be related to one of the three effects: (1) for both conditions, activity is correlated with increasing PE, but for the reward-seeking condition the relationship is significantly stronger, (2) for both conditions activity is correlated with decreasing PE, but for reward-seeking condition the relationship is significantly weaker or (3) activity is correlated with increasing PE in reward-seeking condition and correlated with decreasing PE in punishment-avoiding condition. A post hoc test on the cluster level was performed to discriminate between these effects. Response profiles were extracted for each subject, task condition, and significant cluster exhibiting increased or decreased response to prediction error in reward-seeking condition compared to punishment-avoiding condition. These profiles were calculated as first-level parameter estimates, i.e., GLM beta values, for peak voxel

for the PE regressor effect. Then, average beta values across subjects were calculated for both conditions. To further confirm the significant change in response profiles, a paired t-test between reward-seeking and punishment-avoiding betas was performed for all clusters.

4.5.3 Context-independent prediction error processing

I wanted to identify brain regions in which activity increases with increasing PEs and decreases with increasing PEs. I used individually calculated prediction errors to construct PE regressors modeling latent PE computation time-locked to the outcome onset. My first aim was to identify brain regions encoding increasing and decreasing PEs regardless of the task condition. These regions form the core of the learning network enabling the processing of positive and negative feedback information used to update values of chosen stimuli.

A broad network of regions in which activity was correlated with increasing and decreasing PEs was found (**Fig. 4.5A**). For increasing PEs, significant activity was found in clusters in the vmPFC, superior temporal gyrus, bilateral orbitofrontal cortex, bilateral putamen, left postcentral gyrus, right amygdala and VS, left posterior cingulate cortex (PCC), left angular gyrus, right precentral gyrus, left middle temporal gyrus (**Table 4.1**; increasing PE). Other significant clusters were located in the left caudal middle frontal gyrus, bilateral planum temporale, left parahippocampal gyrus, and right superior frontal gyrus. For decreasing PEs, significant activity was found in clusters in the left dorsomedial cingulate cortex, bilateral anterior insula (aINS), right pars opercularis, left dorsolateral prefrontal cortex (dlPFC), left cerebellum, left precentral gyrus, and left superior frontal gyrus (**Table 4.1**; decreasing PE). Generally, activity related to increasing PEs was more widespread than activity related to decreasing PEs.

4.5.4 Context-dependent prediction error processing

I wanted to test whether PE signaling in essential parts of increasing PE network like striatum and vmPFC undergo valence-related changes. To investigate differences in PE processing between task conditions, I evaluated a second-level GLM testing for higher response to increasing PEs in reward-seeking than punishment-avoiding conditions.

A set of brain regions was found for which the regression slope between voxel activity and the prediction error was higher in the reward-seeking compared to punishment-avoiding condition (**Fig. 4.5B**). These regions included bilateral visual areas V3 and V4, right supramarginal gyrus, right superior parietal lobule, right precuneus, right

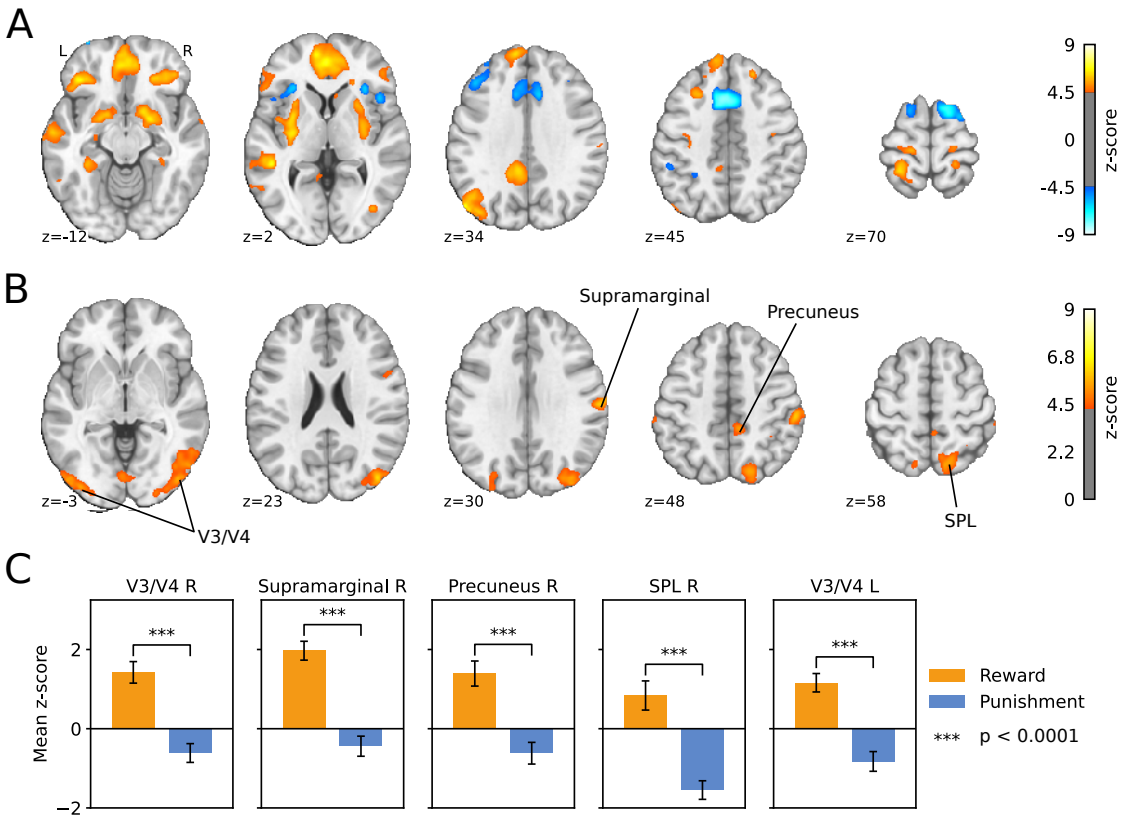

Fig. 4.5 Model-based fMRI results. (**A**) Brain regions processing increasing and decreasing prediction errors. Orange regions exhibit increased activity with increasing prediction error in both task conditions. Blue regions decrease their BOLD response with increasing prediction error in both task conditions. Statistical maps were FDR corrected with a threshold $p < 0.0001$. (**B**) Brain regions responding differently to prediction errors in rewards-seeking and punishment-avoiding conditions. For orange clusters slope between voxel activity and PE was higher in reward-seeking compared to punishment-avoiding conditions. (**C**) Post-hoc test for the clusters in **B** (only five clusters with highest z-scores are shown). First-level parameter estimates (z-scores) for the PE regressor were extracted individually for each cluster, subject, and condition. Each region in **B** increased its activity with increasing PE in reward-seeking condition and decreased its activity with increasing PE in punishment-avoiding condition (paired t-test; $p<0.0001$). Abbreviations: SPL - superior temporal lobule.

Table 4.1 Condition-independent PE signaling. Regions signaling increasing and decreasing prediction errors across reward-seeking and punishment-avoiding conditions.

Region	**L/R**	**X**	**Y**	**Z**	**Z-score peak**	**Cluster Size (mm^3)**
Increasing PE						
Ventromedial Prefrontal Cortex	L	-3	48	-1	7.23	30397
Superior Temporal Gyrus	L	-48	-39	3	6.93	1417
Lateral Orbitofrontal Cortex	L	-39	33	-12	6.79	5071
Putamen	L	-30	-12	6	6.63	8473
Postcentral Gyrus	L	-27	-42	66	6.62	1732
Amygdala, Ventral Striatum, Putamen	R	21	0	-12	6.52	8347
Posterior Cingulate Cortex	L	-6	-45	38	6.47	9985
Angular Gyrus	L	-54	-75	34	6.32	9009
Precentral Gyrus	R	24	-24	59	6.30	2803
Middle Temporal Gyrus	L	-60	-12	-15	6.18	2425
Caudal Middle Frontal Gyrus	L	-27	24	48	6.07	1039
Lateral Orbitofrontal Cortex	R	27	36	-12	6.00	3213
Planum Temporale, Supramarginal Gyrus	R	45	-36	17	5.86	3244
Parahippocampal Gyrus	L	-33	-39	-12	5.80	1417
Precentral Gyrus	L	-15	-27	73	5.69	1039
Superior Frontal Gyrus	R	21	39	55	5.69	1386
Planum Temporale	L	-57	-33	17	5.69	1669
Middle Temporal Gyrus	L	-57	-57	-1	5.39	850
Precentral Gyrus	L	-6	-18	52	5.39	787
Postcentral Gyrus	L	-39	-24	52	5.28	913
Decreasing PE						
Dorsomedial Cingulate Cortex, SMA	L	-6	12	48	8.18	14647
Anterior Insula	L	-33	21	10	6.71	1543
Anterior Insula	R	30	24	6	6.43	756
Pars Opercularis	R	45	18	3	5.93	787
Dorsolateral Prefrontal Cortex	L	-45	30	34	5.89	2803
Cerebellum	L	-33	-60	-26	5.81	630
Precentral Gyrus	L	-30	-3	59	5.62	913
Superior Frontal Gyrus	L	-18	6	69	5.55	882

Table 4.2 Between-condition differences in PE signaling. Regions exhibiting differences in prediction error processing between reward-seeking and punishment-avoiding conditions.

Region	L/R	X	Y	Z	Z-score peak	Cluster Size (mm3)
Increasing PE Reward-seeking > Punishment-avoiding						
V3 / V4	R	42	-81	24	6.28	13608
Supramarginal Gyrus	R	57	-27	52	6.22	3843
V3 / V4	L	-48	-81	-1	5.64	3402
Superior Parietal Lobule	R	21	-72	45	5.60	4410
Precuneus	R	12	-39	52	5.56	1071
V1	R	3	-78	3	5.42	1921
Precentral Gyrus	R	54	9	17	5.29	693
Increasing PE Punishment-avoiding > Reward-seeking						
no significant clusters						

primary visual cortex, and right precentral gyrus (**Table 4.2**). None of these regions were part of the core PE processing network. Since differences in response to PE between reward-seeking and punishment-avoiding conditions cannot be unequivocally interpreted, a posthoc test was performed for all significant clusters by extracting individual subject and condition response profiles as first-level parameter estimates for the peak voxel. Positive activity scaling with increasing PEs in reward-seeking condition and negative activity scaling with increasing PEs in punishment-avoidance condition was found in all seven clusters (paired t-test; $p<0.0001$; **Fig. 4.5C**). There were no brain regions with significantly higher response to increasing PE in punishment-avoiding condition.

4.6 Analysis of functional brain networks

4.6.1 Brain parcellation

One of the first steps in brain network analysis is dividing the brain into regions that will be considered nodes of the functional network (Sporns, 2013). Each specific division is called *brain parcellation* or *brain atlas*. Choice of the parcellation can greatly influence brain network analysis results. Therefore it is critical to choose a suitable parcellation for a given set of research questions. Brain parcellations are usually defined based on anatomical, functional, or multi-modal properties of the brain tissue (Glasser et al., 2016). Many existing parcellations suggest that there is still no gold-standard

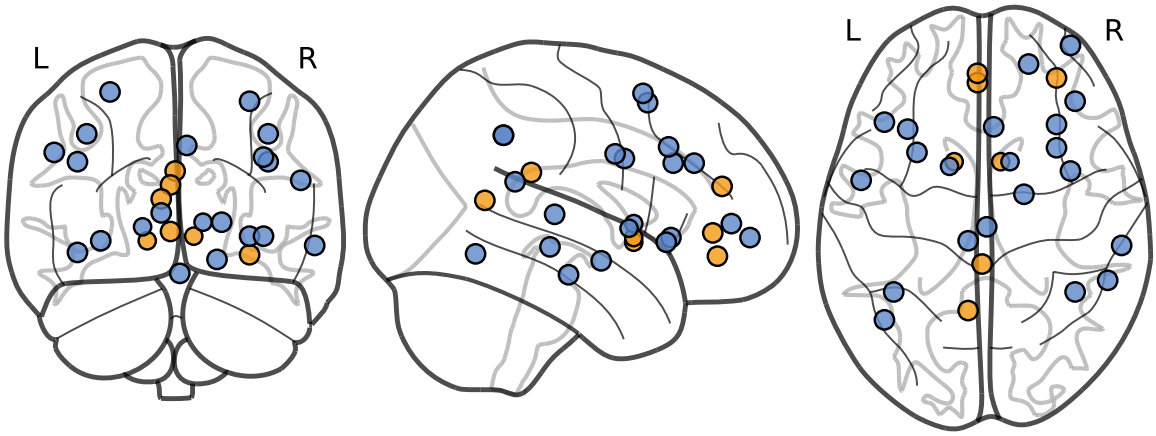

Fig. 4.6 Prediction error signaling ROIs. Regions of interest created from activation likelihood estimation (ALE) meta-analysis clusters from Fouragnan et al. (2018). Network signaling positive prediction errors (+PE; orange spheres) is created from regions with higher BOLD activation for positive than negative PEs. Network signaling negative prediction errors (−PE; blue spheres) is created from regions with higher BOLD activation for negative than positive PEs.

approach to the brain division among the neuroscientific community (Eickhoff et al., 2018). Nonetheless, some brain atlases gained more popularity in the recent decade.

The frequently used brain parcellation in resting-state and task-based fMRI studies is the parcellation introduced by Power et al. (2011). It consists of 264 regions of interest defined based on a meta-analysis of activation patterns from multiple studies. Each ROI is modeled as a 5mm sphere around the specific MNI coordinates. Power's parcellation additionally includes the division of the 264 ROIs into 13 large-scale networks: auditory, cerebellar, somatomotor, cingulo-opercular, default mode, memory, ventral attention, dorsal attention, fronto-parietal, salience, subcortical, uncertain and visual. This division allows examining the dynamic interplay between LSNs in various experimental conditions. However, Power's parcellation does not contain separate prediction-error-related networks; therefore, it cannot be directly used to study how the reward system interacts with the rest of the brain during reinforcement learning. To overcome this issue, the Power's ROIs were combined with regions from the recent meta-analysis on neural representations of prediction error valence (Fouragnan et al., 2018). This procedure allows studying effects specific for networks signaling prediction errors while still preserving reference to well-known large-scale brain systems. Moreover, a recent meta-analysis provided evidence that reward-related brain regions form a stable network observed even during rest (Huckins et al., 2019).

Table 4.3 Prediction error signaling ROIs created from ALE clusters from Fouragnan et al. (2018).

Region	R/L	X	Y	Z	r(mm)	Original network	Strategy
ROIs signaling positive prediction errors (+PE)							
Ventrolateral orbitofrontal cortex	R	32	44	-10	5	Uncertain	shifted
Ventral striatum	L	-12	8	-4	4	-	created
Posterior cingulate cortex	L	0	-36	26	5	Memory	shifted
Medial prefrontal cortex	L	-2	46	20	5	Default mode	shifted
Ventral striatum	R	8	8	-2	4	-	created
Dorsomedial prefrontal cortex	L	-6	-56	14	5	Default mode	shifted
Ventromedial prefrontal cortex	L	-2	42	0	5	Default mode	shifted
ROIs signaling negative prediction errors (−PE)							
Anterior insula	L	-32	22	-4	5	Salience	shifted
Middle frontal gyrus	R	32	14	56	5	Fronto-parietal	relabelled
Inferior parietal lobule	R	40	-48	42	5	Fronto-parietal	shifted
Superior temporal sulcus	R	54	-43	22	5	Ventral attention	relabelled
Fusiform area	L	-42	-60	-9	5	Dorsal attention	relabelled
Pallidum	R	12	8	4	4	Subcortical	shifted
Dorsolateral prefrontal cortex	R	40	34	30	5	Salience	shifted
Dorsolateral prefrontal cortex	L	-42	25	30	5	Fronto-parietal	relabelled
Middle temporal gyrus	R	60	-28	-6	5	Default mode	shifted
Inferior parietal lobule	L	-38	-48	42	5	Dorsal attention	shifted
Thalamus	L	-6	-26	8	5	Subcortical	shifted
Anterior insula	R	32	24	-2	5	Salience	shifted
Pallidum	L	-14	6	2	4	Subcortical	shifted
Dorsomedial prefrontal cortex	R	20	50	4	5	-	created
Dorsomedial orbitofrontal cortex	R	38	58	-2	5	-	created
Precentral cortex	L	-52	0	34	5	-	created
Habenula	R	2	-20	-18	5	-	created
Amygdala	R	18	-6	-12	5	-	created
Middle frontal gyrus	R	38	4	32	5	-	created
Middle frontal gyrus	L	-28	12	60	5	Fronto-parietal	shifted
Dorsomedial cingulate cortex	R	5	23	37	5	Salience	relabelled

The following procedure was employed to create extended brain parcellation. First, two prediction-error-related networks were created based on cluster centers for the valence analysis in a recent meta-analysis on neural correlates of prediction errors (Table 2 in Fouragnan et al. (2018)). The network signaling positive prediction error (+PE) was created from regions exhibiting higher BOLD activation for positive than negative prediction errors (pattern A (ii) in Fouragnan et al. (2018)). Conversely, the network signaling negative prediction error (−PE) consisted of regions with a higher BOLD response for negative than positive prediction errors (pattern A(i) in Fouragnan et al. (2018)). Second, for each cluster center, a new ROI was created using one of the three strategies based on its distance to the closest ROI in Power's atlas (d):

- If $d > 10$mm, there was no overlap between the 5mm sphere created at the cluster center and existing ROIs, so a new ROI was created and added to the ROIs set without any modification of the existing base parcellation.
- If $10\text{mm} > d > 5$mm, there was a slight overlap between the 5mm sphere created at the cluster center and the closest Power's ROI (no more than 31% of sphere volume overlapped). In this case, the closest Power's ROI was shifted into the location of the ALE cluster center and renamed according to the assignment to one of the prediction-error-related networks.
- If $d < 5$mm there was a significant overlap between the 5mm sphere created at the cluster center and the closest Power's ROI (more than 31% of sphere volume overlapped). In this case, the original Power's ROI was retained at its original location and relabelled according to the assignment to one of the prediction-error-related networks.

All newly created ROIs were initially modeled as 5mm spheres around specific MNI coordinates. Radi of four striatal ROIs – left and right pallidum and left and right ventral striatum were decreased to 4mm to avoid overlap between spheres (**Table 4.3**). The extended brain parcellation consisted of 272 ROIs divided into 15 large-scale networks (13 LSNs from Power et al. (2011) and 2 PE networks). Four ROIs from the uncertain network – MNI coordinates (-31, -10, -36), (-56, -45, -24), (8, 41, -24), and (52, -34, -27) – were further excluded due to signal dropout in some participants. The final parcellation consisted of $N_{\text{ROI}} = 268$ regions of interest.

4.6.2 Network construction

Task-related functional connectivity was estimated using a beta-series correlation method introduced by Rissman et al. (2004). The method's name comes from the GLM parameter beta representing linear coefficient reflecting the contribution of modeled signal into an observed BOLD response. The BSC method quantifies event-to-event fluctuations in the activity of different brain areas. These fluctuations allow estimating statistical dependency between regional activations for different event types.

In BSC, each trial is modeled separately in GLM. Then, a series of beta maps representing brain activations for a series of events was used to calculate condition-specific connectivity evoked by the task. An alternative methodology to calculate task-evoked functional connectivity is psychophysiological interaction analysis (Friston et al., 1997). However, it has been suggested that the BSC method can have more statistical power than the PPI method when applied to event-related designs with many trials and short event durations (Cisler et al., 2014). The "single-trial-versus-other-trial" approach introduced by Mumford et al. (2012) was used to compute activation beta maps. This approach assumes creating a separate GLM for each trial in which the trial of interest is modeled as a separate regressor, and all other trials are modeled as a single nuisance regressor. Simulations have shown that the "single-trial-versus-other-trial" approach produces more accurate estimates of trial-wise activation patterns (Mumford et al., 2012).

For each trial, a separate GLM was created. The GLM consisted of:

- One regressor modeling a trial of interest as an event of duration 0s (typically used to model very short cognitive processes) time-locked to the onset of the outcome phase, convolved with a standard SPM hemodynamic response function (HRF).

- Two regressors modeling the rest of the trials of interest as events of duration 0s time-locked to the onset of the outcome phase, convolved with an HRF. One regressor modeled trials with positive prediction errors (gain in reward-seeking condition or no-loss in punishment-avoiding condition). The other modeled trials with negative prediction errors (no-gain in reward-seeking condition or loss in punishment-avoiding condition).

- Six regressors modeling the events of no interest: decision onset, decision onset modulated with expected probability for side for being correct, decision onset for missed trials, left button press, right button press, and decision offset for missed trials.

- Twenty-four head motion parameter regressors modeling motion-related noise.
- Cosine functions modeling low-frequency scanner drift.

The GLM was high-pass filtered with the cut-off frequency $\frac{1}{128}$Hz and fitted to participants data using `FirstLevelModel` class from the Nistats package. The output of this procedure consisted of a three-dimensional beta map for each trial, indicating the voxel-wise estimate of neuronal activity evoked during this trial (**Fig. A.2**).

After GLM estimation, beta values for each trial were averaged in each of 268 ROIs producing 268 beta-series per subject and task condition. Beta-series were then z-transformed and separated by the sign of the prediction error of each trial. This procedure gave four beta-series for each participant: +PE trials in reward-seeking condition, −PE trials in reward-seeking condition, +PE trials in punishment-avoiding condition, and −PE trials in punishment-avoiding condition. Beta-series were then correlated using Pearson's correlation coefficient, generating a $N_{\mathrm{ROI}} \times N_{\mathrm{ROI}}$ symmetrical correlation matrix for each subject, task condition, and trial-wise prediction error sign. Correlation values were then converted to z-scores using Fisher z-transform. Transformed matrices represented the final subject-level estimate of functional connectivity pattern evoked by the specific task condition and type of trial.

4.6.3 Structural resolution parameter selection

The functional brain network is known to exhibit multi-scale community structure (Betzel and Bassett, 2017). Most studies on the topological organization of functional brain networks, however, have examined modular network structure at a single scale (Gu et al., 2015). Different topological scales of a network represent levels of organization at which one can examine a given network. For example, smaller topological scales would reflect the relationship between individuals and smaller groups of friends in social networks. In contrast, larger topological scales would reflect effects observed at scales of hundreds or thousands of individuals and larger populations. Therefore, exploring a network organization across a range of possible scales allows capturing effects that might have been overlooked in a single-scale analysis. For example, it has been shown that resting-state networks display interaction between age and scale, i.e., larger communities become more segregated with age, while smaller communities exhibit an opposite effect (Betzel et al., 2015).

A simple approach to multi-scale modularity analysis is to introduce a structural resolution parameter, γ, into modularity function:

$$Q(\gamma) = \sum_{ij}(A_{ij} - \gamma P_{ij})\delta(\sigma_i\sigma_j)\,, \tag{4.10}$$

where A_{ij} is the connection strength between nodes i and j, P_{ij} is the expected connection strength according to appropriate null model, σ_i indexes the community to which node i is assigned, and δ is the Kronecker delta function. Tuning structural resolution parameter allows uncovering communities of different sizes, otherwise mathematically undetectable using standard modularity function (Fortunato and Barthelemy, 2007; Reichardt and Bornholdt, 2006). Specifically, lower values of γ result in fewer larger communities, and higher values produce more and smaller communities. At the lower end of the γ range, all network nodes are placed within one community, whereas at the higher end, there are as many communities as there are nodes. Different heuristics have been proposed to find an appropriate sampling of structural resolution parameter space. For example, similarity measures between partitions have been used to minimize the variability in the results of modularity maximization techniques (Doron et al., 2012; Lancichinetti et al., 2009). Here, I used heuristics based on the similarity between data-driven partitions and *a priori* reference partition described in the section 4.6.1. This choice has been guided by the intention to describe network reorganization with reference to well-known LSNs. Two constraints have been imposed on the analyzed topological scales:

1. Number of detected modules and mean module size should be similar to the reference partition. In other words, large γ values generating partitions with many *singletons*, i.e., single node communities, should be excluded from the analysis.

2. Detected communities should exhibit high similarity to the reference partition while preserving a high degree of inter-individual variability. High similarity to the reference partition justifies the interpretation of the results in terms of interactions between *a priori* LSNs. High inter-individual variability ensures that the analysis will be more sensitive to topological changes between task conditions and prediction error signs.

To calculate community structure properties across topological scales, the structural resolution parameter space was sampled by varying γ from 0.05 to 3 with a step of 0.05. For each value of γ, the Louvain modularity maximization algorithm was used to quantify optimal network structure (Blondel et al., 2008). Negative connections were separately included in the modularity function and treated asymmetrically as suggested in Rubinov and Sporns (2011):

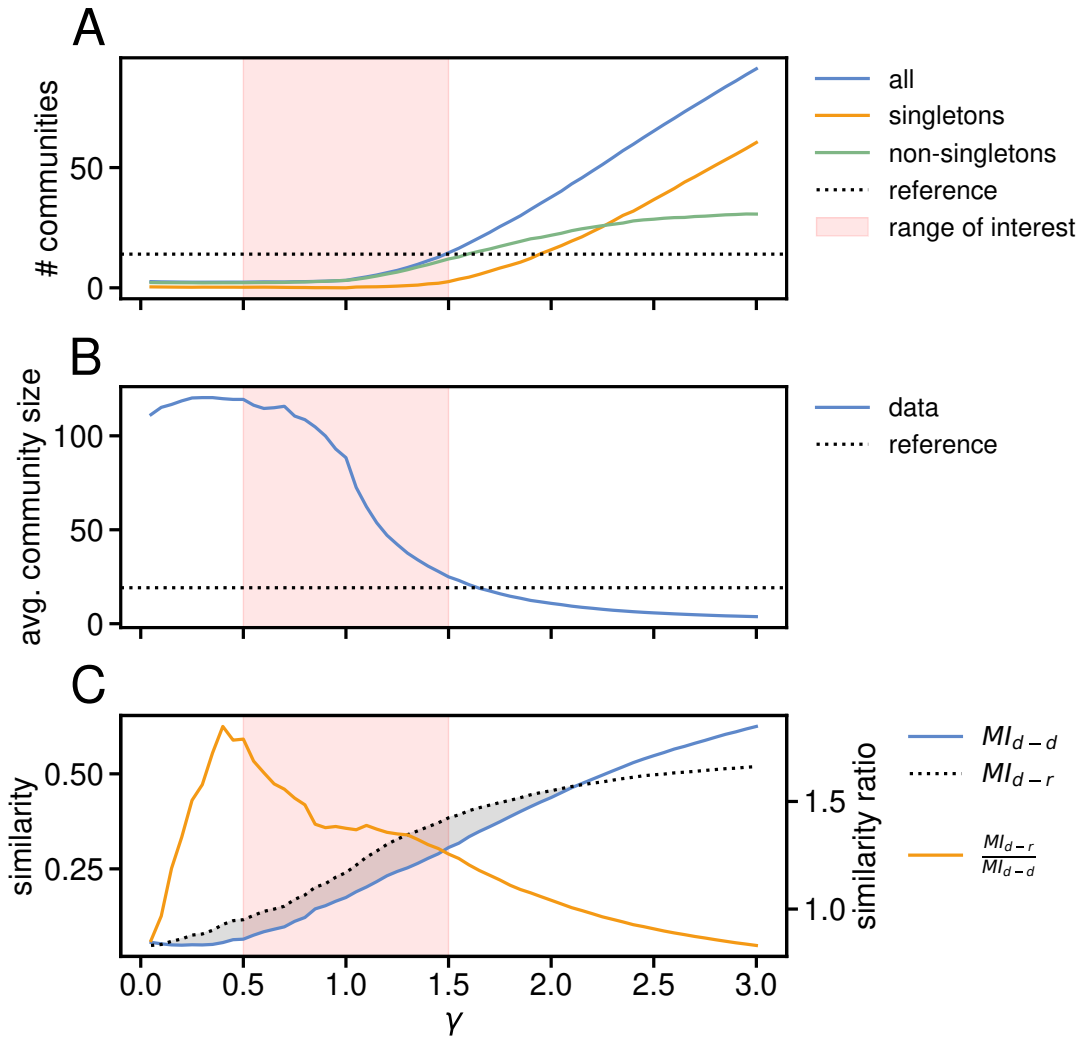

Fig. 4.7 Structural resolution parameter analysis. Structural resolution parameter, γ, was varied from 0.05 to 3 with a step of 0.05. For each value of γ, optimal community structures for empirical networks were calculated using 100 repetitions of Louvain algorithm (Blondel et al., 2008). (**A**) Number of all communities, singleton communities and non-singleton communities. Reference partition based on Power ROIs described in section 4.6.1 has 14 communities. (**B**) Mean data-driven community size compared with average size of 19.14 ROIs per community for reference partition. (**C**) Average similarity between data-driven partitions and between reference and data-driven partitions. Gray area indicates γ range for which partitions are more similar to the reference than to themseleves. Orange line reflects the optimal trade-off between high similarity to reference partition and high inter-individual variability in data-driven partitions. MI_{d-d} - average mutual information between all pairs of data-driven partitions; MI_{d-r} - average mutual information between data-driven partitions and reference partition.

$$Q^{asym}(\gamma) = Q^{+}(\gamma) + \frac{v^{-}}{v^{-} + v^{+}} Q^{-}(\gamma)\,, \tag{4.11}$$

where $Q^{\pm}(\gamma)$ are modularity values from (4.10) calculated separately for positive and negative connections, and $v^{\pm} = \sum_{ij} A^{\pm}_{ij}$ are total connection strengths, i.e., *costs* for positive and negative connections. For each functional network, the Louvain algorithm was run 100 times, and the partition with the highest value of $Q^{asym}(\gamma)$ was kept for further analysis. Several metrics were computed to inspect different topological scales (**Fig. 4.7**): (1) a mean number of communities, singletons, and non-singletons (communities with more than one node), (2) mean community size, (3) mean similarity between data-driven partitions, and (4) mean similarity between reference partition and data-driven partitions. Partition similarity was computed as the normalized mutual information between community partition vectors (Meilă, 2007).

The mean number of data-driven communities was closest to 14 reference communities for $\gamma = 1.5$ (average 14.59 communities per partition). For $\gamma > 1.5$, there was a steep increase in the number of singletons, whereas the number of non-singletons rose to 21.85 for $\gamma = 2$ and plateaued with further increase of structural resolution (**Fig. 4.7A**). The average data-driven community size was above 100 nodes per partition for $0 < \gamma < 0.9$ and rapidly declined for $\gamma \sim 1$. For $\gamma = 1.65$, the average community size was closest to the average reference community size (18.86 nodes per community for data-driven partitions; 19.14 nodes per community for reference partition) (**Fig. 4.7B**). Similarity between partitions measured as normalized mutual information increased as a function of γ (**Fig. 4.7C**). For $0.15 \leq \gamma \leq 2.1$, data-driven partitions exhibited higher similarity to themselves than to reference partition. The ratio between data-driven partition similarity, MI_{d-d}, and similarity between reference and data-driven partition, MI_{d-r}, was highest for $\gamma = 0.4$ ($\frac{MI_{d-d}}{MI_{d-r}} = 1.85$).

According to the first heuristic, the analyzed γ range should include structural resolutions for which the number of communities and the average size of the community resembles characteristics of the reference partition. This consideration suggests including $\gamma \sim 1.5$ since the size and number of data-driven communities were closest to the size and number of reference communities for this topological scale. On the other hand, the second heuristic suggests including $\gamma \sim 0.5$ because partitions observed for this topological scale exhibit the optimal similarity to the reference partition while preserving a high degree of inter-individual variability. For $\gamma > 2$ number of singletons starts to dominate non-singleton communities, and relative inter-individual variability decrease compared to similarity to reference partition, suggesting that this part of the structural resolution scale should be excluded from further analysis. Considering

both heuristics, the final range $0.5 \leq \gamma \leq 1.5$ was adopted for all subsequent analyses. Whereas dense sampling of this range (e.g., with a step of 0.05) would allow tracing how observed effects smoothly appear and disappear when changing a topological resolution, it would also be computationally expensive and challenging to describe statistically. Therefore, a sparse sampling of γ range was employed, enabling to focus on three distinct topological scales: $\gamma = 0.5$ representing the largest scale of few "super-communities" consisting of multiple LSNs, $\gamma = 1$ representing most frequently studied intermediate scale, and $\gamma = 1.5$ representing finer scale approximately resembling well-known division into LSNs.

4.6.4 Network modularity and community structure

To assess subject-level modularity and community structure for both task conditions and trial-wise prediction errors signs, I employed a similar procedure to the one used during structural resolution parameter analysis (section 4.6.3). For each value of $\gamma \in \{0.5, 1, 1.5\}$ and each functional network, the Louvain modularity maximization algorithm was executed 1000 times, and the output of the run with the highest modularity was saved for further analysis (**Fig. 4.8**). The output consisted of weighted modularity with the asymmetric treatment of negative weights, Q^{asym} (defined in equation 4.11; called Q in subsequent analysis), and a $N_{\mathrm{ROI}} \times 1$ community affiliation vector, M, whose i-th element indexed community affiliation of node i.

4.6.5 Consensus partitioning

A *consensus partition* represents a modular structure of functional brain network during the specific experimental condition. Group-level community structure represented by consensus partition can be calculated from subject-level community affiliation vectors. Consensus partitions effectively decrease data dimensionality, allowing a qualitative description of the network reconfiguration. There are two popular approaches for defining a consensus partition: (1) selecting a partition with the highest similarity to the other partitions (Doron et al., 2012), and (2) reclustering an association matrix that stores the information about pairwise nodal co-occurrence within the same community (Bassett et al., 2013; Betzel et al., 2017; Lancichinetti and Fortunato, 2012). Here, the latter approach based on the concept of *agreement* (or *module allegiance*) was used. The agreement, D_{ij}, is a measure defined for the pair of nodes, i and j, and characterizes the extent to which these nodes belong to the same community (Bassett et al., 2015; Bertolero et al., 2015):

$$D_{ij} = \frac{1}{N} \sum_{k=1}^{N} a_{ij}^{k} \,, \tag{4.12}$$

where N is the number of partitions, and a_{ij}^{k} equals 1 if nodes i and j belong to the same community, and 0 otherwise. The elements of the agreement matrix, D_{ij}, reflect the probability that nodes i and j are found within the same community for randomly selected partition given a set of partitions. Despite that information represented by the agreement matrix is not independent of the information contained in the connectivity matrix, an agreement can provide a complementary characteristic of nodal associations. For example, two nodes can share a weak direct connection but many strong indirect connections. In this case, there is a high chance that they will be placed within the same community exhibiting significant agreement and low connectivity at the same time. Furthermore, Bassett et al. (2015) has shown that agreement reduces the noise in connectivity matrices and is more sensitive to differences in network organization compared with information contained in connectivity values alone.

Representative modular structure for task conditions and prediction error signs was calculated in four steps (**Fig. 4.8**). First, subject-level community affiliation vectors were aggregated into four groups: reward-seeking, punishment-avoiding, positive PE, and negative PE. Each group consisted of two vectors per subject, i.e., 58 in total. For example, a reward-seeking group consisted of partitions for both +PE and −PE networks in reward-seeking condition. Similarly, a positive PE group consisted of +PE networks pooled over reward-seeking and punishment-avoiding conditions. This pooling procedure allowed to examine representative community structure independently for both experimental dimensions. Additionally, the fifth group consisting of partitions from all task conditions was created. It consisted of four vectors per subject, i.e., 116 in total. This group was included to represent a modular network structure during prediction errors processing that is stable across task conditions. Second, an agreement matrix was computed for each group. Third, agreement matrices were thresholded to remove weak associations, i.e., all agreement values below $\tau = 0.5$ were set to zero. Finally, thresholded agreement matrices were iteratively reclustered with the Louvain algorithm until convergence, and the output affiliation vectors were considered consensus partitions. This procedure yielded four condition-specific and one condition-independent community vector. The consensus partitioning procedure was independently repeated for three values of structural resolution parameter γ. Agreement matrices and consensus partitions were calculated using `agreement` and

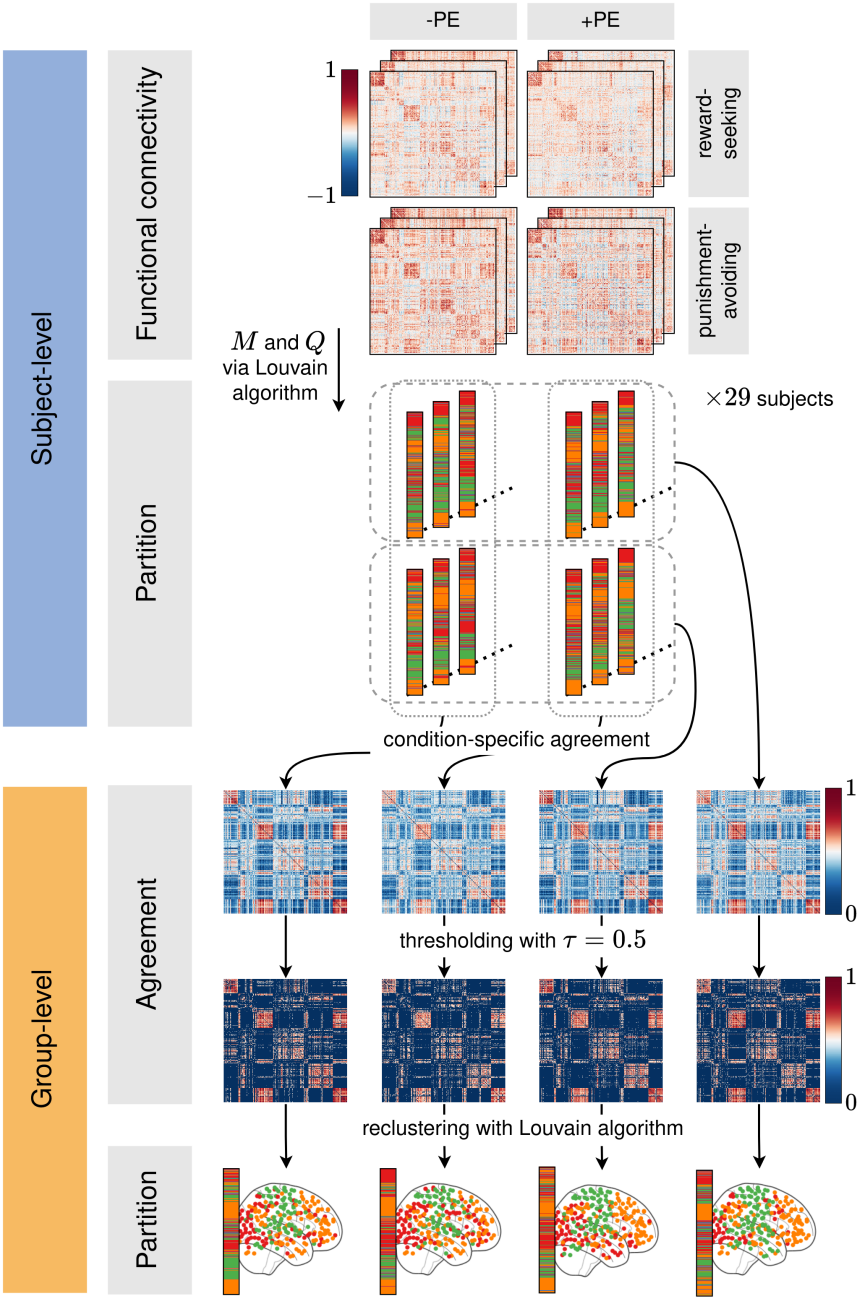

Fig. 4.8 Consensus partitioning pipeline. Multi-step procedure employed to calculate representative network partitions for task conditions and prediction error signs. Subject-level connectivity matrices were first clustered via Louvain algorithm to calculate network modularity, Q, and community affiliation vectors, M (Blondel et al., 2008). Then, subject-level partition were grouped into four groups: reward-seeking, punishment-avoiding, positive PE, and negative PE. For each group, agreement matrix was calculated, thresholded and reclustered yielding final condition-specific consensus partition.

`consensus_und` functions from bctpy package (version 0.5.2) based on the Brain Connectivity Toolbox (Rubinov and Sporns, 2010).

4.6.6 Large-scale network agreement

The brain network continuously adapts its modular architecture to satisfy the demands of various cognitive processes (Bassett et al., 2011; Braun et al., 2015). Therefore, it is essential to explore the pattern of condition-related changes in modular network organization.

I used an approach based on within and between-community agreement to investigate whether the integration and segregation of large-scale networks depend on the outcome valence and prediction error sign. This community-level approach relies on averaging node-level agreement, D_{ij}, derived from data-driven partitions using a reference partition. The reference partition can be either a consensus partition or *a priori* partition representing a stable community structure (Conrad et al., 2020). To better reflect the underlying structure of the data, the condition-independent consensus partitions were used as reference partitions. Results from the extended Power partition (see section 4.6.1) used as a reference partition are showed in **Fig. A.4**.

Community-level agreement values capture the level of association between LSNs. When calculated for a single partition, they represent a subject-level pattern of integration and segregation of different brain systems. They have a clear and straightforward probabilistic interpretation:

- Within-community agreement express the probability that two randomly selected nodes from a given LSN will be found within the same data-driven community. For example, a within-community agreement of 1 would reflect that all LSNs nodes are a part of the same data-driven community. High values of within-community agreement indicate stable and segregated LSN, whereas lower values characterize more unstable and fragmented systems.

- Between-community agreement reflects the probability that a randomly selected pair of nodes from two different LSNs will share the same data-driven community. In most extreme cases, the between-community agreement can be 0 when all nodes from both LSNs belong to different data-driven communities or 1 when both LSNs are merged into a larger community in a data-driven partition. Between-community agreement expresses the extent of integration between LSNs.

To investigate condition-specific changes in the community-level agreement, individual community affiliation vectors described in section 4.6.4 were used to calculate

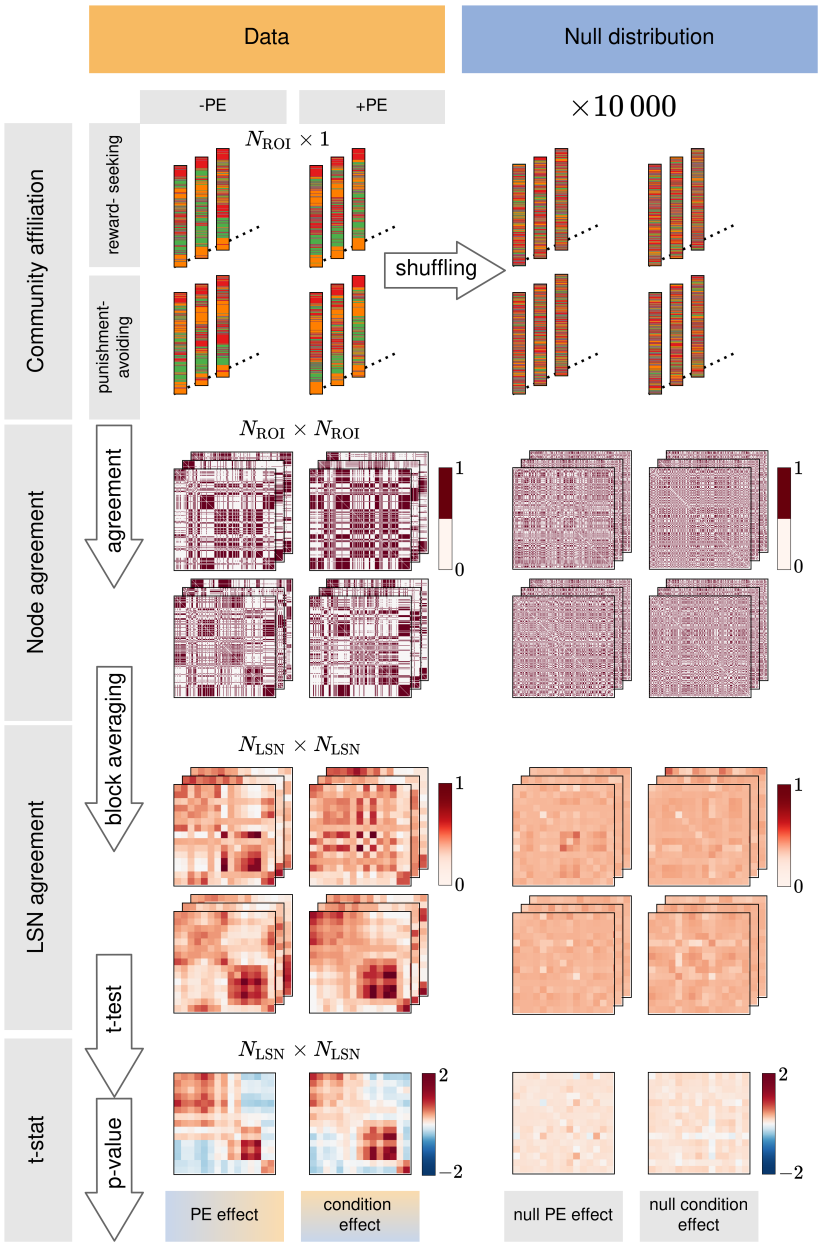

Fig. 4.9 LSN agreement pipeline. Multi-step procedure employed to calculate condition-specific changes in the community-level agreement. Subject-level community affiliation vectors were used to calculate node-level agreement matrices, which were block averaged to produce individual LSN agreement matrices. LSN agreement represents the probability that two randomly selected nodes from reference LSNs belong to the same data-driven community. For each entry in the LSN agreement matrix, two t-tests were conducted testing the effect of task-condition and prediction error sign. T-test significance was tested using the Monte Carlo permutation procedure. Shuffled community affiliation vectors were randomly reordered 10,000 times and subjected to the same analysis procedure as original vectors to produce a null distribution for both t-statistics. P-values were assigned according to the position of the true t-statistic within the null distribution.

individual node-level agreement matrices (**Fig. 4.9**). An individual node-level agreement is an $N_{\mathrm{ROI}} \times N_{\mathrm{ROI}}$ matrix with entries $D_{ij} \in \{0, 1\}$ indicating whether a pair of nodes belongs to the same data-driven community. These node-level matrices were then block-averaged according to the network division into LSNs in condition-independent consensus partition to create individual LSN agreement matrices:

$$D_{mn}^{\mathrm{LSN}} = \frac{1}{NM} \sum_{i,j=1}^{N_{\mathrm{ROI}}} D_{ij} \delta_i^m \delta_j^n \,, \tag{4.13}$$

where N and M are the sizes of communities n and m, and δ_i^m is a community indicator equal to 1 if node i belongs to community m and 0 otherwise. Then, for each community-level agreement value, D_{mn}^{LSN}, a two-sided paired t-test was conducted for task condition and prediction error sign effects. The null hypothesis states that the degree of LSN agreement is the same in reward-seeking and punishment-avoiding conditions for a task condition effect. For a prediction error sign effect, the statistical test aims to answer if there is a significant change of LSN agreement when switching between signaling positive and negative prediction errors.

To determine the significance of the computed t-statistics, a Monte Carlo permutation procedure was employed (Conrad et al., 2020). Permutation-based procedures represent an alternative approach to significance testing that is especially useful in multidimensional neuroimaging data (Nichols and Holmes, 2002). The total number of 10,000 iterations was performed to obtain a reliable estimate of the null distribution for computed t-statistics. For each iteration, the subject-level community allegiance vectors, M, reflecting modular network structure, were randomly shuffled to represent a random topology unrelated to the observed functional connections while preserving the number and size of the original communities. These vectors were then subjected to the same procedure employed for the actual dataset consisting of agreement calculation, block-averaging, and statistical testing. Each iteration outputted two null t-statistic matrices of size $N_{\mathrm{LSN}} \times N_{\mathrm{LSN}}$ for both effects of interest. A set of 10,000 of these matrices formed a null distribution of the t-statistic. P-values were then calculated based on the null distribution percentile corresponding to the true t-statistic. An effect was considered significant if the true t-statistic was lower than 2.5th or higher than 97.5th percentile of the null distribution. A two-sided approach was used to account for both increases and decreases in a community-level agreement between conditions. Permutation-based p-values were then corrected for multiple comparisons using the FDR method (Benjamini and Hochberg, 1995).

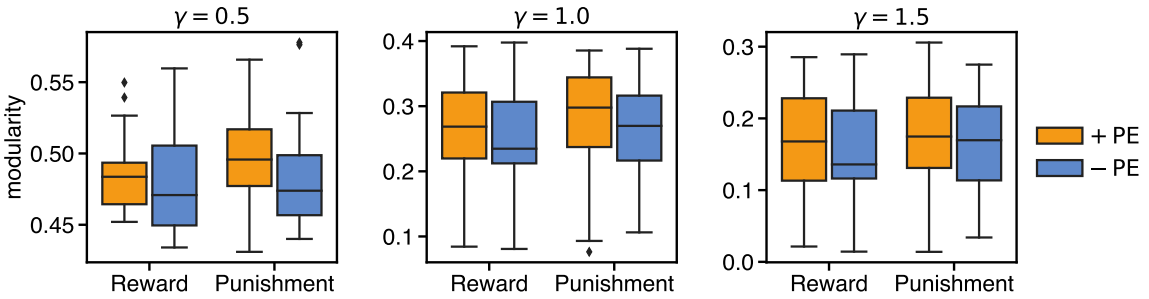

Fig. 4.10 Whole-brain network modularity. Network modularity values for different topological scales and task-conditions. Each plot corresponds to different topological scale characterized by structural resolution parameter γ. Modularity values are grouped according to prediction error sign (orange bars - positive prediction errors; blue bars - negative prediction errors) and task conditions (Reward - reward-seeking condition; Punishment - punishment-avoiding condition). The boxes show the quartiles of modularity distribution while the error bars show lower and upper limits.

4.6.7 Differences in whole-brain modularity

On the level of whole-brain network topology, I was interested in whether the overall degree of modularity changes when subjects switch between processing (1) positive and negative prediction errors in (2) risk-seeking and punishment-avoiding environments. I performed two-way repeated measures ANOVA (rmANOVA) on modularity values with two factors: prediction error sign and task condition. Same statistical testing was applied to all three topological scales.

Average network modularity was highest for the largest topological scale ($Q(\gamma = 0.5) = 0.48 \pm 0.03$) and smallest for the finest topological scale ($Q(\gamma = 0.5) = 0.16 \pm 0.07$). It decreased significantly with increasing structural resolution parameter ($p < 10^{-10}$; paired t-tests between scales). However, the rmANOVA results were not significant. All three effects: main prediction error sign effect, main task condition effect, and interaction (prediction error sign $\times$ task condition) effect, did not reach the significance threshold regardless of the topological scale (**Fig. 4.10**). This shows that the degree of whole-brain network segregation expressed as modularity is stable across task conditions and topological scales. Since the same degree of segregation can characterize a large diversity of network architectures, it is important to analyze networks on finer organizational scales of communities and individual nodes.

4.6.8 Consensus network organization

I was interested in whether large-scale network organization differs between reward-seeking and punishment-avoiding conditions and positive and negative prediction errors. I hypothesized that (1) switching between positive and negative prediction errors would lead to dynamic reorganization of the brain network and the formation of prediction-error-specific subnetworks, (2) regions signaling prediction errors would form separate subnetworks that should be detectable for finer topological scales. To test this hypothesis, I calculated consensus network partitions for each experimental condition and topological scale. I also calculated condition-independent consensus partitions representing stable community structure during prediction error processing. Condition-independent partitions were further used as reference partitions to test condition-specific changes in large-scale networks agreement.

As expected, the number of communities increased, and the average community size for condition-specific consensus partitions decreased with increasing structural resolution parameter γ. For $\gamma = 0.5$, consensus partitions consisted of few "super-communities" (average 2.25 communities across conditions). The mean "super-community" size was 119.1 ± 43.1 nodes (ROIs). For intermediate topological scale, $\gamma = 1$, consensus partitions for all task conditions consisted of three communities with the average size of 89.3 ± 13.2 nodes per community. Consensus partitions detected for the finest topological scale, characterized by the highest structural resolution $\gamma = 1.5$, consisted of 41.2 communities per partition (average across conditions). This proliferation of observed communities resulted from many singletons, i.e., communities with one node. Each partition consisted of 27-37 singletons, which decreased the average community size to 6.5 ± 16.1 nodes per community. A similar number of communities and average community size characterized condition-independent consensus partitions. For the "super-community" scale, the condition-independent consensus partition consisted of two communities with 141 and 127 nodes per community. The intermediate scale partition was composed of three communities of size 100, 94, and 74 nodes per community. In the finest topological scale, the consensus partition divided the network into 47 communities (38 singletons) with an average community size of 5.7 ± 15.2 nodes per community.

In the coarsest topological scale, the functional network consisted of two major "super-communities": task-visual and sensory (**Table 4.4** and **Fig. A.3**, top panel). The task-visual community was composed of task-related networks (PE signaling, fronto-parietal, memory, and salience networks), default mode network, and visual network. The sensory community consisted of cerebellar, somatomotor, cingulo-opercular, dorsal

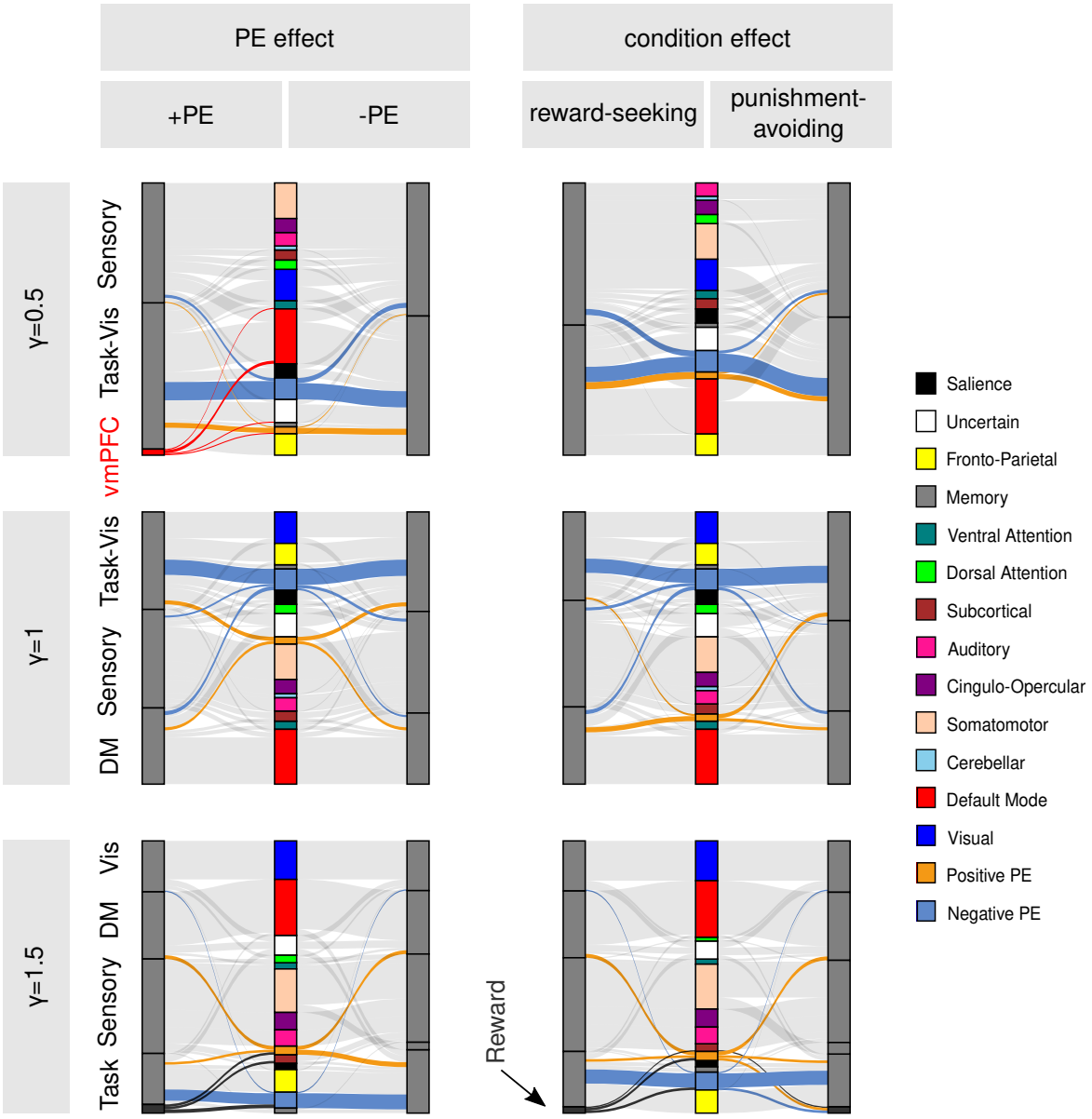

Fig. 4.11 Consensus partitions for different topological scales. Sankey diagrams show composition of consensus partitions with reference to well-known LSNs. First column corresponds to network organization dependent on PE sign, second column shows partition specific to reward-seeking and punishment-avoiding task conditions. Rows correspond to different topological scales characterized by structural resolution parameter γ. Consensus communities: Task-Vis - task-visual; vmPFC - ventromedial prefrontal cortex; DM - default mode; Vis - visual.

Table 4.4 Consensus partition composition for structural resolution parameter $\gamma = 0.5$. Cell values correspond to a percentage of reference network regions belonging to the consensus module. Cell color intensity reflects the underlying percentage value. Reference partition consists of 13 well-known LSN from Power atlas (Power et al., 2011) and two prediction-error-signaling networks (see section 4.6.1). Third column shows consensus module size. Task condition abbreviations: +PE - positive prediction error trials, −PE - negative prediction error trials, RS - risk-seeking condition, PA - punishment-avoiding condition. Reference LSNs abbreviations: DM - default mode, FP - fronto-parietal, MEM - memory, SAL - salience, UNC - uncertain, SUB - subcortical, VAT - ventral attention, CER - cerrebelar, SOM - somatomotor, CO - cingulo-opercular, AUD - auditory, DAT - dorsal attention, VIS - visual.

Module	Condition	n	Large-scale networks														
			+PE	DM	−PE	FP	MEM	SAL	UNC	SUB	VAT	CER	SOM	CO	AUD	DAT	VIS
Task-Visual	+PE	144	0,71	0,70	0,86	1,00	0,75	0,71	0,83	0,40	0,25	0,25	0,00	0,00	0,00	0,33	0,65
	−PE	137	0,86	0,76	0,76	0,95	0,75	0,64	0,78	0,20	0,50	0,25	0,00	0,00	0,00	0,33	0,45
	RS	128	1,00	0,83	0,71	0,95	0,75	0,71	0,74	0,50	0,50	0,00	0,00	0,00	0,00	0,00	0,06
	PA	136	0,71	0,59	0,86	1,00	0,75	0,64	0,70	0,40	0,13	0,25	0,00	0,00	0,00	0,33	0,74
	all	141	0,86	0,70	0,86	1,00	0,75	0,71	0,83	0,30	0,25	0,25	0,00	0,00	0,00	0,33	0,55
Sensory	+PE	118	0,14	0,24	0,14	0,00	0,25	0,29	0,13	0,60	0,63	0,75	1,00	1,00	1,00	0,67	0,35
	−PE	131	0,14	0,24	0,24	0,05	0,25	0,36	0,22	0,80	0,50	0,75	1,00	1,00	1,00	0,67	0,55
	RS	140	0,00	0,17	0,29	0,05	0,25	0,29	0,26	0,50	0,50	1,00	1,00	1,00	1,00	1,00	0,94
	PA	132	0,29	0,41	0,14	0,00	0,25	0,36	0,30	0,60	0,88	0,75	1,00	1,00	1,00	0,67	0,26
	all	127	0,14	0,30	0,14	0,00	0,25	0,29	0,17	0,70	0,75	0,75	1,00	1,00	1,00	0,67	0,45
vmPFC	+PE	6	0,14	0,06	0,00	0,00	0,00	0,00	0,04	0,00	0,13	0,00	0,00	0,00	0,00	0,00	0,00

attention, and auditory networks. Nodes from subcortical and ventral attention networks were divided evenly between two "super-communities." Intriguingly, in the reward-seeking condition, the visual network was detached from the task-related regions and contributed to the sensory community. Moreover, the positive prediction error partition contained a third smaller subnetwork – ventromedial prefrontal cortex community – unobserved for any other condition (**Fig. 4.12**, top panel). Four nodes of this community were located within the ventromedial prefrontal cortex – an area widely known for representing value information to drive choice (Hare et al., 2009; Rangel et al., 2008). The vmPFC community consisted of six regions: three from the default mode network, one from the +PE network, one from the uncertain network, and one from the ventral attention network.

In the intermediate topological scale, a pattern of large-scale communities was similar to the one observed in the "super-community" scale (**Table 4.5** and **Fig. A.3**, middle panel). One notable difference in network organization between these scales

Table 4.5 Consensus partition composition for structural resolution parameter $\gamma = 1$. For abbreviations and additional description see **Table 4.4**.

Module	Condition	n	Large-scale networks														
			+PE	DM	−PE	FP	MEM	SAL	UNC	SUB	VAT	CER	SOM	CO	AUD	DAT	VIS
Task-Visual	+PE	96	0,57	0,06	0,71	0,86	0,75	0,64	0,52	0,10	0,00	0,00	0,03	0,00	0,00	0,44	0,84
	−PE	98	0,57	0,06	0,76	0,90	0,75	0,71	0,57	0,10	0,00	0,25	0,00	0,00	0,00	0,33	0,81
	RS	87	0,29	0,02	0,67	0,86	0,75	0,57	0,48	0,10	0,00	0,00	0,03	0,00	0,00	0,33	0,81
	PA	107	0,57	0,07	0,81	1,00	0,75	0,71	0,57	0,10	0,00	0,25	0,03	0,00	0,00	0,33	0,94
	all	94	0,57	0,04	0,71	0,86	0,75	0,71	0,48	0,10	0,00	0,00	0,03	0,00	0,00	0,33	0,84
Default Mode	+PE	75	0,43	0,85	0,19	0,14	0,00	0,14	0,43	0,30	0,50	0,00	0,00	0,00	0,00	0,00	0,00
	−PE	70	0,43	0,85	0,10	0,10	0,00	0,00	0,43	0,30	0,50	0,00	0,00	0,00	0,00	0,00	0,00
	RS	76	0,71	0,87	0,19	0,14	0,00	0,14	0,43	0,10	0,50	0,00	0,00	0,00	0,00	0,00	0,00
	PA	72	0,43	0,91	0,14	0,00	0,00	0,07	0,39	0,30	0,50	0,00	0,00	0,00	0,00	0,00	0,00
	all	74	0,43	0,87	0,14	0,14	0,00	0,07	0,43	0,30	0,50	0,00	0,00	0,00	0,00	0,00	0,00
Sensory	+PE	97	0,00	0,09	0,10	0,00	0,25	0,21	0,04	0,60	0,50	1,00	0,97	1,00	1,00	0,56	0,16
	−PE	100	0,00	0,09	0,14	0,00	0,25	0,29	0,00	0,60	0,50	0,75	1,00	1,00	1,00	0,67	0,19
	RS	105	0,00	0,11	0,14	0,00	0,25	0,29	0,09	0,80	0,50	1,00	0,97	1,00	1,00	0,67	0,19
	PA	89	0,00	0,02	0,05	0,00	0,25	0,21	0,04	0,60	0,50	0,75	0,97	1,00	1,00	0,67	0,06
	all	100	0,00	0,09	0,14	0,00	0,25	0,21	0,09	0,60	0,50	1,00	0,97	1,00	1,00	0,67	0,16

was a separation of the default mode network and part of the +PE signaling network from the task-visual community. The default mode community consisted of most default mode network nodes (85-91%), around half of the +PE signaling, ventral-attention, and uncertain networks nodes, around 10-30% subcortical and 10-20% of −PE signaling and fronto-parietal network nodes. Interestingly, most −PE signaling network nodes remained coupled with the task-visual community. In contrast to the coarsest topological scale, the visual network was consistently coupled with task-related networks for all task conditions.

The analysis of the network partitions for the highest structural resolution, $\gamma = 1.5$, showed the fine-grained division of the functional network into many smaller communities. Regardless of experimental condition, the four largest communities – visual, default mode, sensory, and task – formed the backbone of each consensus partition (**Table A.1** and **Fig. A.3**, bottom panel). Similar to the intermediate scale, the default mode community consisted of most default mode network nodes and regions from +PE signaling, ventral attention, and uncertain networks. Task community consisted of regions from fronto-parietal, memory, −PE signaling, +PE signaling and salience network. Interestingly, +PE signaling regions were split between default mode, task, and reward communities, whereas −PE signaling regions resided within the task, reward, and other minor communities. In addition to the backbone communities, a

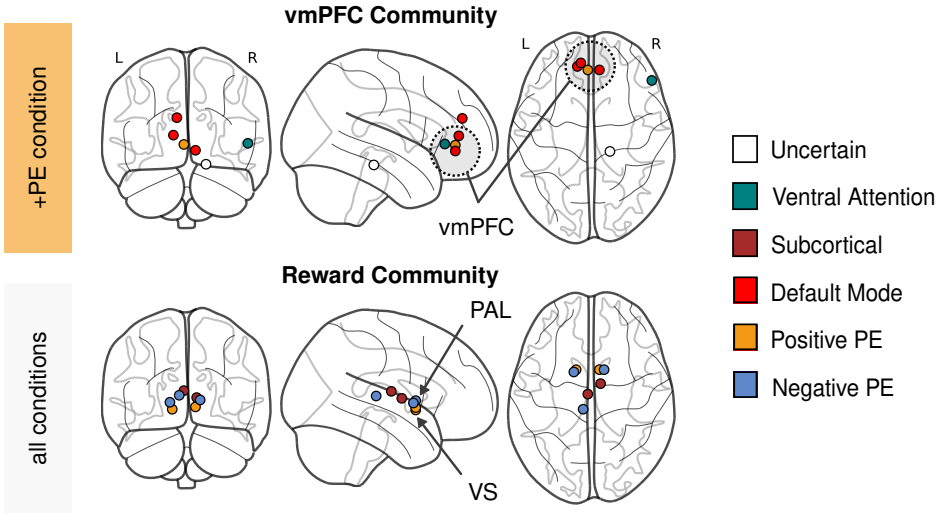

Fig. 4.12 Selected consensus partition communities. (**Top panel**) Ventromedial prefrontal cortex community emerging during processing of positive prediction errors observed for structural resolution $\gamma = 0.5$. (**Bottom panel**) Separate reward community observed for condition-invariant consensus partition for structural resolution $\gamma = 1.5$. It is comprised of PE signaling regions and subcortical network regions and spatially restricted to the areas around striatum. Regional abbreviations: vmPFC - ventromedial prefrontal cortex; VS - ventral striatum; PAL - pallidum.

small cerebellar community was consistently discovered for all experimental conditions. Moreover, a stable reward community was a part of the consensus partition for +PE, risk-seeking, and punishment-avoiding conditions. This community was also detected in condition-independent consensus partition representing general network structure during prediction error processing. The reward community consisted of 5-8 regions from PE signaling and subcortical networks (**Fig. 4.12**, bottom panel). Specifically, the condition-independent reward community consisted of bilateral ventral striatum and pallidum, right striatum, and two regions within the thalamus. All regions constituting the reward community were spatially bounded to subcortical areas located around the striatum. The existence of a separate reward network was recently suggested by Huckins et al. (2019) after examining resting-state connectivity patterns in a large cohort of subjects.

4.6.9 Large-scale networks interactions

I explored condition-specific changes in large-scale networks agreement to test whether modular network architecture fluctuates when subjects switch between processing positive and negative prediction errors. I was also interested if such fluctuations

can be detected when changing between reward-seeking and punishment-avoiding environments. I hypothesized that changes related to the PE sign should be more pronounced than changes related to the task condition regardless of the structural resolution. I also expected that processing negative prediction errors would increase working memory load leading to an increased between-community agreement and decreased within-community agreement, especially for default-mode, task, and reward systems.

Functional networks in the coarsest topological scale consisted of task-visual and sensory "super-communities". Agreement between these "super-communities" was higher during processing −PEs compared with +PEs ($D_{+\mathrm{PE}} = 0.560$, $D_{-\mathrm{PE}} = 0.579$, $t = -0.63$, $p_{\mathrm{FDR}} < 0.0001$) and higher in reward-seeking compared with the punishment-avoiding condition ($D_{\mathrm{RS}} = 0.573$, $D_{\mathrm{PA}} = 0.563$, $t = 0.22$, $p_{\mathrm{FDR}} < 0.0001$) (**Fig. 4.13**). Significant change of within-community agreement was observed only for PE sign dimension. Task-visual community was more segregated during processing +PEs as indicated by increased within-community agreement ($D_{+\mathrm{PE}} = 0.712$, $D_{-\mathrm{PE}} = 0.698$, $t = 0.65$, $p_{\mathrm{FDR}} = 0.006$).

The intermediate topological scale was characterized by three large-scale communities discovered within the consensus partition: default-mode, task-visual and sensory. Three out of six possible changes in the LSN agreement were significant for the PE sign dimension. Agreement within the default mode network increased during processing +PEs compared with −PEs ($D_{+\mathrm{PE}} = 0.576$, $D_{-\mathrm{PE}} = 0.522$, $t = 3.34$, $p_{\mathrm{FDR}} < 0.0001$). Additionally, the sensory community displayed decreased agreement during +PEs processing with both default-mode ($D_{+\mathrm{PE}} = 0.183$, $D_{-\mathrm{PE}} = 0.217$, $t = -2.66$, $p_{\mathrm{FDR}} < 0.0001$) and task-visual communities ($D_{+\mathrm{PE}} = 0.214$, $D_{-\mathrm{PE}} = 0.248$, $t = -3.44$, $p_{\mathrm{FDR}} < 0.0001$). Note that decrease in integration between task-visual and sensory networks was significant for both "super-community" and intermediate scales, whereas increased segregation of task-visual community was confined only to the coarsest topological scale. In contrast to the PE sign effect, only one task-condition change remained significant after correction for multiple comparisons. Agreement within the task-visual community was higher in punishment-avoiding compared with the reward-seeking condition ($D_{\mathrm{RS}} = 0.514$, $D_{\mathrm{PA}} = 0.555$, $t = -2.19$, $p_{\mathrm{FDR}} < 0.0001$). Interestingly, similar effect was not observed for the topological scale characterized by $\gamma = 0.5$. Increased agreement between default-mode and sensory communities during punishment-avoiding condition was initially significant but did not survive multiple comparison correction ($D_{\mathrm{RS}} = 0.192$, $D_{\mathrm{PA}} = 0.208$, $t = -1.11$, $p_{\mathrm{UNC}} = 0.04$).

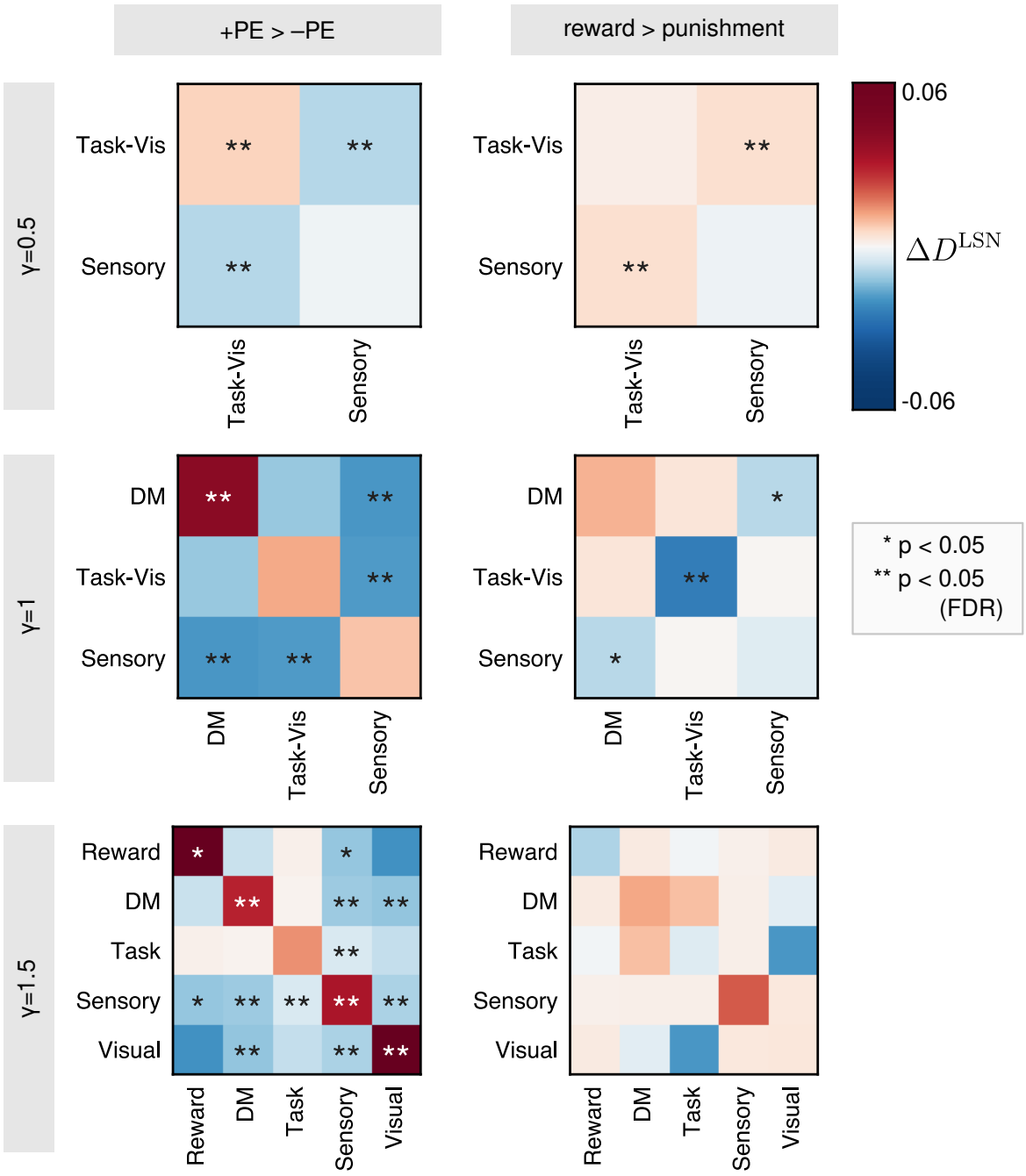

Fig. 4.13 Agreement between large-scale networks. Changes in within and between-community agreement for prediction error sign and task condition. Within-community agreement captures the extent of community segregation whereas between-community agreement quantifies integration between communities. Left column shows difference between agreement during +PE and −PE processing. Right column corresponds to the changes between reward-seeking and punishment-avoiding condition. Rows correspond to topological scales. Reference communities come from condition-independent consensus partitions. LSNs abbreviations: Task-Vis - task-visual; DM - default mode.

The consensus partition for the fine-grained topological scale characterized by $\gamma = 1.5$ consisted of nine non-singleton communities. However, only five of these non-singletons were comprised of more than three network nodes. Since minuscule communities are likely byproducts of a reclustering algorithm without significant biological meaning, they were excluded from further LSN agreement analysis. The remaining communities were labeled as reward, default-mode, task, sensory and visual. None of the 15 possible LSN agreement changes between reward-seeking and punishment-avoiding conditions were significant. Contrary, more than half of PE-sign-related changes turned out to be significant:

- Reward community increased its segregation and decreased its integration with the sensory community with increasing PE. However, any of these changes did not survive multiple comparison correction.

- Default mode community increased its segregation ($D_{+\mathrm{PE}} = 0.549$, $D_{-\mathrm{PE}} = 0.503$, $t = 2.04$, $p_{\mathrm{FDR}} < 0.05$) and decreased its integration with sensory ($D_{+\mathrm{PE}} = 0.061$, $D_{-\mathrm{PE}} = 0.082$, $t = -3.59$, $p_{\mathrm{FDR}} < 0.0001$) and visual communities ($D_{+\mathrm{PE}} = 0.037$, $D_{-\mathrm{PE}} = 0.061$, $t = -3.31$, $p_{\mathrm{FDR}} < 0.0001$) when switching from negative to positive PEs processing.

- Task community had a lower agreement with sensory community during +PE processing ($D_{+\mathrm{PE}} = 0.047$, $D_{-\mathrm{PE}} = 0.056$, $t = -3.54$, $p_{\mathrm{FDR}} < 0.0001$).

- Sensory community increased within-community agreement with increasing PE ($D_{+\mathrm{PE}} = 0.483$, $D_{-\mathrm{PE}} = 0.434$, $t = 2.31$, $p_{\mathrm{FDR}} < 0.0001$) and decreased its integration with all remaining communities.

- Visual community increased segregation during +PE processing ($D_{+\mathrm{PE}} = 0.530$, $D_{-\mathrm{PE}} = 0.441$, $t = 3.41$, $p_{\mathrm{FDR}} < 0.001$) and decreased integration with default-mode and sensory communities.

Almost all changes related to PE sign switching followed a hypothesized pattern of increased within-community agreement and decreased between-community agreement. The only exceptions to that observation were increased agreement during +PE processing between task community and reward/default mode communities for $\gamma = 1.5$ and decreased task-visual community agreement during +PE processing for $\gamma = 0.5$. However, none of these changes reached the significance level, while all significant effects followed the hypothesized pattern. Interestingly, the community separation

effect reflected by increased segregation and decreased integration with other communities was more pronounced for sensory, default-mode, and visual communities and less pronounced for task and reward networks.

PE sign effects were more evident than task condition effects for all topological scales, as suggested by more significant changes observed for PE sign dimension regardless of the topological scale. This observation was most apparent for the fine-grained topological scale, where none of the task condition effects reached significance in contrast to nine significant effects for the PE sign dimension. Moreover, almost all changes related to switching between positive and negative prediction errors were scale-invariant and spanned two or three scales. For example, a decreased agreement between sensory and task/task-visual communities during +PE processing was significant for all three structural resolutions. In contrast, two significant task-condition-related changes were specific to a single structural resolution. Concretely, decreased task-visual community agreement during reward-seeking condition was specific to intermediate scale, and increased agreement between sensory and task-visual communities was significant only for $\gamma = 0.5$.

4.7 Discussion

The overarching goal of my study was to provide a complete description of neural correlates of prediction errors by taking into account three different perspectives on behavioral and neural data. These perspectives included describing behavior, brain activity, and brain connectivity during probabilistic reversal learning.

On the behavioral level, I found that learning speed depends only on the sign of the prediction error and not on the outcome valence. In line with the dual system hypothesis, activation analysis revealed two independent sets of brain regions signaling positive-going and negative-going prediction errors. Contrary to the strict formulation of the reference effect hypothesis, I found significant valence-related changes in PE processing located outside the reward system, mainly in visual and parietal cortices. Whole-brain network analysis did not show significant differences in network modularity. Functional networks revealed a multi-scale community structure with a separate reward network emerging at a finer topological scale. The agreement between detected large-scale networks changed between positive and negative prediction error processing. These changes were characterized by decreased within-network segregation and increased between-network integration during negative prediction error processing. This pattern

of changes, observed mainly for default mode, sensory and visual networks, aligned with the Global Workspace hypothesis.

4.7.1 Opponent system for negative prediction errors processing

From the computational perspective, negative prediction errors are simply negative scalar values located on the same scale as positive prediction errors. This simple observation raises the question - How does the brain represent negative prediction errors? One possible answer to this question takes the form of the dual systems hypothesis stating that negative PEs are signaled by a separate neural system outside of the dopaminergic circuit. I used three independent data analysis perspectives to investigates assumptions of the dual systems hypothesis. Using model-based activation analysis, I observed a clear distinction between dopaminergic, striatal system signaling positive prediction errors and insular-frontal system signaling negative prediction errors. Network perspective revealed differential community membership profiles of both systems with positive PE system linked with the DMN and negative PE system coupled with the FPN. Behavioral data analysis also supported the duality of PE signaling – the learning rate depended on the sign of prediction error.

An electrophysiological, lesion, and pharmacological studies in animals suggested that the negative part of the prediction error signal is encoded in a set of brain regions outside of the dopaminergic system, most notably in the insula and the amygdala (Hayes et al., 2014; Namburi et al., 2016). Recent fMRI meta-analysis supported these observations by showing a widespread network of areas signaling negative prediction errors encompassing dorsomedial cingulate cortex, anterior insula, palladium, middle frontal gyrus (Fouragnan et al., 2018). In line with these observations, I found significant activity correlated with both increasing and decreasing prediction errors. I reported negative PE signaling in the dorsomedial cingulate cortex, bilateral anterior insula, pars opercularis, and a few other areas. The composition of the negative PE signaling network was consistent with meta-analytic findings.

A separate community of reward-related regions was recently recognized as a stable large-scale network at rest (Huckins et al., 2019). This finding suggests that prediction error processing regions should be functionally coupled. Based on that, I expected that the distinction between regions signaling positive and negative PEs would be reflected by the separate community membership of these networks. I analyzed community structure in three different topological scales. On the coarsest scale, I found that both

prediction error systems were a part of a larger task-visual module. Both intermediate and fine-grained topological scales revealed distinct community membership profiles of positive and negative prediction error processing regions. On the intermediate scale, positive PE regions split evenly between task-visual and default mode modules, whereas negative PE regions retained their association with a task-visual module. The fine-grained scale revealed even more complex structural patterns. I found a separate reward module composed mainly of striatal regions belonging to both prediction error networks. The rest of the prediction error signaling regions were divided between default mode and task modules. The default mode module contained almost half of the positive PE regions and no negative PE regions. On the other hand, the task module consisted of both positive and negative PE regions. These findings partially corroborate the dual system hypothesis by showing strong associations between the dopaminergic system and DMN and the insular system and FPN. Interestingly, these associations indicate that the opposition between positive and negative PE systems may be related to the antagonism between task-negative DMN and task-positive FPN. My results show a much more complicated picture of the network community structure in the context of reward-related regions. Most notably, the existence of a separate striatal network comprising both positive and negative PE processing regions suggests a need to integrate information between the opponent systems. I have to note that contrary to the initial hypothesis, I did not find pure modules composed of positive or negative PE processing regions. There are two possible explanations – either some of these regions are irreducibly functionally connected, or the distinction between them exists, but the finer topological scale is needed to uncover them.

Many studies showed that temporal-difference learning models with asymmetric value updates for positive and negative prediction errors are better at explaining learning in humans (Frank et al., 2007; Gershman, 2015, 2016). These models were introduced to take into account the observations supporting the dual systems hypothesis (Frank et al., 2007). If separate neural circuits signal both types of prediction errors, their influence on choice value should be independent. On the behavioral level, this influence is operationalized as a scalar value – the learning rate. I hypothesized that models with separate learning rates for positive and negative PEs should outperform single learning rate models. In line with my prediction and existing literature, I found moderate evidence in favor of this hypothesis. Moreover, the single most frequent model was the model with PE-dependent learning rates. This finding further suggests the separation between positive and negative PE systems on the behavioral level.

4.7.2 Brain systems are organized along prediction error sign axis

We can feel positive emotions in generally negative situations and negative emotions when something good happens. For example, when our car breaks down but repair cost is low, we feel relief, or when our boss praises us, but our colleagues get a raise, we feel envy and frustration. These feelings can be interpreted in the light of the reference effect hypothesis. The reference effect hypothesis states that values are not absolute but are actively constructed based on our previous experience. It offers the solution to the ongoing debate on the primary organizational axis of the brain systems. If values were constructed in absolute terms, brain responses would be organized along the outcome valence axis. However, suppose values are relative to the experienced context. In that case, brain responses should only reflect the prediction error sign axis because prediction mirrors the outcome modulated by expectations acting as a reference. To verify the reference effect hypothesis, I investigated whether brain systems are invariant to the outcome valence and prediction error sign. On the behavioral level, I found that learning rates are invariant to the outcome valence but depend on the prediction error sign. On the activation and connectivity level, I found that the prediction error sign effect sizes are much stronger than the outcome valence effect sizes.

As mentioned in the previous section, temporal-difference learning models incorporating asymmetric learning rates for positive and negative prediction errors, tend to outperform models with symmetric learning rates (Frank et al., 2007; Gershman, 2015, 2016). Although learning rate asymmetry was previously studied, it only reflected the prediction error sign axis and not the outcome valence axis. The question of whether the learning rate shifts between reward-seeking and punishment-avoiding environments remained open. To answer this question, I built a hierarchical latent mixture model with four competing submodels covering an entire space of possibilities regarding outcome valence and prediction error sign asymmetry. Based on the reference effect hypothesis, I expected that the model with separate learning rates for positive and negative prediction errors but constant learning rates for reward-seeking and punishment-avoiding conditions (PDCI model), would outperform other models. In line with this assumption, I found that the PDCI model was the most likely model across participants. This finding suggests that the opponent brain systems reflect positive and negative prediction errors rather than absolute rewards and punishments. Moreover, the learning system's influence on the valuation system, measured as the learning rate, is similar regardless of subjects seeking rewards or avoiding punishments.

The strict version of the reference effect hypothesis states that prediction error processing should be perfectly invariant to the outcome valence. From the activation analysis perspective, this would imply that there should be no difference in BOLD response to PEs between reward-seeking and punishment-avoidance conditions. Contrary to this assumption, I found that visual areas V1, V3, and V4, right supramarginal gyrus, right superior parietal lobule, right precuneus, and right precentral gyrus were more sensitive to increasing prediction errors in reward-seeking compared with the punishment-avoiding condition. The same effect was reported by Meder et al. (2016) in V3, V4, inferior frontal gyrus, supplementary motor area, posterior cingulate, left dorsomedial prefrontal cortex, and right thalamus. The authors suggested that the effect is "congruent with the two-dimension hypothesis," which states that both prediction error sign and outcome valence axes influence learning signals in the brain. Intriguingly, apart from visual areas, regions sensitive to outcome-valence reported by Meder et al. (2016) do not overlap with four non-visual areas found in my study. This divergence may be a result of differences in experimental design, i.e., Meder et al. (2016) used abstract stimuli and no monetary incentives, whereas in my study, participants played for points exchanged for real money. The difference in BOLD response to PE in visual areas can be explained by the slight difference in stimulus presentation during the outcome phase in both versions of the task used by Meder et al. (2016) and in my study. Positive prediction errors are related to the same color of fixation circle background and border in the reward-seeking condition. On the other hand, in the punishment-avoiding condition, positive prediction errors are accompanied by differential colors within the fixation circle. This systematic difference in fixation circle contrast between experimental conditions is likely to influence the activation difference in response to prediction error in visual areas.

Although my activation results do not directly support the strict formulation of the reference effect hypothesis, they can corroborate a more relaxed version of this hypothesis. When I used the same statistical threshold of $p < 0.0001$ for both context-dependent and context-independent effects, the number of significant clusters for the context-dependent effect dropped from seven to two, compared with twenty-eight clusters for the context-independent effect. Two remaining clusters were located within the right V3/V4 and the right supramarginal gyrus. This finding directly suggests that the prediction error sign axis is related to more pronounced changes in the brain activation than the outcome valence axis, supporting the general outcome valence invariance of the reward system.

The organization of large-scale brain systems constantly changes to meet the demands of the environment (Shine et al., 2016). According to the reference effect hypothesis, the functioning of these systems should be invariant to the outcome valence. From the connectivity perspective, this should imply that changes in modular network structure should be predominately associated with switching between positive and negative prediction errors and not between reward-seeking and punishment-avoiding conditions. In agreement with this assumption, I reported more significant tests for the change in prediction error sign than the change in outcome valence regardless of the topological scale. Specifically, I found seven significant PE sign effects for the finer topological scale compared with no significant task condition effects. Moreover, switching between different prediction error signs was associated with a topologically stable pattern of large-scale network reorganization. In contrast, outcome valence effects were less consistent and limited to the two coarsest topological scales. Similar to the observed brain activations, these findings support a relaxed version of the reference effect hypothesis – the prediction error axis is a dominant organizational axis for brain systems except for a few subtle valence effects.

4.7.3 Negative prediction errors elicits stable pattern of network reconfiguration

The higher cognitive effort associated with increased demands of the performed task is thought to require long-distance connections integrating separate neural systems (Dehaene et al., 1998). On the other hand, the brain can process lower cognitive demands within a set of functionally specialized modules. From the network science perspective, these two effects correspond to decreased modularity, increased within-community integration, and decreased between-community segregation. Several studies reported these changes in functional brain networks during cognition (Braun et al., 2015; Finc et al., 2017; Shine et al., 2016; Vatansever et al., 2015). In a reversal learning task, negative prediction errors signal either expected random fluctuations or unexpected changes in reward contingencies. Therefore, negative PE processing may be associated with increased attention and cognitive effort. Similar to previous studies, I expected to observe decreased whole-brain modularity, within-community segregation and increased between-community integration when switching from positive to negative PE processing. In line with these expectations, I observed decreased segregation of default mode, sensory and visual modules, and increased integration between default

mode and sensory/visual modules during negative PE processing. However, overall network modularity was stable across prediction error signs and task conditions.

Although modularity values were lower for negative PEs in all three topological scales, none of these differences reached a significance level. There are two possible explanations for that observation. First, in the case of probabilistic learning, cognitive effort differences between positive and negative PEs could be much lower than similar differences during working memory or attention tasks. Subtle differences in cognitive effort would lead to a smaller modularity breakdown that cannot be detected with the sample size used in my study. This explanation is consistent with the observation of lower modularity for negative PEs in all topological scales and moderately small p-value ($p = 0.18$) for PE effect in the "super-community" topological scale. Second, the reorganization effect may be specific to changes in the composition of network communities without altering the macroscopic level of global modularity (Sporns, 2014). Similar to my results, a recent study on number comparison reported significant alterations of network community structure without changes in whole-brain network modularity (Conrad et al., 2020).

In the coarsest topological scale, the default mode network was a part of the task-visual module. This module decreased its segregation and increased integration with the sensory module during negative PE processing. In intermediate and finer topological scales, DMN formed a separate module. Similar to the task-visual module, this module decreased segregation and increased integration with sensory and visual modules. Recent work demonstrated that "DMN may play an essential role in the formation of an integrated workspace" (Finc et al., 2017). In the context of reinforcement learning, DMN contains a part of the prediction error signaling network, most notably the ventromedial prefrontal cortex commonly associated with value representation (Frank and Claus, 2006; Rangel et al., 2008). My findings suggest that DMN reorganization patterns may reflect (1) Global Workspace formation and (2) enhanced communication between the valuation circuit and other brain systems following negative PE processing.

The sensory network contains the motor cortex responsible for the movement execution following the subject's decision. Some studies suggest that motor regions may implement a winner-take-all mechanism integrating values of different stimuli (Cisek and Kalaska, 2005). Moreover, Horga et al. (2015) demonstrated the importance of sensorimotor connectivity for reinforcement learning during gradual learning in a virtual maze. Interestingly, my results show that the sensory network increases its agreement during negative PE processing with all other systems regardless of the scale.

This suggests a potential need for information integration between the choice and valuation circuit and other large-scale networks.

I observed the separate reward network only for finer topological scale characterized by resolution parameter $\gamma = 1.5$. The reward network consisted of a few striatal regions signaling both types of prediction errors. Consistently with the reconfiguration pattern of the Global Workspace formation, the reward network decreased its segregation and increased its integration with the sensory network during negative PE processing. Decreased segregation of striatal regions may reflect functional decoupling of positive and negative PE processing regions during a more effortful type of prediction error processing. This decoupling possibly allows reducing interference between antagonistic areas when more cognitive resources are required. I have to note that reward system effects did not survive multiple comparison corrections and need to be interpreted with caution. A smaller effect size for the reward network may be related to a smaller community size of only seven nodes.

Agreement analysis revealed only two changes in large-scale network integration and segregation for the outcome valence axis. Both changes were related to the task-visual module and were observed for the two coarser topological scales ($\gamma = 0.5$ and $\gamma = 1$). This module increased its integration with the sensory module and decreased its segregation during the reward-seeking condition. One possible explanation of this effect is the influence of systematic differences in stimulus presentation during the outcome phase described in the previous section. In the activation analysis, I observed differences in PE processing in a large portion of the primary and secondary visual cortex, a part of the task-visual module. It is also important to note that despite statistical significance, both effects were scale-specific and were not present for other topological resolutions, unlike PE-related changes.

4.7.4 Ventromedial prefrontal regions form separate network during positive prediction error processing

The ventromedial prefrontal cortex is one of the main functional hubs of the default mode network (Andrews-Hanna et al., 2014). From the reinforcement learning perspective, vmPFC signals positive prediction errors (Daw et al., 2011; Van den Bos et al., 2012) and represents a subjective value of various stimuli (Levy and Glimcher, 2012; Roy et al., 2012). Using consensus partitioning, I found that regions spanning the vmPFC area formed a separate network community during positive prediction error processing.

Surprisingly, this community was present only for the coarsest topological scale ($\gamma = 1.5$), which inherently favors larger "super-communities." This may suggest remarkably strong connections among vmPFC nodes during positive PE processing. It has been previously suggested that value representation in vmPFC is updated by striatal prediction errors through strong frontostriatal connections (Frank and Claus, 2006). Connectivity studies supported this claim by showing increased functional coupling between ventral striatum and vmPFC during feedback processing (Camara et al., 2009; Münte et al., 2008). Moreover, Van den Bos et al. (2012) showed that connection strength between these two regions is enhanced during positive feedback processing. This value update mechanism offers a possible explanation of my finding – the dopaminergic system may communicate positive prediction errors with vmPFC through frontostriatal connections evoking strong and coherent neural activity in different parts of vmPFC. Strong coherent activity results in strong functional coupling, which results in a separate vmPFC community observed during positive PE processing. The other complementary mechanism for community formation is decreased connectivity with non-community members. From the PE processing perspective, this would reflect the need for the valuation circuit in vmPFC to reduce noise from other brain systems during value updating. Only one or both of the suggested mechanisms may be responsible for the vmPFC community formation. Therefore more studies and fine-grained network structure analysis would be needed to understand this phenomenon better. Since this result was unexpected, it should be treated as exploratory and interpreted with caution.

4.8 Conclusions

Multiple neural systems are engaged to signal prediction errors enabling us to learn through trial-and-error. The results of my study show that these systems are organized along the prediction error sign axis, reflecting the distinction between positive and negative prediction errors. Moreover, the activity and connectivity of these systems are generally invariant to the outcome valence, supporting the idea that the brain adjusts its reference point when the environment is predominantly rewarding or punishing. My results also demonstrate a complex pattern of network interactions following switching between positive and negative prediction errors. These interactions form the pattern observed in studies on cognitive load, suggesting that negative prediction errors require more cognitive resources as they are usually related to the conflict. Intriguingly, I also found some unexpected network community structures. Ventromedial prefrontal

regions formed a small community observed along two large "super-communities," suggesting increased integrity of the valuation system and decreased communication with the rest of the brain during positive prediction error processing. Striatal areas composed of both regions signaling positive and negative prediction errors formed a stable reward network observed for higher topological resolutions supporting the finding of a separate reward network observed at rest.

My findings provide a thorough description of neural correlates of prediction errors from three different perspectives. They support and extend some existing observations and shed new light on the network mechanism behind reinforcement learning. Some of my results raise new fascinating questions and open avenues for future research.

4.9 Limitations

I would like to point out that this thesis has its limitations. I created prediction error signaling regions of interest based on meta-analytic findings, which is less precise than the individual delineation of these regions using anatomical scans. Moreover, I neglected the valuable information about the magnitude of prediction errors by using the beta-series correlation approach. This approach simplified the analysis but could potentially eliminate subtler effects that would be detectable otherwise. Additionally, I used only one brain parcellation to conduct the connectivity analysis. In the future, it would be beneficial to replicate my results using other brain parcellations and ideally other independent datasets. Finally, some hypotheses were not strictly tested quantitively using statistical calculations but were qualitatively verbalized as observed patterns of statistically significant results. This is a difficulty inherent to the approach taken in this thesis, which tries to combine multiple perspectives to answer the same questions.

Summary

Sophisticated methods of modern computational neuroscience and rapid progress in neuroimaging techniques offer unique opportunities to study brain computations during various cognitive processes. These advances provide an unprecedented chance to discover actual algorithms implemented by the neurons and their connections. One of the greatest successes of computational neuroscience were discoveries that led to the formulation of the reward prediction error hypothesis. By many neuroscientists, it is thought to be a deep and elegant explanation of reward-based learning. It links the activity of dopaminergic neurons in the midbrain with computations required to update the value of taken actions. Despite its tremendous success, the reward prediction error hypothesis is relatively simple and, in its original form, fails to address some fundamental questions about learning. One of these questions asks how punishments are processed and how punishment-avoidance learning is biologically implemented. Theorists proposed multiple solutions to this problem, but it is still debatable which one offers the most accurate explanation of available observations.

In this thesis, I had a chance to expand on the reward prediction error hypothesis and fill the gaps related to the brain's implementation of punishment-avoidance learning. The main goal of this thesis was to comprehensively describe all phenomena pertaining to prediction error processing in the brain. To achieve this goal, I took the approach of formulating and testing hypotheses independently using behavioral, activation, and connectivity analysis. This strategy gave me insights into how different regions broadcast prediction errors signals, communicate with other regions and how these interactions manifest in observed behavior. Although finding the commonalities and differences between fundamentally different perspectives on the brain can be challenging, it is required to consolidate our knowledge about the brain.

To fulfill my goal, I designed and conducted an fMRI experiment with a reversal learning task to elicit positive and negative prediction errors independently during reward-seeking and punishment-avoiding conditions. I focused on disentangling the prediction error sign and the outcome valence axes to avoid confusion about whether

certain phenomena are characteristic or invariant to a specific axis. I planned and performed my analyses around three central questions: Is the distinction between signaling better-than-expected and worse-than-expected reflected by two opponent brain systems? Does the brain adjust the reference point to use the same computations in the rewarding and punishing environments? Are negative prediction errors associated with increased integration between different brain systems? Different perspectives allowed me to tackle these questions directly or indirectly and comprehensively describe the neural and behavioral correlates of reward and punishment learning.

I found that learning from outcomes better-than-expected and worse-than-expected is facilitated by two distinct brain systems: dopaminergic corticostriatal and insular-frontal. I observed this distinction directly using activation analysis and indirectly using the behavioral and network approaches. My findings supported the idea that both systems are mainly invariant to the change between purely rewarding and punishing environments. This mechanism is efficient from the brain's perspective because by adjusting the reference point, it can take advantage of two existing circuits regardless of the environmental demands. The reference point reflects the expected number of rewards and punishments in a given environment and allows accurate prediction errors recalculation. Using network analysis, I showed that the interactions between large-scale brain networks fluctuate during prediction error processing. Intriguingly, the pattern of network reconfiguration when switching from positive to negative prediction errors is strikingly similar to the pattern observed when switching from low-demand to high-demand cognitive tasks. This observation suggests that processing worse-than-expected outcomes is cognitively more demanding and evokes some degree of increased integration between brain systems.

The contribution of this thesis to the neuroscience of human choice and learning is twofold. First, it offers firm support for the dual systems hypothesis, demonstrating a clear distinction between positive and negative prediction error signaling. It also shows that the central organizational axis in the learning brain is the prediction error sign axis. Second, it demonstrates that the network approach offers an essential perspective for studying prediction error correlates. It also stimulates to ask novel questions about prediction error processing in the brain, e.g., "What is the connection between prediction error systems and large-scale networks, especially default mode and fronto-parietal?" or "What network interactions facilitate updating values in the ventromedial prefrontal cortex?". Altogether, I believe that my work brings important insights to the computational neuroscience and provides a comprehensive description of neural correlates of prediction errors during reward and punishment learning.

References

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):790–795.

Aghajanian, G. and Liu, R.-J. (2017). Serotonin (5-hydroxytryptamine; 5-ht): Cns pathways and neurophysiology. 2:171–213.

Akaike, H. (1998). Information theory and an extension of the maximum likelihood principle. In *Selected papers of Hirotugu Akaike*, pages 199–213. Springer.

Amara, L., Scala, A., Barthelemy, M., and Stanley, H. E. (2011). Classes of small-world networks. In *The Structure and Dynamics of Networks*, pages 207–210. Princeton University Press.

Andrews-Hanna, J. R. (2012). The brain's default network and its adaptive role in internal mentation. *The Neuroscientist*, 18(3):251–270.

Andrews-Hanna, J. R., Smallwood, J., and Spreng, R. N. (2014). The default network and self-generated thought: component processes, dynamic control, and clinical relevance. *Annals of the New York Academy of Sciences*, 1316(1):29.

Avants, B. B., Epstein, C. L., Grossman, M., and Gee, J. C. (2008). Symmetric diffeomorphic image registration with cross-correlation: evaluating automated labeling of elderly and neurodegenerative brain. *Medical image analysis*, 12(1):26–41.

Azevedo, F. A., Carvalho, L. R., Grinberg, L. T., Farfel, J. M., Ferretti, R. E., Leite, R. E., Filho, W. J., Lent, R., and Herculano-Houzel, S. (2009). Equal numbers of neuronal and nonneuronal cells make the human brain an isometrically scaled-up primate brain. *Journal of Comparative Neurology*, 513(5):532–541.

Baars, B. J. (2002). The conscious access hypothesis: origins and recent evidence. *Trends in cognitive sciences*, 6(1):47–52.

Barabási, A.-L. (2009). Scale-free networks: a decade and beyond. *Science*, 325(5939):412–413.

Barabási, A.-L. and Albert, R. (1999). Emergence of scaling in random networks. *Science*, 286(5439):509–512.

Baram, A. B., Muller, T. H., Nili, H., Garvert, M. M., and Behrens, T. E. J. (2021). Entorhinal and ventromedial prefrontal cortices abstract and generalize the structure of reinforcement learning problems. *Neuron*, 109(4):713–723.

Barnes, J. A. (1969). Graph theory and social networks: A technical comment on connectedness and connectivity. *Sociology*, 3(2):215–232.

Bassett, D. S. and Mattar, M. G. (2017). A network neuroscience of human learning: potential to inform quantitative theories of brain and behavior. *Trends in Cognitive Sciences*, 21(4):250–264.

Bassett, D. S., Porter, M. A., Wymbs, N. F., Grafton, S. T., Carlson, J. M., and Mucha, P. J. (2013). Robust detection of dynamic community structure in networks. *Chaos: An Interdisciplinary Journal of Nonlinear Science*, 23(1):013142.

Bassett, D. S., Wymbs, N. F., Porter, M. A., Mucha, P. J., Carlson, J. M., and Grafton, S. T. (2011). Dynamic reconfiguration of human brain networks during learning. *Proceedings of the National Academy of Sciences*, 108(18):7641–7646.

Bassett, D. S., Yang, M., Wymbs, N. F., and Grafton, S. T. (2015). Learning-induced autonomy of sensorimotor systems. *Nature neuroscience*, 18(5):744–751.

Bavard, S., Lebreton, M., Khamassi, M., Coricelli, G., and Palminteri, S. (2018). Reference-point centering and range-adaptation enhance human reinforcement learning at the cost of irrational preferences. *Nature communications*, 9(1):1–12.

Bayer, H. M. and Glimcher, P. W. (2005). Midbrain dopamine neurons encode a quantitative reward prediction error signal. *Neuron*, 47(1):129–141.

Beissner, F. (2015). Functional mri of the brainstem: common problems and their solutions. *Clinical neuroradiology*, 25(2):251–257.

Benjamini, Y. and Hochberg, Y. (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal statistical society: series B (Methodological)*, 57(1):289–300.

Bertolero, M. A., Yeo, B. T., and D'Esposito, M. (2015). The modular and integrative functional architecture of the human brain. *Proceedings of the National Academy of Sciences*, 112(49):E6798–E6807.

Betzel, R. F. and Bassett, D. S. (2017). Multi-scale brain networks. *Neuroimage*, 160:73–83.

Betzel, R. F., Medaglia, J. D., Papadopoulos, L., Baum, G. L., Gur, R., Gur, R., Roalf, D., Satterthwaite, T. D., and Bassett, D. S. (2017). The modular organization of human anatomical brain networks: Accounting for the cost of wiring. *Network Neuroscience*, 1(1):42–68.

Betzel, R. F., Mišić, B., He, Y., Rumschlag, J., Zuo, X.-N., and Sporns, O. (2015). Functional brain modules reconfigure at multiple scales across the human lifespan. *arXiv preprint arXiv:1510.08045*.

Birn, R. M. (2012). The role of physiological noise in resting-state functional connectivity. *Neuroimage*, 62(2):864–870.

Biswal, B., Zerrin Yetkin, F., Haughton, V. M., and Hyde, J. S. (1995). Functional connectivity in the motor cortex of resting human brain using echo-planar mri. *Magnetic resonance in medicine*, 34(4):537–541.

Blondel, V. D., Guillaume, J.-L., Lambiotte, R., and Lefebvre, E. (2008). Fast unfolding of communities in large networks. *Journal of statistical mechanics: theory and experiment*, 2008(10):P10008.

Braun, U., Schäfer, A., Walter, H., Erk, S., Romanczuk-Seiferth, N., Haddad, L., Schweiger, J. I., Grimm, O., Heinz, A., Tost, H., et al. (2015). Dynamic reconfiguration of frontal brain networks during executive cognition in humans. *Proceedings of the National Academy of Sciences*, 112(37):11678–11683.

Buckner, R. L., Andrews-Hanna, J. R., and Schacter, D. L. (2008). The brain's default network: anatomy, function, and relevance to disease.

Bullmore, E. and Sporns, O. (2009). Complex brain networks: graph theoretical analysis of structural and functional systems. *Nature reviews neuroscience*, 10(3):186–198.

Bullmore, E. and Sporns, O. (2012). The economy of brain network organization. *Nature Reviews Neuroscience*, 13(5):336–349.

Bullmore, E. T. and Bassett, D. S. (2011). Brain graphs: graphical models of the human brain connectome. *Annual review of clinical psychology*, 7:113–140.

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

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

Camerer, C. F. and Loewenstein, G. (2011). Chapter one. behavioral economics: Past, present, future. In *Advances in behavioral economics*, pages 3–52. Princeton University Press.

Cenci, M. A. (2007). Dopamine dysregulation of movement control in l-dopa-induced dyskinesia. *Trends in neurosciences*, 30(5):236–243.

Chavhan, G. B., Babyn, P. S., Thomas, B., Shroff, M. M., and Haacke, E. M. (2009). Principles, techniques, and applications of t2*-based mr imaging and its special applications. *Radiographics*, 29(5):1433–1449.

Choromański, K., Matuszak, M., and Miekisz, J. (2013). Scale-free graph with preferential attachment and evolving internal vertex structure. *Journal of Statistical Physics*, 151(6):1175–1183.

Cisek, P. and Kalaska, J. F. (2005). Neural correlates of reaching decisions in dorsal premotor cortex: specification of multiple direction choices and final selection of action. *Neuron*, 45(5):801–814.

Cisler, J. M., Bush, K., and Steele, J. S. (2014). A comparison of statistical methods for detecting context-modulated functional connectivity in fmri. *Neuroimage*, 84:1042–1052.

Cocchi, L., Zalesky, A., Fornito, A., and Mattingley, J. B. (2013). Dynamic cooperation and competition between brain systems during cognitive control. *Trends in cognitive sciences*, 17(10):493–501.

Cohen, M., Heller, A., and Ranganath, C. (2005). Functional connectivity with anterior cingulate and orbitofrontal cortices during decision-making. *Cognitive Brain Research*, 23(1):61–70.

Cole, M. W., Bassett, D. S., Power, J. D., Braver, T. S., and Petersen, S. E. (2014). Intrinsic and task-evoked network architectures of the human brain. *Neuron*, 83(1):238–251.

Colombo, M. (2014). Deep and beautiful. the reward prediction error hypothesis of dopamine. *Studies in history and philosophy of science part C: Studies in history and philosophy of biological and biomedical sciences*, 45:57–67.

Conrad, B. N., Wilkey, E. D., Yeo, D. J., and Price, G. R. (2020). Network topology of symbolic and nonsymbolic number comparison. *Network Neuroscience*, 4(3):714–745.

Cools, R., Clark, L., Owen, A. M., and Robbins, T. W. (2002). Defining the neural mechanisms of probabilistic reversal learning using event-related functional magnetic resonance imaging. *Journal of Neuroscience*, 22(11):4563–4567.

Corbetta, M. and Shulman, G. L. (2002). Control of goal-directed and stimulus-driven attention in the brain. *Nature reviews neuroscience*, 3(3):201–215.

Cordes, D., Haughton, V. M., Arfanakis, K., Carew, J. D., Turski, P. A., Moritz, C. H., Quigley, M. A., and Meyerand, M. E. (2001). Frequencies contributing to functional connectivity in the cerebral cortex in "resting-state" data. *American journal of neuroradiology*, 22(7):1326–1333.

Coste, C. P. and Kleinschmidt, A. (2016). Cingulo-opercular network activity maintains alertness. *Neuroimage*, 128:264–272.

Cox, R. W. and Hyde, J. S. (1997). Software tools for analysis and visualization of fmri data. *NMR in Biomedicine: An International Journal Devoted to the Development and Application of Magnetic Resonance In Vivo*, 10(4-5):171–178.

Dale, A. M., Fischl, B., and Sereno, M. I. (1999). Cortical surface-based analysis: I. segmentation and surface reconstruction. *Neuroimage*, 9(2):179–194.

Damoiseaux, J. S., Beckmann, C., Arigita, E. S., Barkhof, F., Scheltens, P., Stam, C., Smith, S., and Rombouts, S. (2008). Reduced resting-state brain activity in the "default network" in normal aging. *Cerebral cortex*, 18(8):1856–1864.

Damoiseaux, J. S., Rombouts, S., Barkhof, F., Scheltens, P., Stam, C. J., Smith, S. M., and Beckmann, C. F. (2006). Consistent resting-state networks across healthy subjects. *Proceedings of the national academy of sciences*, 103(37):13848–13853.

Daw, N. D., Gershman, S. J., Seymour, B., Dayan, P., and Dolan, R. J. (2011). Model-based influences on humans' choices and striatal prediction errors. *Neuron*, 69(6):1204–1215.

De Luca, M., Beckmann, C. F., De Stefano, N., Matthews, P. M., and Smith, S. M. (2006). fmri resting state networks define distinct modes of long-distance interactions in the human brain. *Neuroimage*, 29(4):1359–1367.

de Wit, S., Watson, P., Harsay, H. A., Cohen, M. X., van de Vijver, I., and Ridderinkhof, K. R. (2012). Corticostriatal connectivity underlies individual differences in the balance between habitual and goal-directed action control. *Journal of Neuroscience*, 32(35):12066–12075.

Deco, G., Tononi, G., Boly, M., and Kringelbach, M. L. (2015). Rethinking segregation and integration: contributions of whole-brain modelling. *Nature Reviews Neuroscience*, 16(7):430–439.

Dehaene, S., Kerszberg, M., and Changeux, J.-P. (1998). A neuronal model of a global workspace in effortful cognitive tasks. *Proceedings of the national Academy of Sciences*, 95(24):14529–14534.

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. *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.

DeSalvo, M. N., Douw, L., Takaya, S., Liu, H., and Stufflebeam, S. M. (2014). Task-dependent reorganization of functional connectivity networks during visual semantic decision making. *Brain and behavior*, 4(6):877–885.

Di, X. and Biswal, B. B. (2019). Toward task connectomics: examining whole-brain task modulated connectivity in different task domains. *Cerebral Cortex*, 29(4):1572–1583.

Di, X., Zhang, Z., and Biswal, B. B. (2021). Understanding psychophysiological interaction and its relations to beta series correlation. *Brain Imaging and Behavior*, 15(2):958–973.

Diuk, C., Tsai, K., Wallis, J., Botvinick, M., and Niv, Y. (2013). Hierarchical learning induces two simultaneous, but separable, prediction errors in human basal ganglia. *Journal of Neuroscience*, 33(13):5797–5805.

Dohmatob, E., Dumas, G., and Bzdok, D. (2020). Dark control: The default mode network as a reinforcement learning agent. *Human brain mapping*, 41(12):3318–3341.

Doron, K. W., Bassett, D. S., and Gazzaniga, M. S. (2012). Dynamic network structure of interhemispheric coordination. *Proceedings of the National Academy of Sciences*, 109(46):18661–18668.

Doucet, G., Naveau, M., Petit, L., Delcroix, N., Zago, L., Crivello, F., Jobard, G., Tzourio-Mazoyer, N., Mazoyer, B., Mellet, E., et al. (2011). Brain activity at rest: a multiscale hierarchical functional organization. *Journal of neurophysiology*, 105(6):2753–2763.

Eickhoff, S. B., Jbabdi, S., Caspers, S., Laird, A. R., Fox, P. T., Zilles, K., and Behrens, T. E. (2010). Anatomical and functional connectivity of cytoarchitectonic areas within the human parietal operculum. *Journal of Neuroscience*, 30(18):6409–6421.

Eickhoff, S. B., Yeo, B. T., and Genon, S. (2018). Imaging-based parcellations of the human brain. *Nature Reviews Neuroscience*, 19(11):672–686.

Erdeniz, B. and Done, J. (2019). Common and distinct functional brain networks for intuitive and deliberate decision making. *Brain sciences*, 9(7):174.

Esteban, O., Markiewicz, C. J., Blair, R. W., Moodie, C. A., Isik, A. I., Erramuzpe, A., Kent, J. D., Goncalves, M., DuPre, E., Snyder, M., et al. (2019). fmriprep: a robust preprocessing pipeline for functional mri. *Nature methods*, 16(1):111–116.

Farahani, F. V., Karwowski, W., and Lighthall, N. R. (2019). Application of graph theory for identifying connectivity patterns in human brain networks: a systematic review. *Frontiers in Neuroscience*, 13:585.

Fazeli, S. and Büchel, C. (2018). Pain-related expectation and prediction error signals in the anterior insula are not related to aversiveness. *Journal of Neuroscience*, 38(29):6461–6474.

Finc, K., Bonna, K., He, X., Lydon-Staley, D. M., Kühn, S., Duch, W., and Bassett, D. S. (2020). Dynamic reconfiguration of functional brain networks during working memory training. *Nature communications*, 11(1):1–15.

Finc, K., Bonna, K., Lewandowska, M., Wolak, T., Nikadon, J., Dreszer, J., Duch, W., and Kühn, S. (2017). Transition of the functional brain network related to increasing cognitive demands. *Human brain mapping*, 38(7):3659–3674.

Fiorillo, C. D., Tobler, P. N., and Schultz, W. (2003). Discrete coding of reward probability and uncertainty by dopamine neurons. *Science*, 299(5614):1898–1902.

Fornito, A., Harrison, B. J., Zalesky, A., and Simons, J. S. (2012). Competitive and cooperative dynamics of large-scale brain functional networks supporting recollection. *Proceedings of the National Academy of Sciences*, 109(31):12788–12793.

Fornito, A., Zalesky, A., and Bullmore, E. T. (2010). Network scaling effects in graph analytic studies of human resting-state fmri data. *Frontiers in systems neuroscience*, 4:22.

Fortunato, S. and Barthelemy, M. (2007). Resolution limit in community detection. *Proceedings of the national academy of sciences*, 104(1):36–41.

Fouragnan, E., Retzler, C., Mullinger, K., and Philiastides, M. G. (2015). Two spatiotemporally distinct value systems shape reward-based learning in the human brain. *Nature communications*, 6(1):1–11.

Fouragnan, E., Retzler, C., and Philiastides, M. G. (2018). Separate neural representations of prediction error valence and surprise: Evidence from an fmri meta-analysis. *Human brain mapping*, 39(7):2887–2906.

Fox, M. D., Snyder, A. Z., Vincent, J. L., Corbetta, M., Van Essen, D. C., and Raichle, M. E. (2005). The human brain is intrinsically organized into dynamic, anticorrelated functional networks. *Proceedings of the National Academy of Sciences*, 102(27):9673–9678.

Frank, M. J. and Claus, E. D. (2006). Anatomy of a decision: striato-orbitofrontal interactions in reinforcement learning, decision making, and reversal. *Psychological review*, 113(2):300.

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. *Proceedings of the National Academy of Sciences*, 104(41):16311–16316.

Freeman, L. C. (1977). A set of measures of centrality based on betweenness. *Sociometry*, pages 35–41.

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

Friston, K., Frith, C., Liddle, P., and Frackowiak, R. (1993). Functional connectivity: the principal-component analysis of large (pet) data sets. *Journal of Cerebral Blood Flow & Metabolism*, 13(1):5–14.

Friston, K. J. (2011). Functional and effective connectivity: a review. *Brain connectivity*, 1(1):13–36.

Gaissmaier, W. and Schooler, L. J. (2008). The smart potential behind probability matching. *Cognition*, 109(3):416–422.

Gamerman, D. and Lopes, H. F. (2006). *Markov chain Monte Carlo: stochastic simulation for Bayesian inference*. CRC Press.

Garrison, J., Erdeniz, B., and Done, J. (2013). Prediction error in reinforcement learning: a meta-analysis of neuroimaging studies. *Neuroscience & Biobehavioral Reviews*, 37(7):1297–1310.

Gerchen, M. F., Bernal-Casas, D., and Kirsch, P. (2014). Analyzing task-dependent brain network changes by whole-brain psychophysiological interactions: A comparison to conventional analysis. *Human brain mapping*, 35(10):5071–5082.

Gerraty, R. T., Davidow, J. Y., Foerde, K., Galvan, A., Bassett, D. S., and Shohamy, D. (2018). Dynamic flexibility in striatal-cortical circuits supports reinforcement learning. *Journal of Neuroscience*, 38(10):2442–2453.

Gershman, S. J. (2015). Do learning rates adapt to the distribution of rewards? *Psychonomic bulletin & review*, 22(5):1320–1327.

Gershman, S. J. (2016). Empirical priors for reinforcement learning models. *Journal of Mathematical Psychology*, 71:1–6.

Gershman, S. J., Pesaran, B., and Daw, N. D. (2009). Human reinforcement learning subdivides structured action spaces by learning effector-specific values. *Journal of Neuroscience*, 29(43):13524–13531.

Gläscher, J., Daw, N., Dayan, P., and O'Doherty, J. P. (2010). States versus rewards: dissociable neural prediction error signals underlying model-based and model-free reinforcement learning. *Neuron*, 66(4):585–595.

Gläscher, J., Hampton, A. N., and O'Doherty, J. P. (2009). Determining a role for ventromedial prefrontal cortex in encoding action-based value signals during reward-related decision making. *Cerebral cortex*, 19(2):483–495.

Glasser, M. F., Coalson, T. S., Robinson, E. C., Hacker, C. D., Harwell, J., Yacoub, E., Ugurbil, K., Andersson, J., Beckmann, C. F., Jenkinson, M., et al. (2016). A multi-modal parcellation of human cerebral cortex. *Nature*, 536(7615):171–178.

Glover, G. H. (2011). Overview of functional magnetic resonance imaging. *Neurosurgery Clinics*, 22(2):133–139.

Goldenberg, D. and Galván, A. (2015). The use of functional and effective connectivity techniques to understand the developing brain. *Developmental cognitive neuroscience*, 12:155–164.

Gordon, E. M., Laumann, T. O., Marek, S., Raut, R. V., Gratton, C., Newbold, D. J., Greene, D. J., Coalson, R. S., Snyder, A. Z., Schlaggar, B. L., et al. (2020). Default-mode network streams for coupling to language and control systems. *Proceedings of the National Academy of Sciences*, 117(29):17308–17319.

Gorgolewski, K., Burns, C. D., Madison, C., Clark, D., Halchenko, Y. O., Waskom, M. L., and Ghosh, S. S. (2011). Nipype: a flexible, lightweight and extensible neuroimaging data processing framework in python. *Frontiers in neuroinformatics*, 5:13.

Gorgolewski, K. J., Auer, T., Calhoun, V. D., Craddock, R. C., Das, S., Duff, E. P., Flandin, G., Ghosh, S. S., Glatard, T., Halchenko, Y. O., et al. (2016). The brain imaging data structure, a format for organizing and describing outputs of neuroimaging experiments. *Scientific data*, 3(1):1–9.

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.

Gratton, C., Laumann, T. O., Nielsen, A. N., Greene, D. J., Gordon, E. M., Gilmore, A. W., Nelson, S. M., Coalson, R. S., Snyder, A. Z., Schlaggar, B. L., et al. (2018). Functional brain networks are dominated by stable group and individual factors, not cognitive or daily variation. *Neuron*, 98(2):439–452.

Greicius, M. D., Krasnow, B., Reiss, A. L., and Menon, V. (2003). Functional connectivity in the resting brain: a network analysis of the default mode hypothesis. *Proceedings of the National Academy of Sciences*, 100(1):253–258.

Greicius, M. D., Supekar, K., Menon, V., and Dougherty, R. F. (2009). Resting-state functional connectivity reflects structural connectivity in the default mode network. *Cerebral cortex*, 19(1):72–78.

Greve, D. N. and Fischl, B. (2009). Accurate and robust brain image alignment using boundary-based registration. *Neuroimage*, 48(1):63–72.

Gu, S., Satterthwaite, T. D., Medaglia, J. D., Yang, M., Gur, R. E., Gur, R. C., and Bassett, D. S. (2015). Emergence of system roles in normative neurodevelopment. *Proceedings of the National Academy of Sciences*, 112(44):13681–13686.

Guimera, R. and Amaral, L. A. N. (2005). Functional cartography of complex metabolic networks. *nature*, 433(7028):895–900.

Guo, R., Böhmer, W., Hebart, M., Chien, S., Sommer, T., Obermayer, K., and Gläscher, J. (2016). Interaction of instrumental and goal-directed learning modulates prediction error representations in the ventral striatum. *Journal of Neuroscience*, 36(50):12650–12660.

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

Hare, T. A., Camerer, C. F., and Rangel, A. (2009). Self-control in decision-making involves modulation of the vmpfc valuation system. *Science*, 324(5927):646–648.

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. *Journal of neuroscience*, 28(22):5623–5630.

Hauser, T. U., Iannaccone, R., Walitza, S., Brandeis, D., and Brem, S. (2015). Cognitive flexibility in adolescence: neural and behavioral mechanisms of reward prediction error processing in adaptive decision making during development. *Neuroimage*, 104:347–354.

Hayes, D. J., Duncan, N. W., Xu, J., and Northoff, G. (2014). A comparison of neural responses to appetitive and aversive stimuli in humans and other mammals. *Neuroscience & Biobehavioral Reviews*, 45:350–368.

Hilger, K., Ekman, M., Fiebach, C. J., and Basten, U. (2017). Efficient hubs in the intelligent brain: Nodal efficiency of hub regions in the salience network is associated with general intelligence. *Intelligence*, 60:10–25.

Homberg, J. R. (2012). Serotonin and decision making processes. *Neuroscience & Biobehavioral Reviews*, 36(1):218–236.

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.

Huckins, J. F., Adeyemo, B., Miezin, F. M., Power, J. D., Gordon, E. M., Laumann, T. O., Heatherton, T. F., Petersen, S. E., and Kelley, W. M. (2019). Reward-related regions form a preferentially coupled system at rest. *Human brain mapping*, 40(2):361–376.

Humphries, M. D. and Gurney, K. (2008). Network 'small-world-ness': a quantitative method for determining canonical network equivalence. *PloS one*, 3(4):e0002051.

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

Jenkinson, M. (2003). Fast, automated, n-dimensional phase-unwrapping algorithm. *Magnetic Resonance in Medicine: An Official Journal of the International Society for Magnetic Resonance in Medicine*, 49(1):193–197.

Jung, R. E. and Haier, R. J. (2007). The parieto-frontal integration theory (p-fit) of intelligence: converging neuroimaging evidence. *Behavioral and Brain Sciences*, 30(2):135–154.

Kahneman, D. and Tversky, A. (1979). Prospect theory: An analysis of decision under risk. *Econometrica*, 47(2):363–391.

Kahnt, T., Heinzle, J., Park, S. Q., and Haynes, J.-D. (2011). Decoding the formation of reward predictions across learning. *Journal of Neuroscience*, 31(41):14624–14630.

Kahnt, T., Park, S. Q., Cohen, M. X., Beck, A., Heinz, A., and Wrase, J. (2009). Dorsal striatal–midbrain connectivity in humans predicts how reinforcements are used to guide decisions. *Journal of cognitive neuroscience*, 21(7):1332–1345.

Kamin, L. (1969). Predictability, surprise, attention, and conditioning. in ba campbell & rm church (eds.), punishment and aversive behavior (pp. 279-296). *New York: Appleton-Century-Crofts*.

Kass, R. E., Carlin, B. P., Gelman, A., and Neal, R. M. (1998). Markov chain monte carlo in practice: a roundtable discussion. *The American Statistician*, 52(2):93–100.

Katahira, K. (2015). The relation between reinforcement learning parameters and the influence of reinforcement history on choice behavior. *Journal of Mathematical Psychology*, 66:59–69.

Katahira, K. and Toyama, A. (2021). Revisiting the importance of model fitting for model-based fmri: It does matter in computational psychiatry. *PLoS computational biology*, 17(2):e1008738.

Khaw, M. W., Glimcher, P. W., and Louie, K. (2017). Normalized value coding explains dynamic adaptation in the human valuation process. *Proceedings of the National Academy of Sciences*, 114(48):12696–12701.

Klein, A., Ghosh, S. S., Bao, F. S., Giard, J., Häme, Y., Stavsky, E., Lee, N., Rossa, B., Reuter, M., Chaibub Neto, E., et al. (2017). Mindboggling morphometry of human brains. *PLoS computational biology*, 13(2):e1005350.

Kumar, P., Goer, F., Murray, L., Dillon, D. G., Beltzer, M. L., Cohen, A. L., Brooks, N. H., and Pizzagalli, D. A. (2018). Impaired reward prediction error encoding and striatal-midbrain connectivity in depression. *Neuropsychopharmacology*, 43(7):1581–1588.

Lancichinetti, A. and Fortunato, S. (2012). Consensus clustering in complex networks. *Scientific reports*, 2(1):1–7.

Lancichinetti, A., Fortunato, S., and Kertész, J. (2009). Detecting the overlapping and hierarchical community structure in complex networks. *New journal of physics*, 11(3):033015.

Landini, L., Positano, V., and Santarelli, M. (2018). *Advanced image processing in magnetic resonance imaging*. CRC press.

Langer, N., Pedroni, A., Gianotti, L. R., Hänggi, J., Knoch, D., and Jäncke, L. (2012). Functional brain network efficiency predicts intelligence. *Human brain mapping*, 33(6):1393–1406.

Lee, M. D. (2011). How cognitive modeling can benefit from hierarchical bayesian models. *Journal of Mathematical Psychology*, 55(1):1–7.

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

Lei, W., Liu, K., Chen, G., Tolomeo, S., Liu, C., Peng, Z., Liu, B., Liang, X., Huang, C., Xiang, B., et al. (2020). Blunted reward prediction error signals in internet gaming disorder. *Psychological Medicine*, pages 1–10.

Levy, D. J. and Glimcher, P. W. (2012). The root of all value: a neural common currency for choice. *Current opinion in neurobiology*, 22(6):1027–1038.

Li, J., McClure, S. M., King-Casas, B., and Read Montague, P. (2006). Policy adjustment in a dynamic economic game. *PLoS One*, 1(1):e103.

Lin, A., Adolphs, R., and Rangel, A. (2012). Social and monetary reward learning engage overlapping neural substrates. *Social cognitive and affective neuroscience*, 7(3):274–281.

Liu, X., Hairston, J., Schrier, M., and Fan, J. (2011). Common and distinct networks underlying reward valence and processing stages: a meta-analysis of functional neuroimaging studies. *Neuroscience & Biobehavioral Reviews*, 35(5):1219–1236.

Louie, K., Khaw, M. W., and Glimcher, P. W. (2013). Normalization is a general neural mechanism for context-dependent decision making. *Proceedings of the National Academy of Sciences*, 110(15):6139–6144.

Lu, H., Zou, Q., Gu, H., Raichle, M. E., Stein, E. A., and Yang, Y. (2012). Rat brains also have a default mode network. *Proceedings of the National Academy of Sciences*, 109(10):3979–3984.

Luce, R. D. (1957). A theory of individual choice behavior. Technical report, Columbia University New York Bureau of Applied Social Research.

Maia, T. V. and Frank, M. J. (2011). From reinforcement learning models to psychiatric and neurological disorders. *Nature neuroscience*, 14(2):154–162.

Mantini, D., Gerits, A., Nelissen, K., Durand, J.-B., Joly, O., Simone, L., Sawamura, H., Wardak, C., Orban, G. A., Buckner, R. L., et al. (2011). Default mode of brain function in monkeys. *Journal of Neuroscience*, 31(36):12954–12962.

Mason, O. and Verwoerd, M. (2007). Graph theory and networks in biology. *IET systems biology*, 1(2):89–119.

Mattar, M. G., Thompson-Schill, S. L., and Bassett, D. S. (2018). The network architecture of value learning. *Network Neuroscience*, 2(02):128–149.

Mattfeld, A. T., Gluck, M. A., and Stark, C. E. (2011). Functional specialization within the striatum along both the dorsal/ventral and anterior/posterior axes during associative learning via reward and punishment. *Learning & Memory*, 18(11):703–711.

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.

Meder, D., Madsen, K. H., Hulme, O., and Siebner, H. R. (2016). Chasing probabilities—signaling negative and positive prediction errors across domains. *Neuroimage*, 134:180–191.

Meder, D., Rabe, F., Morville, T., Madsen, K. H., Koudahl, M. T., Dolan, R. J., Siebner, H. R., and Hulme, O. J. (2019). Ergodicity-breaking reveals time optimal decision making in humans. *arXiv preprint arXiv:1906.04652*.

Meilă, M. (2007). Comparing clusterings—an information based distance. *Journal of multivariate analysis*, 98(5):873–895.

Metereau, E. and Dreher, J.-C. (2013). Cerebral correlates of salient prediction error for different rewards and punishments. *Cerebral Cortex*, 23(2):477–487.

Mhuircheartaigh, R. N., Rosenorn-Lanng, D., Wise, R., Jbabdi, S., Rogers, R., and Tracey, I. (2010). Cortical and subcortical connectivity changes during decreasing levels of consciousness in humans: a functional magnetic resonance imaging study using propofol. *Journal of Neuroscience*, 30(27):9095–9102.

Milgram, S. (1967). The small world problem. *Psychology today*, 2(1):60–67.

Mitchell, H., Hamilton, T., Steggerda, F., and Bean, H. (1945). The chemical composition of the adult human body and its bearing on the biochemistry of growth. *Journal of Biological Chemistry*, 158(3):625–637.

Montague, P. R. (1999). Reinforcement learning: an introduction, by sutton, rs and barto, ag. *Trends in cognitive sciences*, 3(9):360.

Montague, P. R., Dayan, P., Nowlan, S. J., Pouget, A., and Sejnowski, T. (1993). Using aperiodic reinforcement for directed self-organization during development. In *Advances in neural information processing systems*, pages 969–976.

Montague, P. R., Dayan, P., and Sejnowski, T. J. (1996). A framework for mesencephalic dopamine systems based on predictive hebbian learning. *Journal of neuroscience*, 16(5):1936–1947.

Montague, R. (2007). *Your brain is (almost) perfect: How we make decisions*. Penguin.

Mumford, J. A., Turner, B. O., Ashby, F. G., and Poldrack, R. A. (2012). Deconvolving bold activation in event-related designs for multivoxel pattern classification analyses. *Neuroimage*, 59(3):2636–2643.

Münte, T. F., Heldmann, M., Hinrichs, H., Marco-Pallares, J., Krämer, U. M., Sturm, V., and Heinze, H.-J. (2008). Nucleus accumbens is involved in human action monitoring: evidence from invasive electrophysiological recordings. *Frontiers in Human Neuroscience*, 2:11.

Nakahara, H., Itoh, H., Kawagoe, R., Takikawa, Y., and Hikosaka, O. (2004). Dopamine neurons can represent context-dependent prediction error. *Neuron*, 41(2):269–280.

Namburi, P., Al-Hasani, R., Calhoon, G. G., Bruchas, M. R., and Tye, K. M. (2016). Architectural representation of valence in the limbic system. *Neuropsychopharmacology*, 41(7):1697–1715.

Newman, M. E. and Girvan, M. (2004). Finding and evaluating community structure in networks. *Physical review E*, 69(2):026113.

Nichols, T. E. and Holmes, A. P. (2002). Nonparametric permutation tests for functional neuroimaging: a primer with examples. *Human brain mapping*, 15(1):1–25.

Nickchen, K., Boehme, R., del Mar Amador, M., Hälbig, T. D., Dehnicke, K., Panneck, P., Behr, J., Prass, K., Heinz, A., Deserno, L., et al. (2017). Reversal learning reveals cognitive deficits and altered prediction error encoding in the ventral striatum in huntington's disease. *Brain imaging and behavior*, 11(6):1862–1872.

Nieuwenhuis, S., Aston-Jones, G., and Cohen, J. D. (2005). Decision making, the p3, and the locus coeruleus–norepinephrine system. *Psychological bulletin*, 131(4):510.

Nilsson, H., Rieskamp, J., and Wagenmakers, E.-J. (2011). Hierarchical bayesian parameter estimation for cumulative prospect theory. *Journal of Mathematical Psychology*, 55(1):84–93.

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. *Journal of Neuroscience*, 32(2):551–562.

O'Doherty, J., Dayan, P., Schultz, J., Deichmann, R., Friston, K., and Dolan, R. J. (2004). Dissociable roles of ventral and dorsal striatum in instrumental conditioning. *science*, 304(5669):452–454.

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.

O'Sullivan, N., Szczepanowski, R., El-Deredy, W., Mason, L., and Bentall, R. P. (2011). fmri evidence of a relationship between hypomania and both increased goal-sensitivity and positive outcome-expectancy bias. *Neuropsychologia*, 49(10):2825–2835.

Oyama, K., Hernádi, I., Iijima, T., and Tsutsui, K.-I. (2010). Reward prediction error coding in dorsal striatal neurons. *Journal of neuroscience*, 30(34):11447–11457.

Palminteri, S., Czernecki, V., Justo, D., Jauffret, C., Karachi, C., Capelle, L., Durr, A., and Pessiglione, M. (2010). Brain opponent systems for reward and punishment learning. In *Front. Comput. Neurosci. Conference Abstract: Computations, Decisions and Movement. doi: 10.3389/conf. fnins*, volume 18.

Palminteri, S., Khamassi, M., Joffily, M., and Coricelli, G. (2015). Contextual modulation of value signals in reward and punishment learning. *Nature communications*, 6(1):1–14.

Palminteri, S. and Pessiglione, M. (2017). Opponent brain systems for reward and punishment learning: causal evidence from drug and lesion studies in humans. In *Decision neuroscience*, pages 291–303. Elsevier.

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.

Pavlov, I. P. (1928). *Lectures on conditioned reflexes*. Lawrence.

Peirce, J. W. (2007). Psychopy—psychophysics software in python. *Journal of neuroscience methods*, 162(1-2):8–13.

Power, J. D., Cohen, A. L., Nelson, S. M., Wig, G. S., Barnes, K. A., Church, J. A., Vogel, A. C., Laumann, T. O., Miezin, F. M., Schlaggar, B. L., et al. (2011). Functional network organization of the human brain. *Neuron*, 72(4):665–678.

Raichle, M. E. (2015). The brain's default mode network. *Annual review of neuroscience*, 38:433–447.

Raichle, M. E., MacLeod, A. M., Snyder, A. Z., Powers, W. J., Gusnard, D. A., and Shulman, G. L. (2001). A default mode of brain function. *Proceedings of the National Academy of Sciences*, 98(2):676–682.

Rangel, A., Camerer, C., and Montague, P. R. (2008). A framework for studying the neurobiology of value-based decision making. *Nature reviews neuroscience*, 9(7):545–556.

Rangel, A. and Clithero, J. A. (2012). Value normalization in decision making: theory and evidence. *Current opinion in neurobiology*, 22(6):970–981.

Reichardt, J. and Bornholdt, S. (2006). Statistical mechanics of community detection. *Physical review E*, 74(1):016110.

Reiter, A. M., Koch, S. P., Schröger, E., Hinrichs, H., Heinze, H.-J., Deserno, L., and Schlagenhauf, F. (2016). The feedback-related negativity codes components of abstract inference during reward-based decision-making. *Journal of cognitive neuroscience*, 28(8):1127–1138.

Rescorla, R. A. (2002). Comparison of the rates of associative change during acquisition and extinction. *Journal of Experimental Psychology: Animal Behavior Processes*, 28(4):406.

Rescorla, R. A. and Wagner, A. R. (1972). A theory of Pavlovian conditioning: Variations on the effectiveness of reinforcement and non-reinforcement. In Black, A. H. and Prokasy, W. F., editors, *Classical conditioning II: Current research and theory*, pages 64–99. Appleton-Century-Crofts, New York.

Riaz, F. and Ali, K. M. (2011). Applications of graph theory in computer science. In *2011 Third International Conference on Computational Intelligence, Communication Systems and Networks*, pages 142–145. IEEE.

Ribas-Fernandes, J. J., Solway, A., Diuk, C., McGuire, J. T., Barto, A. G., Niv, Y., and Botvinick, M. M. (2011). A neural signature of hierarchical reinforcement learning. *Neuron*, 71(2):370–379.

Rigoli, F. (2019). Reference effects on decision-making elicited by previous rewards. *Cognition*, 192:104034.

Rigoli, F., Friston, K. J., and Dolan, R. J. (2016). Neural processes mediating contextual influences on human choice behaviour. *Nature communications*, 7(1):1–11.

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

Rissman, J., Gazzaley, A., and D'Esposito, M. (2004). Measuring functional connectivity during distinct stages of a cognitive task. *Neuroimage*, 23(2):752–763.

Roesch, M. R., Calu, D. J., and Schoenbaum, G. (2007). Dopamine neurons encode the better option in rats deciding between differently delayed or sized rewards. *Nature neuroscience*, 10(12):1615–1624.

Rolls, E. T., McCabe, C., and Redoute, J. (2008). Expected value, reward outcome, and temporal difference error representations in a probabilistic decision task. *Cerebral cortex*, 18(3):652–663.

Roy, M., Shohamy, D., and Wager, T. D. (2012). Ventromedial prefrontal-subcortical systems and the generation of affective meaning. *Trends in cognitive sciences*, 16(3):147–156.

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

Rubinov, M. and Sporns, O. (2011). Weight-conserving characterization of complex functional brain networks. *Neuroimage*, 56(4):2068–2079.

Sadaghiani, S. and D'Esposito, M. (2015). Functional characterization of the cingulo-opercular network in the maintenance of tonic alertness. *Cerebral Cortex*, 25(9):2763–2773.

Sadler, J. R., Shearrer, G. E., Acosta, N. T., Papantoni, A., Cohen, J. R., Small, D. M., Park, S. Q., Gordon-Larsen, P., and Burger, K. S. (2020). Network organization during probabilistic learning via taste outcomes. *Physiology & Behavior*, 223:112962.

Sammut, Claudeand Webb, G. I., editor (2010). *Bellman Equation*, pages 97–97. Springer US, Boston, MA.

Satoh, T., Nakai, S., Sato, T., and Kimura, M. (2003). Correlated coding of motivation and outcome of decision by dopamine neurons. *Journal of neuroscience*, 23(30):9913–9923.

Satterthwaite, T. D., Elliott, M. A., Gerraty, R. T., Ruparel, K., Loughead, J., Calkins, M. E., Eickhoff, S. B., Hakonarson, H., Gur, R. C., Gur, R. E., et al. (2013). An improved framework for confound regression and filtering for control of motion artifact in the preprocessing of resting-state functional connectivity data. *Neuroimage*, 64:240–256.

Schlagenhauf, F., Huys, Q. J., Deserno, L., Rapp, M. A., Beck, A., Heinze, H.-J., Dolan, R., and Heinz, A. (2014). Striatal dysfunction during reversal learning in unmedicated schizophrenia patients. *Neuroimage*, 89:171–180.

Schonberg, T., O'Doherty, J. P., Joel, D., Inzelberg, R., Segev, Y., and Daw, N. D. (2010). Selective impairment of prediction error signaling in human dorsolateral but not ventral striatum in parkinson's disease patients: evidence from a model-based fmri study. *Neuroimage*, 49(1):772–781.

Schultz, W. (1986). Responses of midbrain dopamine neurons to behavioral trigger stimuli in the monkey. *Journal of neurophysiology*, 56(5):1439–1461.

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. *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.

Schwarz, G. (1978). Estimating the dimension of a model. *The annals of statistics*, pages 461–464.

Seger, C. A., Peterson, E. J., Cincotta, C. M., Lopez-Paniagua, D., and Anderson, C. W. (2010). Dissociating the contributions of independent corticostriatal systems to visual categorization learning through the use of reinforcement learning modeling and granger causality modeling. *Neuroimage*, 50(2):644–656.

Seymour, B., Daw, N., Dayan, P., Singer, T., and Dolan, R. (2007). Differential encoding of losses and gains in the human striatum. *Journal of Neuroscience*, 27(18):4826–4831.

Seymour, B., O'Doherty, J. P., Dayan, P., Koltzenburg, M., Jones, A. K., Dolan, R. J., Friston, K. J., and Frackowiak, R. S. (2004). Temporal difference models describe higher-order learning in humans. *Nature*, 429(6992):664–667.

Shine, J. M., Bissett, P. G., Bell, P. T., Koyejo, O., Balsters, J. H., Gorgolewski, K. J., Moodie, C. A., and Poldrack, R. A. (2016). The dynamics of functional brain networks: integrated network states during cognitive task performance. *Neuron*, 92(2):544–554.

Shteingart, H. and Loewenstein, Y. (2014). Reinforcement learning and human behavior. *Current Opinion in Neurobiology*, 25:93–98.

Smith, S. M., Fox, P. T., Miller, K. L., Glahn, D. C., Fox, P. M., Mackay, C. E., Filippini, N., Watkins, K. E., Toro, R., Laird, A. R., et al. (2009). Correspondence of the brain's functional architecture during activation and rest. *Proceedings of the national academy of sciences*, 106(31):13040–13045.

Sporns, O. (2013). Network attributes for segregation and integration in the human brain. *Current opinion in neurobiology*, 23(2):162–171.

Sporns, O. (2014). Contributions and challenges for network models in cognitive neuroscience. *Nature neuroscience*, 17(5):652–660.

Sporns, O. and Betzel, R. F. (2016). Modular brain networks. *Annual review of psychology*, 67:613–640.

Spreng, R. N., Stevens, W. D., Chamberlain, J. P., Gilmore, A. W., and Schacter, D. L. (2010). Default network activity, coupled with the frontoparietal control network, supports goal-directed cognition. *Neuroimage*, 53(1):303–317.

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

Sutton, R. S. and Barto, A. G. (2018). *Reinforcement learning: An introduction*. MIT press.

Tagliazucchi, E. and Laufs, H. (2014). Decoding wakefulness levels from typical fmri resting-state data reveals reliable drifts between wakefulness and sleep. *Neuron*, 82(3):695–708.

Tanaka, S. C., Doya, K., Okada, G., Ueda, K., Okamoto, Y., and Yamawaki, S. (2016). Prediction of immediate and future rewards differentially recruits cortico-basal ganglia loops. In *Behavioral economics of preferences, choices, and happiness*, pages 593–616. Springer.

Thomas Yeo, B., Krienen, F. M., Sepulcre, J., Sabuncu, M. R., Lashkari, D., Hollinshead, M., Roffman, J. L., Smoller, J. W., Zöllei, L., Polimeni, J. R., et al. (2011). The organization of the human cerebral cortex estimated by intrinsic functional connectivity. *Journal of neurophysiology*, 106(3):1125–1165.

Tosh, C. R. and McNally, L. (2015). The relative efficiency of modular and non-modular networks of different size. *Proceedings of the Royal Society B: Biological Sciences*, 282(1802):20142568.

Tustison, N. J., Avants, B. B., Cook, P. A., Zheng, Y., Egan, A., Yushkevich, P. A., and Gee, J. C. (2010). N4itk: improved n3 bias correction. *IEEE transactions on medical imaging*, 29(6):1310–1320.

Valentin, V. V. and O'Doherty, J. P. (2009). Overlapping prediction errors in dorsal striatum during instrumental learning with juice and money reward in the human brain. *Journal of neurophysiology*, 102(6):3384–3391.

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.

Van Den Heuvel, M., Mandl, R., and Hulshoff Pol, H. (2008). Normalized cut group clustering of resting-state fmri data. *PloS one*, 3(4):e2001.

Van Den Heuvel, M. P., Stam, C. J., Kahn, R. S., and Pol, H. E. H. (2009). Efficiency of functional brain networks and intellectual performance. *Journal of Neuroscience*, 29(23):7619–7624.

van Doorn, J., van den Bergh, D., Böhm, U., Dablander, F., Derks, K., Draws, T., Etz, A., Evans, N. J., Gronau, Q. F., Haaf, J. M., et al. (2021). The jasp guidelines for conducting and reporting a bayesian analysis. *Psychonomic Bulletin & Review*, 28(3):813–826.

Vanderveldt, A., Oliveira, L., and Green, L. (2016). Delay discounting: pigeon, rat, human—does it matter? *Journal of Experimental Psychology: Animal learning and cognition*, 42(2):141.

Vatansever, D., Menon, D. K., Manktelow, A. E., Sahakian, B. J., and Stamatakis, E. A. (2015). Default mode dynamics for global functional integration. *Journal of Neuroscience*, 35(46):15254–15262.

Vincent, J. L., Kahn, I., Snyder, A. Z., Raichle, M. E., and Buckner, R. L. (2008). Evidence for a frontoparietal control system revealed by intrinsic functional connectivity. *Journal of neurophysiology*, 100(6):3328–3342.

Vossel, S., Geng, J. J., and Fink, G. R. (2014). Dorsal and ventral attention systems: distinct neural circuits but collaborative roles. *The Neuroscientist*, 20(2):150–159.

Waelti, P., Dickinson, A., and Schultz, W. (2001). Dopamine responses comply with basic assumptions of formal learning theory. *Nature*, 412(6842):43–48.

Wagner, A. and Rescorla, R. (1972). Inhibition in pavlovian conditioning: Application of a theory. *Inhibition and learning*, pages 301–336.

Watanabe, N., Sakagami, M., and Haruno, M. (2013). Reward prediction error signal enhanced by striatum–amygdala interaction explains the acceleration of probabilistic reward learning by emotion. *Journal of Neuroscience*, 33(10):4487–4493.

Watson, J. B. (1930). *Behaviorism*. Phoenix books.

Watts, D. J. and Strogatz, S. H. (1998). Collective dynamics of 'small-world'networks. *nature*, 393(6684):440–442.

Wilson, R. C. and Niv, Y. (2015). Is model fitting necessary for model-based fmri? *PLoS computational biology*, 11(6):e1004237.

Wise, R. A. (1982). Neuroleptics and operant behavior: the anhedonia hypothesis. *Behavioral and brain sciences*, 5(1):39–53.

Wise, R. A. (1996). Addictive drugs and brain stimulation reward. *Annual review of neuroscience*, 19(1):319–340.

Wu, K., Taki, Y., Sato, K., Hashizume, H., Sassa, Y., Takeuchi, H., Thyreau, B., He, Y., Evans, A. C., Li, X., et al. (2013). Topological organization of functional brain networks in healthy children: differences in relation to age, sex, and intelligence. *PloS one*, 8(2):e55347.

Xiong, J., Parsons, L. M., Gao, J.-H., and Fox, P. T. (1999). Interregional connectivity to primary motor cortex revealed using mri resting state images. *Human brain mapping*, 8(2-3):151–156.

Yacubian, J., Gläscher, J., Schroeder, K., Sommer, T., Braus, D. F., and Büchel, C. (2006). Dissociable systems for gain-and loss-related value predictions and errors of prediction in the human brain. *Journal of Neuroscience*, 26(37):9530–9537.

Yarkoni, T., Markiewicz, C. J., de la Vega, A., Gorgolewski, K. J., Salo, T., Halchenko, Y. O., McNamara, Q., DeStasio, K., Poline, J.-B., Petrov, D., et al. (2019). Pybids: Python tools for bids datasets. *Journal of open source software*, 4(40).

Yoon, U., Fonov, V. S., Perusse, D., Evans, A. C., Group, B. D. C., et al. (2009). The effect of template choice on morphometric analysis of pediatric brain data. *Neuroimage*, 45(3):769–777.

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.

Zhang, Y., Brady, J. M., and Smith, S. (2000). Hidden markov random field model for segmentation of brain mr image. In *Medical Imaging 2000: Image Processing*, volume 3979, pages 1126–1137. International Society for Optics and Photonics.

Appendix A

Supplementary information

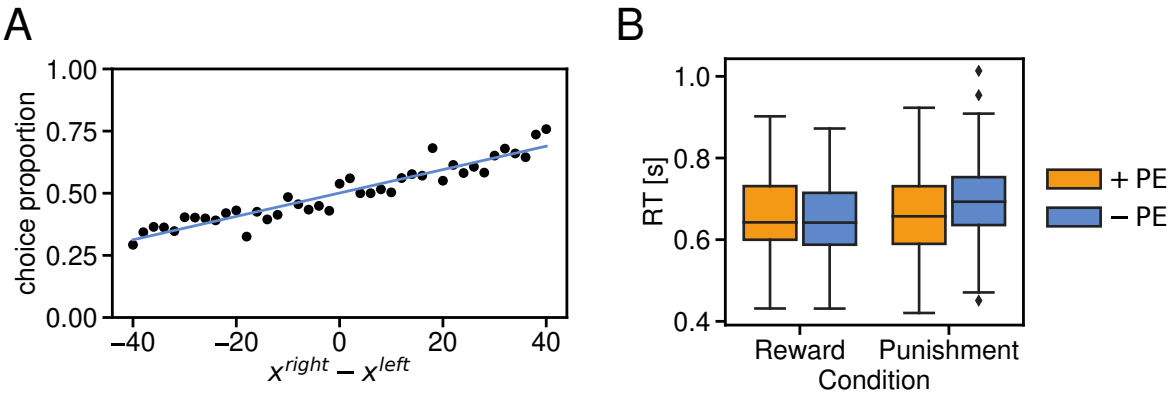

Fig. A.1 Simple behavioral measures. (**A**) Relationship between the difference in punishment/reward magnitude and choice proportion for right box. Higher difference between reward/punishment magnitudes led to higher probability of choosing more profitable box. (**B**) Mean choice time (reaction time) in seconds for both task condition and prediction error signs.

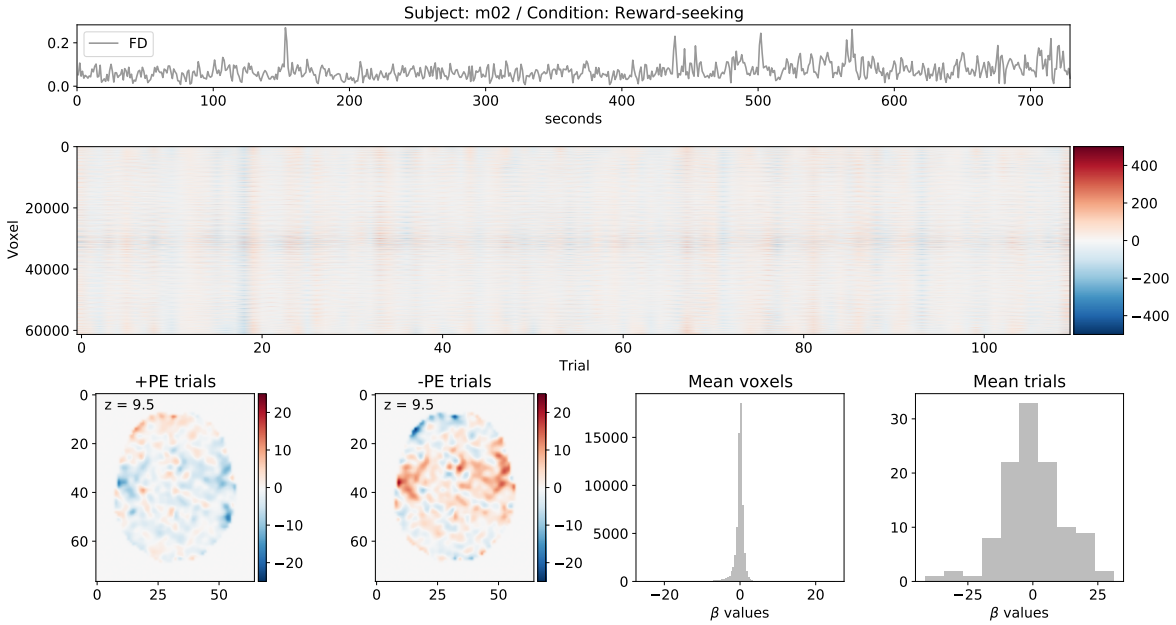

Fig. A.2 Example beta maps summary for single subject and task condition. Different characteristics of trial-wise beta maps in BSC analysis. The top panel shows the timecourse of framewise displacement FD. The middle panel shows the carpet plot of beta estimates – x-axis corresponds to 110 trials of the PRL task, y-axis corresponds to all voxels within the brain mask, color indicates trial-evoked activation. Two bottom-left panels show a coronal slice for z=9.5 for mean brain activation during +PE and -PE trials. Two bottom-right panels show beta-values distributions averaged across trials and voxels.

Table A.1 Consensus partition composition for structural resolution parameter $\gamma = 1.5$. For abbreviations see **Table 4.4**.

Module	Condition	n	Large-scale networks														
			+PE	DM	−PE	FP	MEM	SAL	UNC	SUB	VAT	CER	SOM	CO	AUD	DAT	VIS
Defualt mode	+PE	54	0,43	0,80	0,00	0,00	0,00	0,00	0,26	0,00	0,25	0,00	0,00	0,00	0,00	0,00	0,00
	−PE	52	0,43	0,76	0,00	0,00	0,00	0,00	0,26	0,00	0,25	0,00	0,00	0,00	0,00	0,00	0,00
	RS	56	0,43	0,83	0,00	0,00	0,00	0,00	0,26	0,00	0,25	0,00	0,00	0,00	0,00	0,00	0,00
	PA	57	0,43	0,85	0,00	0,00	0,00	0,00	0,26	0,00	0,25	0,00	0,00	0,00	0,00	0,00	0,00
	all	54	0,43	0,80	0,00	0,00	0,00	0,00	0,26	0,00	0,25	0,00	0,00	0,00	0,00	0,00	0,00
Sensory	+PE	78	0,00	0,00	0,00	0,00	0,25	0,21	0,00	0,50	0,50	0,00	1,00	1,00	1,00	0,33	0,00
	−PE	71	0,00	0,00	0,00	0,00	0,25	0,14	0,00	0,50	0,38	0,00	0,83	1,00	1,00	0,44	0,00
	RS	75	0,00	0,00	0,00	0,00	0,25	0,14	0,00	0,50	0,25	0,00	1,00	1,00	1,00	0,33	0,00
	PA	68	0,00	0,00	0,00	0,00	0,25	0,21	0,00	0,50	0,50	0,25	0,74	1,00	1,00	0,11	0,00
	all	76	0,00	0,00	0,00	0,00	0,25	0,14	0,00	0,50	0,38	0,00	1,00	1,00	1,00	0,33	0,00
Visual	+PE	43	0,00	0,00	0,05	0,05	0,00	0,00	0,22	0,00	0,00	0,00	0,00	0,00	0,00	0,56	1,00
	−PE	40	0,00	0,00	0,05	0,05	0,00	0,00	0,22	0,00	0,00	0,00	0,00	0,00	0,00	0,22	1,00
	RS	43	0,00	0,00	0,05	0,00	0,00	0,00	0,26	0,00	0,00	0,00	0,00	0,00	0,00	0,56	1,00
	PA	40	0,00	0,00	0,05	0,05	0,00	0,00	0,22	0,00	0,00	0,00	0,00	0,00	0,00	0,22	1,00
	all	39	0,00	0,00	0,05	0,00	0,00	0,00	0,22	0,00	0,00	0,00	0,00	0,00	0,00	0,22	1,00
Task	+PE	45	0,29	0,04	0,48	0,86	0,75	0,36	0,22	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00
	−PE	53	0,57	0,09	0,67	0,81	0,75	0,21	0,22	0,20	0,00	0,00	0,00	0,00	0,00	0,00	0,00
	RS	44	0,29	0,04	0,52	0,86	0,75	0,21	0,22	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00
	PA	45	0,29	0,04	0,57	0,86	0,75	0,36	0,13	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00
	all	43	0,29	0,04	0,43	0,86	0,75	0,29	0,22	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00
Reward	+PE	8	0,29	0,00	0,14	0,00	0,00	0,00	0,00	0,30	0,00	0,00	0,00	0,00	0,00	0,00	0,00
	RS	5	0,29	0,00	0,10	0,00	0,00	0,00	0,00	0,10	0,00	0,00	0,00	0,00	0,00	0,00	0,00
	PA	7	0,29	0,00	0,14	0,00	0,00	0,00	0,00	0,20	0,00	0,00	0,00	0,00	0,00	0,00	0,00
	all	7	0,29	0,00	0,14	0,00	0,00	0,00	0,00	0,20	0,00	0,00	0,00	0,00	0,00	0,00	0,00
Cerebellar	+PE	4	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	1,00	0,00	0,00	0,00	0,00	0,00
	−PE	2	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,50	0,00	0,00	0,00	0,00	0,00
	RS	3	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,75	0,00	0,00	0,00	0,00	0,00
	PA	3	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,75	0,00	0,00	0,00	0,00	0,00
	all	3	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,75	0,00	0,00	0,00	0,00	0,00
Other	+PE	3	0,00	0,00	0,05	0,00	0,00	0,14	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00
	−PE	6	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,17	0,00	0,00	0,00	0,00
	−PE	3	0,00	0,06	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00
	RS	3	0,00	0,00	0,05	0,00	0,00	0,00	0,00	0,20	0,00	0,00	0,00	0,00	0,00	0,00	0,00
	RS	3	0,00	0,00	0,05	0,00	0,00	0,14	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00
	PA	9	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,26	0,00	0,00	0,00	0,00
	PA	5	0,00	0,00	0,05	0,00	0,00	0,29	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00
	PA	3	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,33	0,00
	all	3	0,00	0,00	0,05	0,00	0,00	0,14	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00
	all	3	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,33	0,00

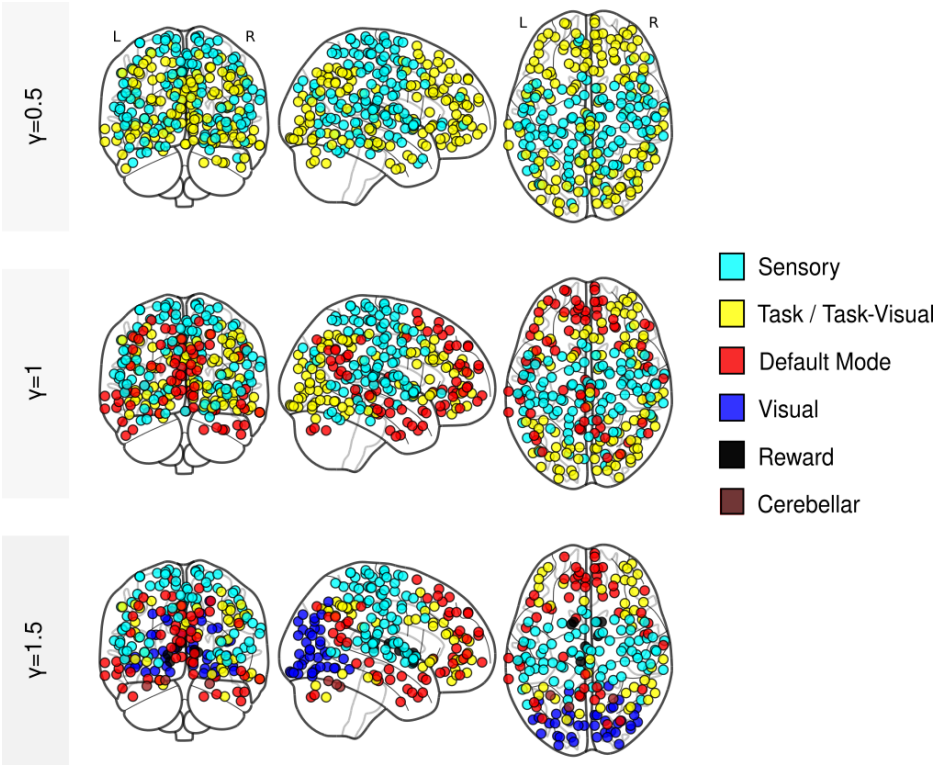

Fig. A.3 Condition-independent consensus partitions. Data-driven network division into large-scale networks for different topological scales. These divisions represent stable community structure during prediction error processing regardless of prediction error sign and task condition. For the "super-community" scale, characterized by $\gamma = 0.5$, network is divided into task-visual and sensory communities. For the intermediate scale, $\gamma = 1$, third default-mode community emerges. For the finest scale, $\gamma = 1.5$, network is divided into nine non-singleton communities. Only five of them with more than three nodes are shown: sensory, task, visual, reward and cerebellar.

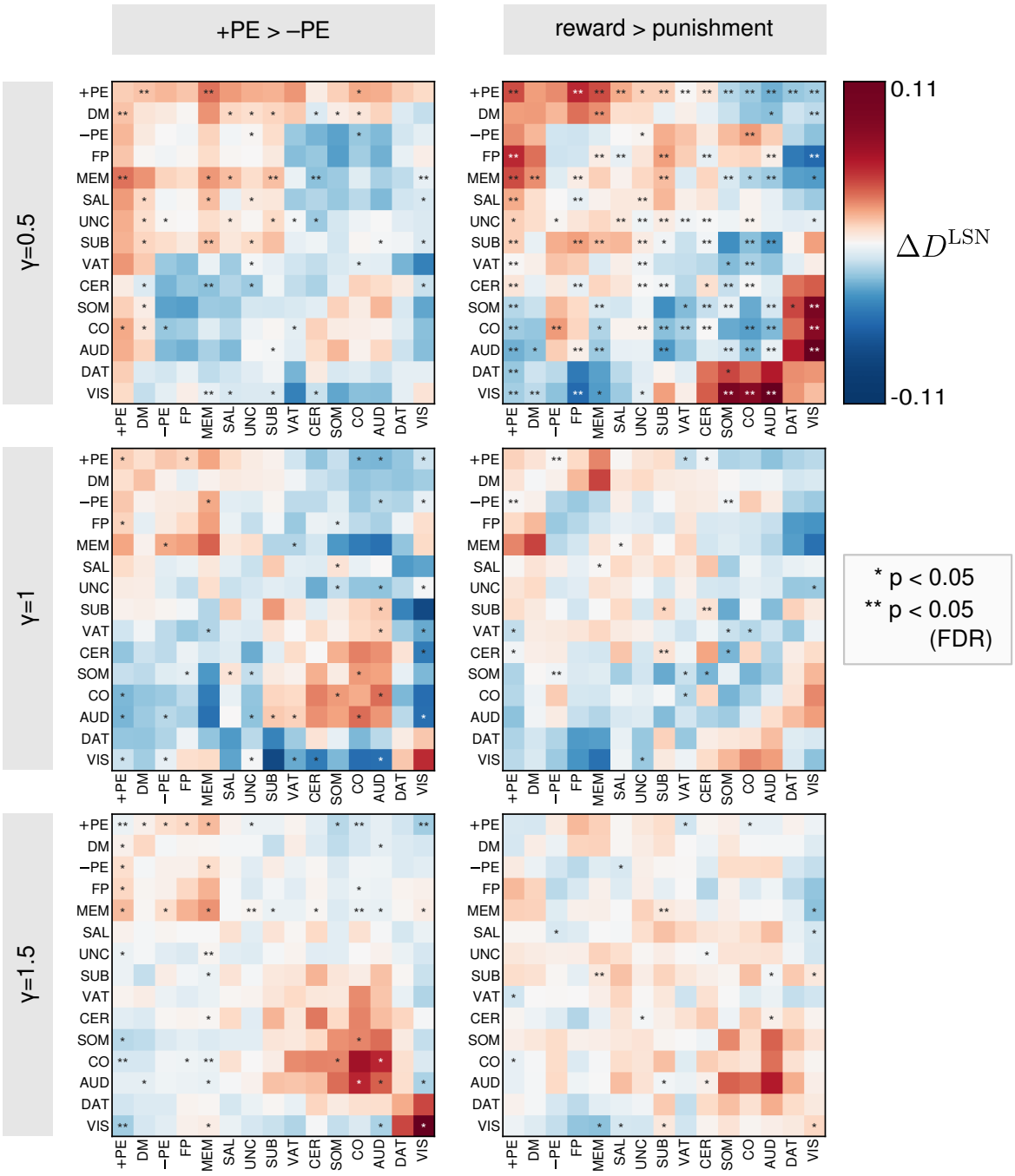

Fig. A.4 Agreement between large-scale networks for extended Power atlas. Changes in within and between-community agreement for prediction error sign and task condition. Reference communities come from extended Power partition described in section 4.6.2. LSNs abbreviations: ±PE - networks signaling positive and negative PEs; DM - default mode; FP - fronto-parietal; MEM - memory; SAL - salience; UNC - uncertain; SUB - subcortical; VAT - ventral attention; CER - cerebellar; SOM - somatomotor; CO - cingulo-opercular; AUD - auditory; DAT - dorsal attention; VIS - visual.