DISS. ETH NO. 27039

Brain function and mechanisms in adolescent depression: A computational psychiatry approach

A thesis submitted to attain the degree of
DOCTOR OF SCIENCES of ETH ZURICH
(Dr. sc. ETH Zurich)

presented by

David Willinger

MSc in Cognitive Science, University of Vienna

born on June 1, 1990
citizen of Austria

accepted on the recommendation of
Prof. Dr. Nicole Wenderoth
Prof. Dr. Silvia Brem
Prof. Dr. med. Susanne Walitza

2020

Meiner Familie gewidmet

Contents

Summary

Adolescent major depressive disorder (MDD) is associated with major impairments in the quality of life and a dramatically heightened suicidality. To improve understanding of the etiology of adolescent MDD and advance the efficacy of treatments for affected individuals, computational cognitive neuroscience has developed novel methods that find a growing number of application in psychiatric research. This dissertation project examined the neurobiology of incentive and emotion processing in adolescent MDD with functional magnetic resonance imaging (fMRI) and provides a framework for integrating methods for studying healthy and impaired neurodevelopment.

The first study (Chapter 2) presented in this thesis investigated the neurodevelopment of the functional coupling within the incentive network in reward and loss contexts. In this study, participants with an age range from 11 to 35 years covering early adolescence until young adulthood performed a monetary incentive delay task, where they had to press a button to either receive a monetary reward or had to avoid a monetary loss during individual trials. The aim of the analysis was to (1) develop a computational model capturing effects of expected values and predictions errors on response vigor and (2) their association to the developmental, age-related state of corticostriatal connectivity. With increasing age, participants improved their behavioral discrimination of low and high expected values in tandem with keeping more stable value representations. This suggests that adult participants were less prone to be negatively affected by feedback when performing the task. Transient corticostriatal connectivity changes were associated with the developmental change. Particularly, a stronger influence from the prefrontal cortex over the striatum was found and suggests a shift towards a more incentive-driven, motivated behavior in adulthood.

In the second study (Chapter 3), we employed the same monetary incentive delay task in two matched groups of adolescents (11-18 years), of which one was a group of adolescents with major depressive disorder. The focus of this investigation was to identify possible behavioral or neural disruptions of incentive processing in adolescent MDD within corticostriatal networks. Behaviorally, we established that healthy adolescents used a more complex dual learning model to learn cue-outcome associations throughout the task. The behavior of participants with MDD was best described with a simpler learning model, with the learning rate of depressed individuals being lower compared to healthy controls. This suggests a limited capacity to update representations of value in adolescent MDD. Analysis of neural correlates during loss processing revealed that differential encoding of errors in the orbitofrontal cortex in depression was linked to aberrant gain control of this region. Previous reports of disrupted reward processing in depression were not confirmed in this work.

While the first two studies were testing incentive processing, the second focus of the thesis was to develop a task for the assessment of the emotion processing circuitry in adolescent MDD. In Chapter 4, the functional architecture of the prefrontal-amygdala network in a group of 33 healthy adults with a newly developed, ecologically more valid dynamic face- and shape-matching task was examined. The aim of this study was to identify valence-sensitive connectivity patterns within the prefrontal-amygdala network that could serve as candidate pathways underlying valence-specific dysfunctional processes in (adolescent) MDD. We identified valence-dependent coupling between the amygdala and the medial prefrontal cortex that showed sensitivity to aversive and ambiguous emotional information.

In a preliminary fourth study (Chapter 5), a similar face- and shape-matching task was performed by two groups of adolescents with and without a diagnosis of MDD. We used the linear ballistic accumulator model to split the decision process into mechanistically interpretable components (e.g. processing efficiency reflected as drift rate of the model). We found strong evidence that adolescents with MDD exhibited slower evidence accumulation of ambiguously/neutrally valenced faces. For processing ambiguous faces, this less efficient information processing was associated with hypoactivity in the subgenual part of the medial prefrontal cortex. We conclude that deficient perception and evaluation of ambiguous social cues underlie adolescent MDD, providing insights into a dysfunctional emotion processing mechanism of the disorder.

This thesis extends the knowledge about typical and aberrant development of functional brain circuits during adolescence in the domains of incentive and emotion processing significantly. First, we found how the emergence of corticostriatal connectivity give rise to motivated behavior across adolescence. Second, we advanced the understanding of adolescent MDD by demonstrating that it is associated with (1) a maladaptive learning mechanism from loss and (2) inefficient processing from ambiguous emotional faces. Thus, this work provides novel mechanistic insights into one of the most debilitating psychiatric disorders and exemplifies a methodological framework of how the integration of behavioral and neural models can be harnessed to study the human brain during development and in psychiatric disorders.

Zusammenfassung

Depression im Kindes- und Jugendalter geht oft mit starken Einschränkungen der Lebensqualität und einer dramatisch erhöhten Suizidalität einher. Um das Verständnis der Ätiologie der Depression im Kindes- und Jugendalter zu verbessern und die Effektivität der Behandlungen von Betroffenen zu steigern, hat die komputationale kognitive Neurowissenschaft neue Methoden entwickelt, die eine wachsende Anzahl an Anwendungen in der psychiatrischen Forschung findet. Die vorliegende Dissertation untersuchte mit Hilfe der funktionellen Magnetresonanztomographie die Neurobiologie der Anreiz- und Emotionsverarbeitung in Jugendlichen mit Depression und stellt ein integratives Framework vor, um die gesunde und beeinträchtigte Entwicklung des Gehirns zu untersuchen.

In der ersten Studie (Kapitel 2) der vorliegenden Arbeit wurde die Entwicklung der funktionellen Kopplung innerhalb des kortiko-striatalen Netzwerks bei Erhalt von Belohnungen und Erleiden von Verlust untersucht. In dieser Studie führten die Teilnehmer im Alter von 11 bis 35 Jahren, und damit von der frühen Adoleszenz bis ins junge Erwachsenenalter, den Monetary-Incentive-Delay-Task durch. Die Ziele der Studie waren (1) ein komputationales Modell zu entwickeln, das den Effekt von Erwartungswerten («expected value») und Vorhersagefehlern («prediction error») auf die Antwortzeit beschreibt und (2) die Assoziation mit dem altersbezogenen Entwicklungsstand der kortiko-striatalen Konnektivität zu bestimmen. Die Analyse ergab, dass die Teilnehmer mit zunehmendem Alter ihre Reaktionszeit verstärkt an höhere Erwartungswerte anpassten und gleichzeitig eine stabilere Repräsentation der Erwartungswerte behielten. Das legt nahe, dass das Verhalten erwachsener Teilnehmer während der Aufgabe weniger durch das Feedback beeinflusst wurden. Diese Verhaltensänderung über das Alter hinweg hing mit einer Verstärkung der kortiko-striatalen Konnektivität zusammen. Speziell wurde eine Verstärkung der Konnektivität vom präfrontalen Kortex zum Striatum gefunden, was eine Verschiebung in Richtung eines anreizbetonten motivierten Verhaltens im Erwachsenenalter kennzeichnet.

In der zweiten Studie (Kapitel 3) wurde die gleiche Aufgabe mit zwei passenden Gruppen von Jugendlichen (11-18 Jahre) mit und ohne Depression durchgeführt. Der Fokus dieser Untersuchung lag auf der Identifikation möglicher Abweichungen auf der Verhaltens- sowie der neuralen Ebene im kortiko-striatalen Netzwerk bei der Verarbeitung von Belohnung und Verlust. Auf der Verhaltensebene wurde erhoben, dass gesunde Jugendliche ein komplexeres Dual-Lernmodell verwendeten, um Assoziationen zwischen dem Hinweisreiz und dem Resultat in der Aufgabe zu lernen. Das Verhalten der Teilnehmer mit Depression wurde am besten mit einem einfacheren Lernmodell beschrieben, in welchem die einzige Lernrate der Patienten niedriger war als die der Kontrollprobanden. Das deutet darauf hin, dass Jugendliche mit Depression eine eingeschränkte Fähigkeit besitzen, die Wert-Repräsentationen zu ändern. Eine Analyse der Bildgebungsdaten bei der Verarbeitung von Geldverlust zeigte, dass die Fehlerkodierung im orbitofrontalen Cortex in der Depression mit mangelnder Kontrolle der Signalverstärkung in dieser Region in

Verbindung stand. Die Ergebnisse von früheren Studien, in denen Änderungen in der Belohnungsverarbeitung gezeigt wurden, konnten in dieser Studie nicht bestätigt werden.

Während in den Kapiteln 2 und 3 die Motivationsprozesse untersucht wurden, war der Fokus der nächsten Studie (Kapitel 4) die Entwicklung einer Aufgabe zur Testung der Emotionsverarbeitung im Gehirn von Jugendlichen mit Depression. Hier wurde die funktionelle Architektur des Präfrontal-Amygdala-Netzwerks in 33 gesunden Erwachsenen mit einer neuentwickelten, ökologisch valideren Zuordnungsaufgabe mit dynamischen Gesichtern und Formen untersucht. Das Ziel dieser Studie war valenz-sensitive Konnektivitätsmuster innerhalb des Präfrontal-Amygdala-Netzwerks zu identifizieren, die möglicherweise valenzspezifischen dysfunktionalen Prozessen in Jugendlichen mit Depression potenziell zu Grunde liegen. Unsere Ergebnisse zeigten eine valenzabhängige Kopplung zwischen der Amygdala und dem medialen präfrontalen Cortex, die eine Sensitivität zu aversiver und uneindeutiger emotionaler Information besass.

In der vorläufigen Studie in Kapitel 5 wurde eine ähnliche Zuordnungsaufgabe wie in Kapitel 4 von zwei Gruppen Jugendlicher mit und ohne Depression durchgeführt. Wir modellierten den Entscheidungsprozess mit dem Linear-Ballistic-Accumulator Modell, um ihn in mechanistisch interpretierbare Komponenten aufzutrennen (z.B. in die in der «drift rate» des Modells widerspiegelte Verarbeitungseffizienz). Wir fanden Hinweise, dass Jugendliche mit Depression langsamere Evidenzakkumulation bei Gesichtern mit uneindeutiger Emotion zeigten. Bei der Verarbeitung dieser Gesichter war die weniger effiziente Informationsverarbeitung mit Hypoaktivität im subgenualen Teil des medialen präfrontalen Cortex assoziiert. Dies lässt darauf schliessen, dass mangelndes Wahrnehmungsvermögen und eine veränderte Evaluation von undeutlichen sozialen Reizen der Depression im Kindes- und Jugendalter zu Grunde liegen. Dies liefert neue Einblicke in die dysfunktionalen Mechanismen der Emotionsverarbeitung bei der Depression.

Diese Arbeit liefert neue bedeutende Erkenntnisse über die funktionellen Hirnnetzwerke bei Kinder und Jugendlichen mit und ohne Depression in den Gebieten der Anreiz- und Emotionsverarbeitung. Wir fanden neue Befunde, wie eine Veränderung der kortikostriatalen Konnektivität das motivierte Verhalten über das Jugendalter hinweg beeinflusst. Ausserdem konnten wir das Verständnis der Depression im Kindes- und Jugendalter verbessern, indem wir (1) einen maladpativen Lernmechanismus von Verlust und (2) ineffiziente Verarbeitung von Gesichtern mit uneindeutiger Emotion zeigten. Auf diese Weise liefert diese Arbeit neue mechanistische Erkenntnisse in eine der kräftezehrendsten psychiatrischen Störungen und veranschaulicht einen methodologischen Rahmen zur Integration von Modellen der Verhaltensebene und der Neurobiologie, um das menschliche Gehirn bei der Entwicklung oder bei psychiatrischen Störungen zu charakterisieren.

1 General Introduction

Major depressive disorder (MDD) is a severe mood disorder with drastic impairments in psychosocial functioning causing high social and economic costs (R. C. Kessler, 2012; R. C. Kessler et al., 2005). It is characterized by anhedonia (i.e. loss of pleasure or interest in activities) and persistent sad or "empty" mood (Diagnostic and Statistical Manual of Mental Disorders, DSM, American Psychiatric Association, 2013, and International Statistical Classification of Diseases and Related Health Problems, ICD, World Health Organization, 1993). These core symptoms are commonly accompanied by cognitive (e.g. feeling of hopelessness or guilt, problems in decision making or concentration, suicidal ideation) and physical disturbances (e.g. aches, decreased energy, psychomotor retardation, sleep patterns; Thapar, Collishaw, Pine, & Thapar, 2012). MDD affects more than 264,000,000 people worldwide (James et al., 2018), is one of the most prevalent disorders during adolescence with an estimated 12 months prevalence of 7.5% in mid to late adolescence (Avenevoli, Swendsen, He, Burstein, & Merikangas, 2015) and has one of the highest disease burden in youth (Gore et al., 2011; Whiteford et al., 2013). Symptoms of adolescent and adult depression tend to be similar, although adolescents more often express increased irritability or anger without overt sadness (Emslie, Mayes, & Ruberu, 2005). Adolescents with MDD have an increased risk for psychosocial and educational problems (Fergusson & Woodward, 2002), substance abuse (Keenan-Miller, Hammen, & Brennan, 2007), and another depressive episode in adult life (Pine, Cohen, Gurley, Brook, & Ma, 1998). Critically, they show dramatically increased suicidal behavior (Gould et al., 1998), which is a leading cause of death and hospitalization in adolescence (Wilkinson, Kelvin, Roberts, Dubicka, & Goodyer, 2011). Effective treatments and methods for prevention are urgently needed to reduce the disease burden, nevertheless, in order to achieve this, novel approaches can help to understand the etiology of adolescent MDD.

1.1 Computational neuroscience to advance psychiatric research

A main challenge in psychiatric research arises due to the nature of current diagnostic classification schemes (DSM, ICD). They were developed for clinical use to aid diagnosis and treatment based on syndromes (i.e. set of medical signs and symptoms). This nosology is based mainly on subjective experience as opposed to more objective neurobiological or psychological dysfunction. However, a symptom-based classification allows patients with the same diagnostic label to exert a certain grade of heterogeneity (i.e. patients present with different or not-overlapping symptoms) and comorbidity (i.e. unclarity about one or separate true disease mechanisms; Buckholtz & Meyer-Lindenberg, 2012). To overcome some of the limitations of symptom-based classification for research, the Research Domain Criteria approach (RDoC, T. R. Insel, 2014) is one attempt to create a transdiagnostic taxonomy for mental disorders based on integrated evidence from several different levels of analysis (e.g. imaging and behavior). Here, it has been

proposed that computational models are a key step to link the levels and provide a more precise biological understanding of psychiatric disorders (Ferrante et al., 2019).

In the field of cognitive neuroscience, computational modeling had great success in explaining the association of cognitive processes and behavior and neural activity, thus integrating different levels of analysis (Kriegeskorte & Douglas, 2018). Based on Marr (1982), the models in different level of analysis can be distinguished in implementation level (e.g. neural network models), and more abstract algorithmic models (e.g. reinforcement learning models). Although algorithmic models are agnostic about the underlying implementation, they allow for a formal description of a cognitive process that generated the observed data. Nevertheless, when there is knowledge about how a process is implemented in neural networks, levels of analysis can be synergistically linked to test hypotheses about the neural substrate. Such analyses increase the level of specificity to investigate how cognitive processes are realized on a neuronal level (den Ouden, Daunizeau, Roiser, Friston, & Stephan, 2010).

Computational psychiatry is a growing field that develops and applies computational models to study dysfunctional processes in the brain and advance clinically useful classification and prediction (Montague, Dolan, Friston, & Dayan, 2012). Characterizing mental disorders across different levels of analysis using methods from computational neuroscience holds promise to move from symptom-based to mechanistic and quantitative representations of disease states (Huys, Maia, & Frank, 2016). This has the potential to bridge the gap between new developments of cognitive neuroscience and clinical applications by advancing insights on neurobiology and cognitive processes.

In the next sections, it is shown where the application of computational modeling has already provided important contributions in the field of cognitive neuroscience and how they allow mechanistic insight into adolescent MDD by means of providing formal descriptions of reinforcement learning and perceptual decision making.

1.2 Motivated behavior across development

Learning the consequences of one's action is critical to aid value-guided behavior. In the last years, research has begun to uncover how the brain performs computations for value-guidance, learning and behavioral adaptation. Influential work as the reinforcement learning model of Rescorla and Wagner (1972) or the Temporal Difference model of Sutton (1988) has addressed how an agent learns the actions that maximize reward in the future by interacting with its environment. Central to these models is that an agent performs an action or makes a decision to obtain a certain good associated with a value. When the agent observes the outcome, it receives feedback about the actual quality of the good. The difference between the agent's *expected value* and the outcome it observes is called *prediction error* (Rescorla & Wagner, 1972). The crucial part in reinforcement learning is that the agent will update its expectation based on the prediction error in each observation. The prediction error is weighted by a learning rate, which determines the step

size of the agent's adjustment. Importantly, such an update mechanism allows for establishing and tracking the value in environments, where obtaining a reward is probabilistic. These models have been applied to describe changes of cue-outcome associations in decision making or instrumental learning paradigms. In the latter, it is important to note that expected value (Dudman & Krakauer, 2016) and prediction errors (Bestmann, Ruge, Rothwell, & Galea, 2014) have been shown to be able to modulate motivated behavior in terms of the response vigor.

In their groundbreaking work, Schultz, Dayan, and Montague (1997) could demonstrate that in the primate brain, dopamine neurons (1) phasically increase firing when a reward is received, which (2) shift to a response to a reward predicting cue after repeated exposure (as opposed to the reward itself), and (3) phasically decrease firing when the cue appears but a reward is not obtained, thus resembling a firing pattern similar to a temporal difference reward prediction error. Ensuing human neuroimaging studies (McClure, Berns, & Montague, 2003; O'Doherty, Dayan, Friston, Critchley, & Dolan, 2003; O'Doherty et al., 2004) have found activation patterns that correspond to expected values and prediction errors in brain regions innervated by dopaminergic projections (Björklund & Dunnett, 2007). Later, studies found that these signals play critical roles in approach and avoidance learning (Palminteri et al., 2012; Rigoli, Chew, Dayan, & Dolan, 2016). Over the last years, the interest increased in how maturation of the corticostriatal networks give rise to motivated behavior in adolescence (Davidow, Insel, & Somerville, 2018).

Across development from childhood into adulthood, the brain's cognitive networks are subject to major remodeling (Andersen & Teicher, 2008; Casey, Giedd, & Thomas, 2000; Casey, Jones, & Hare, 2008). Several maturational events moderate brain development and maturation such as changes in gonadal hormones (e.g. Andersen & Teicher, 2008; Markham, Morris, & Juraska, 2007) and rearrangements in structural networks including synaptic pruning and increasing myelination during adolescence (Casey et al., 2000; Casey et al., 2008; Giedd, 2004; Giedd et al., 1999). Many cognitive functions mature alongside with adolescent neurodevelopment (Somerville, 2016) and previous work has provided a complex picture of how incentives drive changes in behavior (e.g. increase of response speed) and neural sensitivity (e.g. striatal response to incentives) across this developmental phase. Influential models proposed that a late prefrontal maturation compared to subcortical regions might in result in less efficient cognitive control to employ motivated action effortfully (Casey, Jones, & Somerville, 2011; Ernst, Pine, & Hardin, 2006; Steinberg, 2010). However, a new wave of research has yielded inconsistent reports on the effect of brain maturation on behavior, with some studies reporting better (Cohen et al., 2010; Geier, Terwilliger, Teslovich, Velanova, & Luna, 2009) weaker (C. Insel, Kastman, Glenn, & Somerville, 2017) or no age-specific difference on (Barkley-Levenson, Van Leijenhorst, & Galván, 2013; Bjork, Smith, Chen, & Hommer, 2010; Cho et al., 2013; Lamm et al., 2014; Paulsen, Hallquist, Geier, & Luna, 2015; Strang & Pollak, 2014) performance in a wide array of incentivized tasks. Thus, the role of functional maturation

of corticostriatal networks to support motivated behavior driven by incentives in reward and avoidance contexts is still unclear and requires further clarification.

1.3 Motivated behavior in adolescent MDD

MDD has been associated with different psychological phenomena related to impaired learning, thus it is important to shed more light on how the learning process is affected (Huys, Daw, & Dayan, 2015; Kube, Schwarting, Rozenkrantz, Glombiewski, & Rief, 2020; Scholl & Klein-Flügge, 2018). Maladaptive learning from positive experiences (e.g. receiving a reward or avoiding a loss) could represent a potential mechanism of the persistence of MDD. To quantify learning from the environment, computational modeling of behavior can be used to infer on latent variables that modulate motivated behavior. For example, a large expected value might speed up response speed, yet the expected value depends on the reward history within a task.

In adult MDD, blunting of reward prediction error signals in the ventral striatum (Dombrovski, Szanto, Clark, Reynolds, & Siegle, 2013; Gradin et al., 2011; Kumar et al., 2018; Kumar et al., 2008) and the orbitofrontal cortex (Rothkirch, Tonn, Köhler, & Sterzer, 2017) was reported in various neuroimaging studies. Moreover, studies have associated adult MDD with deficits in reinforcement learning in approach and avoidance context (Dombrovski et al., 2010; Vrieze et al., 2013), with prominent proposals that explain MDD as a form of maladaptive expectations (Huys et al., 2015; Kube et al., 2020). Only a few studies investigated neural sensitivity during incentive processing in adolescent MDD (see Kerestes, Davey, Stephanou, Whittle, & Harrison, 2014, for review). To date, research on neural mechanisms of incentive processing in adolescent MDD has focused on reward processing (Forbes et al., 2009; Forbes et al., 2010; Stringaris et al., 2015). Findings of these studies are generally in concordance with studies of adult MDD, demonstrating decreased striatal activity during reward anticipation and feedback processing, which was suggested to reflect a dysfunction in regions critical for reinforcement learning (Pizzagalli et al., 2009). However, none of these studies employed a computational model to describe learning trajectories across a task and systematically investigated it in adolescent MDD.

Thus, a critical gap in the literature is to find computational mechanisms of reinforcement learning in adolescent MDD. By fitting a model to the behavior (i.e. response speed), we can compare the model parameters embodying the factors that drive behavioral adaptation across the task and gain insights into learning processes in the disorder. In addition, such a model allows us to enter the derived the learning signal in a model-based neuroimaging analysis (see below) to investigate neural alterations underlying possible aberrant computations.

1.4 Emotion processing in adolescent MDD

In addition to disturbed incentive processing, a hallmark of adolescent MDD is an attentional negative bias (Beck, 1967). It has been proposed to be a vulnerability factor in adolescence and might have a causal role in the onset of depression (Jacobs, Reinecke, Gollan, & Kane, 2008). Thus, gaining more insights in the underlying neural mechanisms supporting a persistent negative attribution will advance the understanding of adolescent MDD.

The neural basis of the attentional negative bias commonly found in MDD remains only poorly understood. Prominent models propose that the bias is associated with dysfunctional affective processing (Disner, Beevers, Haigh, & Beck, 2011; Mayberg, 1997). In this view, the negativity bias stems from maladaptive bottom-up processes (from subcortical structures to higher-order cortical regions) that persist due to inefficient cognitive control and regulation processes, which can facilitate the maintenance of negative attributions. In line with that, elevated amygdala activity has been detected consistently in adolescent MDD (Gaffrey, Barch, Singer, Shenoy, & Luby, 2013; Hall et al., 2014; Mingtian et al., 2012; Redlich et al., 2018; Tao et al., 2012; Yang et al., 2010) and in minors at familial risk for MDD (Chai et al., 2015; Monk, Klein, et al., 2008), along with hypoactivation in the medial prefrontal cortex (Miller, Hamilton, Sacchet, & Gotlib, 2015) and aberrant prefrontal-amygdala connectivity (Musgrove et al., 2015).

Previous work has applied sequential sampling models (SSM) to study emotional perception and attentional bias in MDD (Ho et al., 2016; Pe, Vandekerckhove, & Kuppens, 2013). SSMs describe the decision process in a high level of abstraction making its application suitable in a wide array of simple decision-making tasks to cover different phenomena. They allow a description of the decision process by taking into account not only the distribution of reaction times, but also their interaction with error rates (Brown & Heathcote, 2008). A particular instance of an SSM is the linear ballistic accumulator (LBA) model. It conceptualizes the decision process with two (or more) separate evidence accumulators that gather evidence for all options across time. The speed of evidence accumulation is called the drift rate. As soon as one accumulator reaches the response threshold, a response is made. A higher threshold reflects more cautious responses (slower but more accurate), while a lower threshold decreases the decision time in trade for accuracy. Finally, all processes not associated with decision processing (e.g. motor preparation and execution) are reflected in the non-decision time parameter. While some parameters might be proactively adjusted during the task by an individual (e.g. response threshold), others reflect the information processing ability (e.g. drift rate, Forstmann et al., 2010). Taken together, SSMs can aid the characterization of cognitive impairments during emotion processing and allow for the association of any such impairments and their neural correlates. Here, it will be critical to understand how processing of different valenced stimuli is altered in adolescent MDD and how this may be related to aberrant processing in the prefrontal-amygdala network.

1.5 Methods to study development of brain function in healthy and patients

Functional magnetic resonance imaging (fMRI) is able to measure physiological changes in the brain that indirectly reflect neural activity. The activation of a neuronal population yields to changes in blood flow and volume and an increasing demand of oxygen, which is provided by the nearby vascular system. The difference in susceptibility effects of oxyhemoglobin and deoxyhemoglobin is called the blood-oxygenation-level dependent (BOLD) response (Ogawa, Lee, Kay, & Tank, 1990). The BOLD response has been linked to local field potentials, thus reflecting primarily input and local neural processing and not spiking output (Logothetis, Pauls, Augath, Trinath, & Oeltermann, 2001).

The variation and time course of the BOLD response depends on the specific experimental stimulation. A single stimulation (*event*) leads to an increase neural activity and oxygen supply and thereby to an increase of the BOLD signal, which peaks approximately after 4-6 seconds. When the stimulation ceases, neural activity decreases again and the BOLD signal returns to baseline levels after around 25 seconds. Blocked designs, where stimuli are repeatedly presented in longer time blocks, have been employed frequently since they have high statistical power (Friston, Holmes, Price, Büchel, & Worsley, 1999). However, more recently, event-related experimental designs that allow for the analysis of individual trials are gaining popularity in the field of cognitive neuroscience (Huettel, 2012). A particular instantiation of this design is called model-based fMRI, where the BOLD response is modulated on a trial-by-trial basis in a parametric fashion (Gläscher & O'Doherty, 2010; O'Doherty et al., 2003; O'Doherty, Hampton, & Kim, 2007). This approach allows to integrate trial-by-trial estimations of variables derived from computational models and relate them to brain activation on a single-voxel level to investigate which brain areas implement a specific computational process (Stephan, Iglesias, Heinzle, & Diaconescu, 2015).

In contrast to more conventional analyses of neuroimaging data, dynamic causal modeling (DCM; Friston, Harrison, & Penny, 2003) is a hypothesis-driven analysis approach. It uses models to assess the network dynamics of brain regions regarding their directed influences (effective connectivity) under the influence of experimental conditions. DCM treats the brain as a deterministic system, where the response in one region is determined completely by its inputs. The dynamic system is expressed by differential equations, that model latent neural states that are probabilistically mapped to the hemodynamic response by a biophysical forward model (Buxton, Wong, & Frank, 1998). Over the last years, the DCM framework has been continuously developed and advanced to application for resting-state fMRI (Friston, Kahan, Biswal, & Razi, 2014), or to Bayesian model reduction and a hierarchical framework using empirical priors (Friston et al., 2016). The DCM parameters are optimized in a way that the predicted and the observed BOLD time courses match as good as possible. The marginal likelihood of the model parameters, i.e. the model evidence, is a relative score for a model that can be used to compare competing hypotheses about the functional neural architecture embodied by the DCM (Stephan et

al., 2010). By allowing inference on how neuronal interactions between regions most likely generate the observed BOLD signal, DCM is a useful tool to study pathophysiological mechanisms on a (coarse) implementation level (Heinzle & Stephan, 2018).

1.6 General aims and hypotheses

To date, still many aspects of the etiology of adolescent MDD are unclear. In the last years, attempts to improve characterization of psychiatric disorders by scrutinizing affected cognitive and neural processes using computational models have received growing attention (Stephan et al., 2015). The overarching goal of this thesis is to gain mechanistic insights in alterations of incentive and emotion processing in adolescent MDD by applying computational models of behavior in combination with fMRI.

In the first study, we assessed the emergence of the ability to adjust instrumental vigor (as a measure for motivated behavior) based on incentives with varying magnitude and valence and investigated its relation to age-dependent changes in connectivity within the corticostriatal network. While one study reported that neurodevelopment facilitates goal-direct behavior across adolescence (C. Insel et al., 2017), the association between connectivity and ability to adjust response speed remains to be established. The aim of this study was to clarify whether such functional development of the corticostriatal network gives rise to the ability to modulate response speed as a function of incentives. We hypothesized that a strengthening of the frontostriatal connectivity across development would be associated with increased motivated behavior. Furthermore, we explored the influence of sex on the neural representations of reinforcement learning.

The second study delved into the investigation of the adaptation of motivated behavior in adolescent MDD. Here, the objective was to identify whether incentive-based adjustments of instrumental vigor or brain responses to monetary incentives are altered in individuals with MDD using model-based behavioral and neuroimaging analyses. As there is evidence from an adolescent sample at familial risk for MDD for decreased loss signaling in the orbitofrontal cortex (Jin et al., 2017), we hypothesized impaired functioning during avoidance learning in adolescent MDD. In addition, based on the previously established impairments in regions crucial for reinforcement learning (Kerestes et al., 2014), we assumed decreased reward prediction error signaling in the dopaminergic targets, particularly the striatum.

In the third study, our aim was to characterize the change of effective connectivity within the prefrontal-amygdala network while processing different emotional valence. While it has been established that this network is important for emotion processing (Fusar-Poli et al., 2009), it is still unclear how emotional valence modulates the functional coupling within the network. The goal of this study was to reveal neural mechanisms of the dynamic integration of affective information in healthy adults. We hypothesized that affective information is encoded in the coupling between the amygdala and the medial prefrontal cortex, and the medial and lateral prefrontal cortex.

The fourth preliminary study focused on shedding light on the emotional perception and its neural systems that might underlie biased processing in adolescent MDD (Disner et al., 2011). To this end, we applied a LBA model to investigate the decision-making process during a face- and shape-matching task. Our aim was to determine how valence-dependent modifications of evidence accumulation and response threshold are altered in adolescent MDD. Our hypothesis was that we would find evidence for impaired information processing during emotion processing reflected in the drift rate. In addition, we hypothesized that possible impairments in information processing are related to brain activity in the amygdala and the medial prefrontal cortex.

2 Neurodevelopment of the incentive network facilitates motivated behavior from adolescence to adulthood

David Willinger[a,b], Iliana I. Karipidis[a,b,c], Plamina Dimanova[a], Susanne Walitza[a,b], Silvia Brem[a,b]

2.1 Overview

The ability to enhance motivated performance through incentives is crucial to guide and ultimately optimize the outcome of goal-directed behavior. It remains largely unclear how motivated behavior and performance develops particularly across adolescence. Here, we used computational fMRI to assess how response vigor and its underlying neural circuitry are modulated by reward and loss in a monetary incentive delay paradigm. We demonstrate that maturational fine-tuning of functional coupling within the corticostriatal incentive circuitry from adolescence to adulthood facilitates the ability to enhance performance selectively for higher subjective values. Additionally, during feedback, we found developmental sex differences of striatal representations of reward prediction errors. Our findings suggest that a failure to utilize subjective value for motivated behavior in adolescence is rooted in immature information processing in the incentive system. This indicates that the neurocircuitry for coordination of incentivized, motivated cognitive control acts as a bottleneck for behavioral adjustments in adolescence.

2.2 Introduction

Goal-directed behavior depends fundamentally on the capacity to attribute significance to stimuli in the environment and adapt performance accordingly. This does not only include choosing between available options but also deciding about how much effort and speed to dedicate to an action (Dayan, 2012; Niv, Daw, Joel, & Dayan, 2007). While acting too slow can result in a lost opportunity, acting too fast can lead to excessive opportunity costs. Thus, optimal goal-directed behavior is an adjustment of e.g. motor control as a function of subjective relevance (Manohar et al., 2015). The ability for continuous, flexible behavioral adjustments to achieve goals is supported by cognitive control systems that can selectively improve performance by integrating motivational outcome values and available resources (Kool, Gershman, & Cushman, 2017; Kouneiher, Charron, & Koechlin, 2009; Mir et al., 2011).

Ample evidence suggests that motivated action depends on the interactions within cortico-striato-thalamic networks. The prefrontal cortex supports complex cognitive control processes including action selection, performance monitoring, and feedback-based learning (Botvinick & Braver, 2015). In turn, striatum and insula might have opponent

[a] Department of Child and Adolescent Psychiatry and Psychotherapy, University Hospital of Psychiatry Zurich, University of Zurich, Switzerland
[b] Neuroscience Center Zurich, University of Zurich and ETH Zurich, Switzerland
[c] Center for Interdisciplinary Brain Sciences Research, Stanford University School of Medicine, Stanford, CA, USA

We thank Selina Neuer for helping with recruitment and testing and Philipp Stämpfli for technical assistance.

roles in encoding the motivational value of cues ("expected value") and prediction errors in reward and loss avoidance contexts, respectively, to guide learning of action-outcome contingencies and to facilitate the selection of a candidate action in the premotor cortex (Averbeck & Costa, 2017; Niv et al., 2007; Palminteri et al., 2012). Prospective outcome has been linked to motivation by demonstrating that motoric vigor increases as a function of an unsigned expected value (Dudman & Krakauer, 2016; Manohar et al., 2015; Niv et al., 2007; Pessiglione et al., 2007; Rigoli et al., 2016). This delineates the importance of striatal and insular modulation of action selection as a function of motivational salience. In other words, the anticipated outcome value of a specific action may facilitate its selection and its ensuing execution. The signal integrated in the striatum pass via the basal ganglia to the thalamus and the cortex, where they guide motivated behavior (Haber & Knutson, 2010).

The ability of selective exertion of cognitive control based on prospective outcomes to improve performance has been well-documented in adults (Chiew & Braver, 2016; Locke & Braver, 2008; Pfabigan et al., 2014; Wrase et al., 2007; Wu, Samanez-Larkin, Katovich, & Knutson, 2014). Nevertheless, studies using incentivized tasks in adolescents have yielded a complex picture of neurodevelopmental patterns and their manifestation on a behavioral level (Davidow et al., 2018). Prior work has established a functional remodeling of this key circuit in the domains of decision-making (Barkley-Levenson & Galván, 2014; Cohen et al., 2010; Hauser, Iannaccone, Walitza, Brandeis, & Brem, 2015; Van Den Bos, Cohen, Kahnt, & Crone, 2012; Van Den Bos, Rodriguez, Schweitzer, & McClure, 2015), inhibitory control (C. Insel et al., 2017; Somerville, Hare, & Casey, 2011), and incentive anticipation (Cho et al., 2013; Lamm et al., 2014) across adolescence. Influential work of Ernst et al. (2006), Steinberg (2010), and Casey et al. (2011) has suggested that prefrontal cortex maturation improves exertion of cognitive control to perform motivated behavior and self-control adaptively. However, there are conflicting reports on behavioral manifestations of the immature adolescent control system. While some studies demonstrated improvements in behavioral performance with reward in adolescents compared to adults (Cohen et al., 2010; Geier et al., 2009), other studies investigating reward and loss processing (Bjork et al., 2010; Cho et al., 2013; Joseph, Zhu, Lynam, & Kelly, 2016; Lamm et al., 2014) did not find age-specific differences in either reaction time (RT), performance in inhibition (Paulsen et al., 2015; Strang & Pollak, 2014), or choices and loss aversion in decision-making (Barkley-Levenson et al., 2013). A recent study reported selective performance improvements in reward contexts across development (C. Insel et al., 2017), which is consistent with the idea that the deployment of cognitive resources supporting motivated behavior emerges along with the maturation of corticostriatal networks. These varying accounts indicate that more work is necessary to establish an enhanced understanding of the functional architecture of the corticostriatal system that supports the integration of control and value signals to shape behavior during incentivized processing.

Here, we performed a functional magnetic resonance imaging (fMRI) study to investigate incentive processing with varying magnitude (low, high) and valence (reward, loss) across development (11-35 years) using a well-validated monetary incentive delay (MID) paradigm (Knutson, Westdorp, Kaiser, & Hommer, 2000). We utilized a computational learning model to investigate (1) factors affecting trial-by-trial response vigor of adults and adolescents and (2) model-based brain activity and effective connectivity patterns of expected value and prediction error in reward and loss contexts. We hypothesized that the immature corticostriatal circuitry of adolescents would demonstrate less incentive-guided behavioral adaptation, i.e. lower reaction time for high incentives compared to adults (C. Insel et al., 2017). Based on studies showing elevated striatal activity during adolescence (Barkley-Levenson & Galván, 2014; Cohen et al., 2010), we predicted that activity in the ventral striatum (VS) correlates with the expected values and reward prediction errors and is increased in adolescence compared to adulthood. A weaker effective connectivity between the prefrontal cortex and the VS during anticipation in adolescents than in adults could be indicative of a protracted/late maturation of corticostriatal circuits across adolescence.

2.3 Results

First, we examined the log reaction time (logRT) and the accuracy of 67 participants that performed the MID task in the scanner (Figure 2.1). The hit rate across participants was 61.8% (*SD* = 2.0%) and thus close to the hit rate of 66% that we aimed for in the task design (Table 4.1). A mixed model analysis of log RT showed a significant main effect of condition, $F(4, 7304.1) = 2.785$, $p = 0.025$, and an age-by-condition interaction, $F(4, 7304.1) = 4.240$, $p = 0.002$. The main effect of age was not significant, $F(1, 63) = 2.397$, $p = 0.127$. Post hoc Tukey tests showed that during the high loss condition, subjects responded faster than in the neutral ($p = 0.015$), low loss ($p = 0.0002$), and low reward ($p = 0.025$) condition. During the high reward condition, participants responded faster than during the neutral ($p = 0.0001$), low loss ($p < 0.0001$) and low reward ($p = 0.0002$) condition. This indicated that subjects were not pressing the button with random speed but exerted more vigorous responses in trials with high magnitude outcomes. Correlation analysis of mean logRT for each condition revealed that logRT to high rewards decreased with the age of the participants ($r(63) = -0.256$, $p = 0.040$). LogRT in other conditions were not related to age.

Table 2.1. Behavioral data of the monetary incentive delay task. Note that the hit rate was experimentally manipulated such that a hit rate of around 66% was achieved.

	High loss	Low loss	Neutral	Low Gain	High Gain
Hit rate	63 (8)%	59 (10) %	-	60 (9) %	63 (9) %
Response time	258 (9) *ms*	267 (11) *ms*	263 (8) *ms*	263 (8) *ms*	254 (8) *ms*
Liking rating	8.7 (15.5)	22.5 (24.0)	47.9 (17.3)	74.0 (14.2)	91.8 (12.2)
Arousal rating	61.1 (30.7)	47.7 (27.8)	32.3 (20.9)	51.5 (23.7)	63.6 (29.3)

Mean (SD) across subjects. Rating range was 0-100.

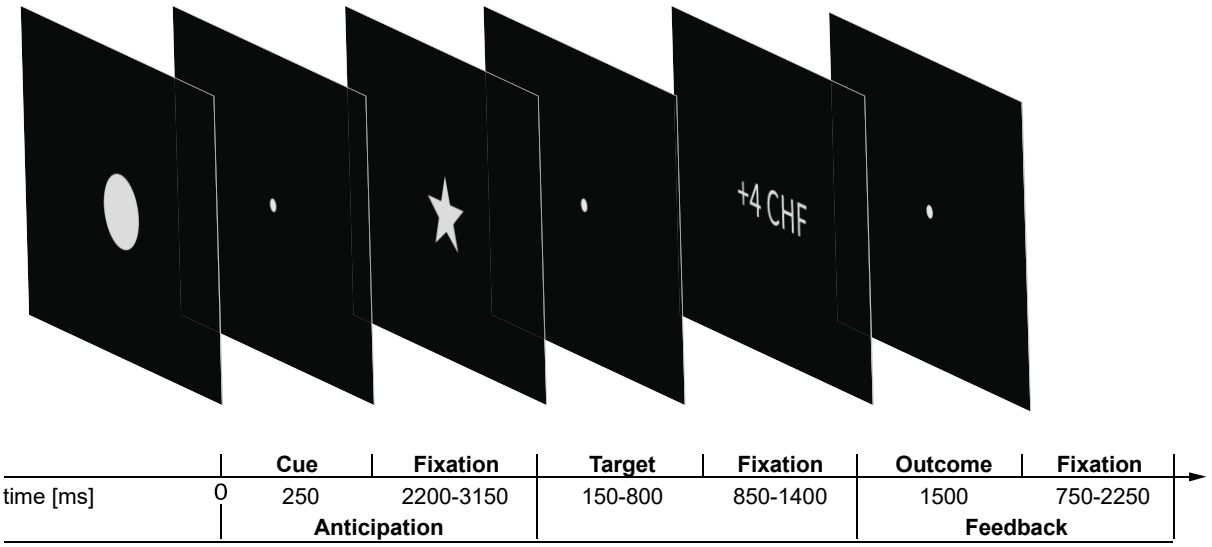

Figure 2.1. The monetary incentive delay (MID) task examines incentive anticipation and feedback processing. In the beginning of each trial, cues indicated level of magnitude (low, high) and valence (reward, loss, null) of the possible outcome. After a variable delay a target (star) was presented, that was used as go-signal where subjects were instructed to respond as fast as possible. After the next fixation period, the actual outcome was presented for 1500ms. Altogether, the task comprised 24 trials per cue, i.e. 120 trials in total. An adaptive algorithm ensured a hit rate of ≈ 66%.

2.3.1 Modeling response vigor

To assess in more detail how response vigor was modulated across the entire task, we employed a computational reinforcement learning model that predicted logRTs for each trial. This modeling approach extends the standard analysis by allowing us to track individual representations of reward and loss and their respective modulation of response vigor at each trial. We used a Rescorla-Wagner-like model (1972), where expected values of reward (Q^+) and loss (Q^-) were updated according to reward (δ^+) and loss (δ^-) prediction errors weighted by the learning rate (α). Then, we defined several alternative response models that described the mapping of the variables derived from the learning model onto the logRT responses (Figure 2.2, see Methods for full details). Model comparison showed that among all five tested response models, the model (M3) including cue salience (i.e. $|Q|$, merged $|Q^+|$ and $|Q^-|$) and novelty (i.e. $|\delta|$) terms (M1: *XP* = 0%, *PP* = 1.5%; M2: *XP* = 0%, *PP* = 1.5%; M3: *XP* = 100%; *PP* = 90.0%; M4: *XP* = 0%, *PP* = 1.5%; M5: *XP* = 0%, *PP* = 5.5%; XP, exceedance probability; PP, posterior probability) outperformed all other models (Figure 2.2, Table 2.2). This means that in our task (1) average reward and loss rates and (2) (signed) prediction error signaling did not contribute to explaining the response data. Based on this result, we used the trial-by-trial predictions of the best model in the subsequent fMRI analysis.

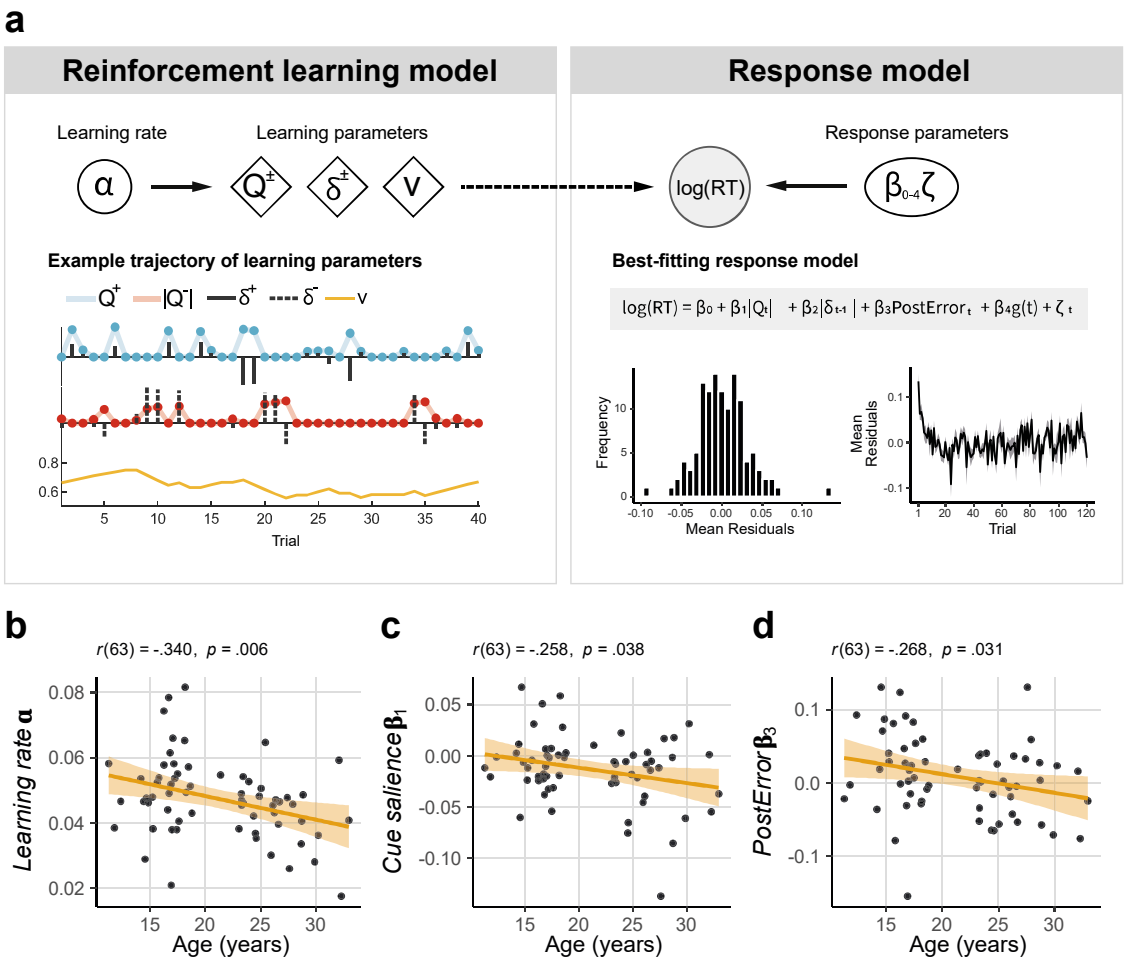

Figure 2.2. (a) A computational learning model was employed to estimate latent variables of interest (left panel). This resulted in trial-by-trial predictions for expected values and prediction errors for reward (Q^+, δ^+) and loss (Q^-, δ^-) trials. We show example trajectories of the estimated variables for a single subject; for clarity, only the first 40 trials are depicted. The learning parameters of the reinforcement learning model were entered in the response model to predict logRT for an individual trial (right panel). Inspection of the model residuals suggested that the model captured the observed data well. Shaded area indicates the SEM. (b-d) Trial-by-trial analysis of response time revealed moderators of vigour in our cohort. Learning rate decreased across age (b), i.e. adolescents changed their predictions about expected outcomes faster. Moreover, we found an age-related increase of response vigour in trials with higher cue salience (c) and in post-error trials (d). *r*, Pearson correlation coefficient.

To check the model fit of our behavioral model, we performed a linear mixed model analysis for simulated logRTs of the winning model. For this, we excluded two participants, for which the model fitting procedure did not converge. This analysis revealed comparable effects to the behavioral effects observed in the data. As in the raw data analysis, we found a significant main effect of condition, $F(4,7727) = 5.39$, $p = 0.0002$, an age-by-condition interaction, $F(4, 7727) = 8.59$, $p < 10^{-6}$, but no main effect of age, $F(1, 63.1) = 0.26$, $p = 0.61$. The model captured the differences between the conditions very well and reproduced effects found in the raw data analysis, namely that in the high loss condition participants increase vigor compared to neutral, low loss and low reward condition ($p <$

0.0001) and that they respond faster with high rewards at stake compared to neutral, low loss and low reward condition ($p < 0.0001$). Simulated mean logRTs did also significantly correlate with age in the high reward condition ($r(63) = -0.300$, $p = 0.03$).

2.3.2 Response model parameters are related to age

The parameters of the winning behavioral model are summarized in Table 2.2. There was no evidence, that quadratic or inverse-age models fitted the individual parameters better than the linear model ($\Delta BIC < 6.2$ for all model comparisons, Table 2.2) and therefore they were not investigated further. To see, if there is a relationship between the model parameters and (1) the age of participants or (2) the subjective liking and arousal ratings, we performed correlation analyses with the posterior mean of the parameters of the winning model. First, we found that the parameters for learning rate (α), cue salience (β_1) and $PostError$ (β_3) were significantly negatively correlated with age, i.e. adults were more flexible in their response vigor (speeding up for high expected values) and less prone to post error slowing (Figure 2.2; see Figure S2.1 for all parameter correlations). We did not find any correlation between behavioral parameters and post-scan outcome liking ratings (all $p > .05$). The response model parameters were only moderately correlated across age with all absolute $r < 0.562$ (Figure S2.2). Of the parameters showing significant correlation with age (α, β_1, β_3), the correlation of the learning rate α and cue salience β_1 was significant ($r(63) = 0.302$, $p = 0.01$).

Table 2.2. Winning model (M3) results and model comparison of age association for response model parameters

Model parameter	**Mean (SD)**	**BIC_{linear}**	**$BIC_{inverse}$**	**$BIC_{quadratic}$**	**$\Delta BIC_{(best-linear)}$**
α	0.047 (0.012)	-583.22	-580.32	-585.73	2.51
β_0	-1.365 (0.144)	-253.17	-253.20	-254.56	1.39
β_1	-0.013 (0.033)	-449.57	-448.64	-449.62	0.05
β_2	0.002 (0.026)	-478.50	-477.93	-478.91	0.41
β_3	0.009 (0.054)	-385.55	-385.32	-385.56	0.01
β_4	-0.019 (0.071)	-345.04	-344.36	-351.30	6.26

2.3.3 Incentive valuation remains constant across development

Additionally, we performed an analysis of post-scan ratings, to assess whether any age-related behavioral differences are related to different incentive valuation. The analysis of outcome liking ratings revealed a significant main effect of condition, $F(4, 225.55) = 24.51$, $p < 10^{-15}$, the main effect of age, $F(4, 54.19) = 0.24$, $p = 0.62$, and the age-by-condition interaction, $F(4, 225.02) = 0.37$, $p = 0.83$, were not significant. Liking scores differed across all experimental conditions significantly ($p < 0.001$), increasing from high loss to high reward outcome. Arousal ratings showed no significant main effect nor interaction (main effect condition, $F(4, 232) = 0.95$, $p = 0.435$, main effect age, $F(4, 58) = 2.889$, $p = 0.094$, age-by-condition interaction, $F(4, 232) = 0.629$, $p = 0.643$). This suggests that there was no age-related difference of motivational value immanent in the monetary outcomes.

2.3.4 Cue salience representation in incentive networks changes with age

Using fMRI, the first key question we sought to answer was whether representations of expected value and prediction error vary across age. We carried out parametric whole-brain analyses using the computed signals from the behavioral analysis as (nonorthogonalized) predictors for the BOLD signal to examine how they modulate brain activity during the task. The resulting first-level maps were entered into separate multiple regression analyses to determine the effects of age, sex and age × sex on neural signatures of expected value and prediction error processing (Methods).

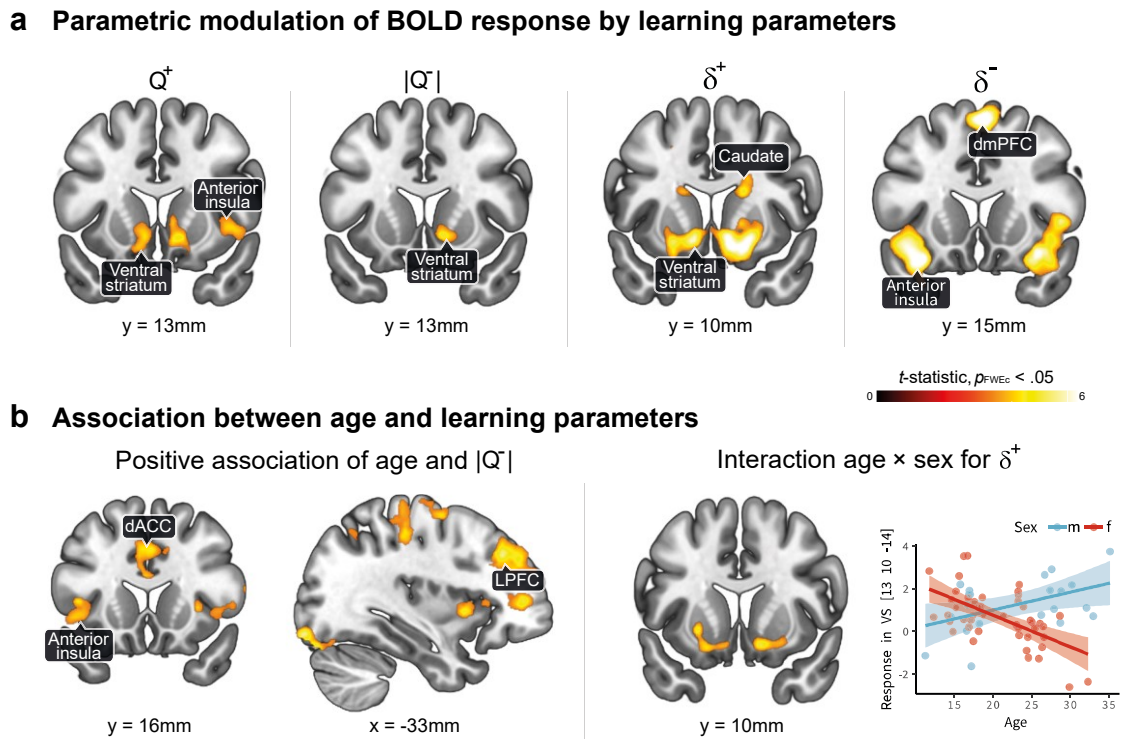

Figure 2.3. Summary of the results of the second-level multiple regression analyses. (a) We present the average effect of the parametric regressors derived from the reinforcement learning model. Responses in the ventral striatum correlated with the expected values Q^+ and $|Q^-|$. Prediction error signals correlated with responses in the ventral striatum and caudate (δ^+), and the anterior insula and the dorsomedial PFC (δ^-). (b) In addition, we show the statistical map of the brain responses to |Q-| that was moderated by age (left panel). Reward prediction error signals showed a significant age-by-sex interaction (right panel), with all clusters showing an age-dependent decrease in females and an increase in males. Details of the results can be found in Table S2.2. N = 67, p_{CDT} < .001, p_{FWEc} < .05.

The average effect of the Q^+ signal was significant in the bilateral VS, the left ventrolateral prefrontal cortex, anterior cingulate, bilateral angular gyri, the right insula, middle temporal gyrus, the thalamus and the cerebellum (Figure 2.3, Table S2.2). These effects were not modulated by age or sex. On the other hand, an average effect of $|Q^-|$ was observed in the right ventral striatum, thalamus, anterior cingulate, supplementary motor area, postcentral gyrus, lingual gyrus and fusiform gyrus (Figure 2.3, Table S2.2). Age had a significant positive effect on $|Q^-|$ activity in a cluster comprising dorsal anterior

cingulate and supplementary motor cortex, as well as the bilateral prefrontal cortex, the bilateral insulae, the supramarginal gyrus and the occipital fusiform gyrus. No significant effect of sex was observed.

2.3.5 Prediction error signaling depends on age-by-sex interaction

Representations of δ^+ were detected in the bilateral ventral striatum, the bilateral caudate nuclei, the ventromedial prefrontal cortex, the left dorsolateral prefrontal cortex, supplementary motor area, a cluster spanning dorsal hippocampus and the lateral thalamus, and the occipital cortex (Figure 2.3, Table S2.2). We found that δ^+ activity was positively correlated with age in the bilateral fusiform gyrus, but, contrary to our hypothesis, no negative correlation was found in the VS. However, we observed a significant age-by-sex interaction in the right ventral striatum and the superior temporal gyrus. In particular, older females exhibited reduced activity related to δ^+, while in males the activity increased. The average effect of δ^- was located in the bilateral ventral striatum peaking in the putamen, bilateral caudate nucleus, anterior cingulate, bilateral posterior orbital gyri, bilateral anterior insula, thalamus, pre-/postcentral gyri. In addition, a main effect of sex was found in the supramarginal gyrus (Figure 2.3, Table S2.2).

2.3.6 Fine-tuning of corticostriatal connectivity from adolescence to adulthood

Results from behavioral and the whole-brain analyses indicated, that during reward and loss anticipation there is a significant effect of age on (1) response vigor to salient cues (but not prediction error signaling) and (2) activity in core regions of the incentive processing circuitry during the anticipation phase. Given this association of age with response vigor and neural responses to cue salience, we evaluated how age differences of processing reinforcement learning signals manifested in the incentive processing network during the anticipation of incentives. For this, we performed an analysis of effective connectivity (dynamic causal modeling, DCM; see Methods) that determined the model that fitted the neural dynamics best.

We estimated each first-level DCM and analyzed the (1) group average and (2) the effect of age on each connection with a second-level Parametric Empirical Bayes (PEB) model (see Methods). Then, we examined the effect on the average connectivity between regions and the self-inhibition parameters of the DCM. In the DCM framework, self-inhibition parameters reflect a region's sensitivity to inputs for a given task context. The averaged connectivity strength of each connection is presented in Table 2.3. Note, that some connections have been removed in the Bayesian model reduction procedure (Friston et al., 2016; Zeidman et al., 2019), as they did not contribute to the model evidence. Across all participants, we found that the input region thalamus has excitatory influence on all other regions in our modelled network. The VS received input from the thalamus, the insula and the lateral prefrontal cortex. In addition, we found inhibitory connectivity from the striatal region to the thalamus. Connections originating in the insula showed negative connectivity to the thalamus, the VS, and the lateral prefrontal cortex (LPFC).

Dorsal anterior cingulate connectivity was targeting thalamus, the LPFC and the insula. The LPFC exhibited negative efferent connectivity to the thalamus, the dorsal anterior cingulate cortex (dACC) and the insula, and positive efferent connectivity to the VS. In the VS and the dACC, we found significant modulatory effects of Q_t^+ and $|Q_t^-|$. In the dACC, the self-inhibition correlated with Q_t^+ positively and with $|Q_t^-|$ negatively (Figure 2.4).

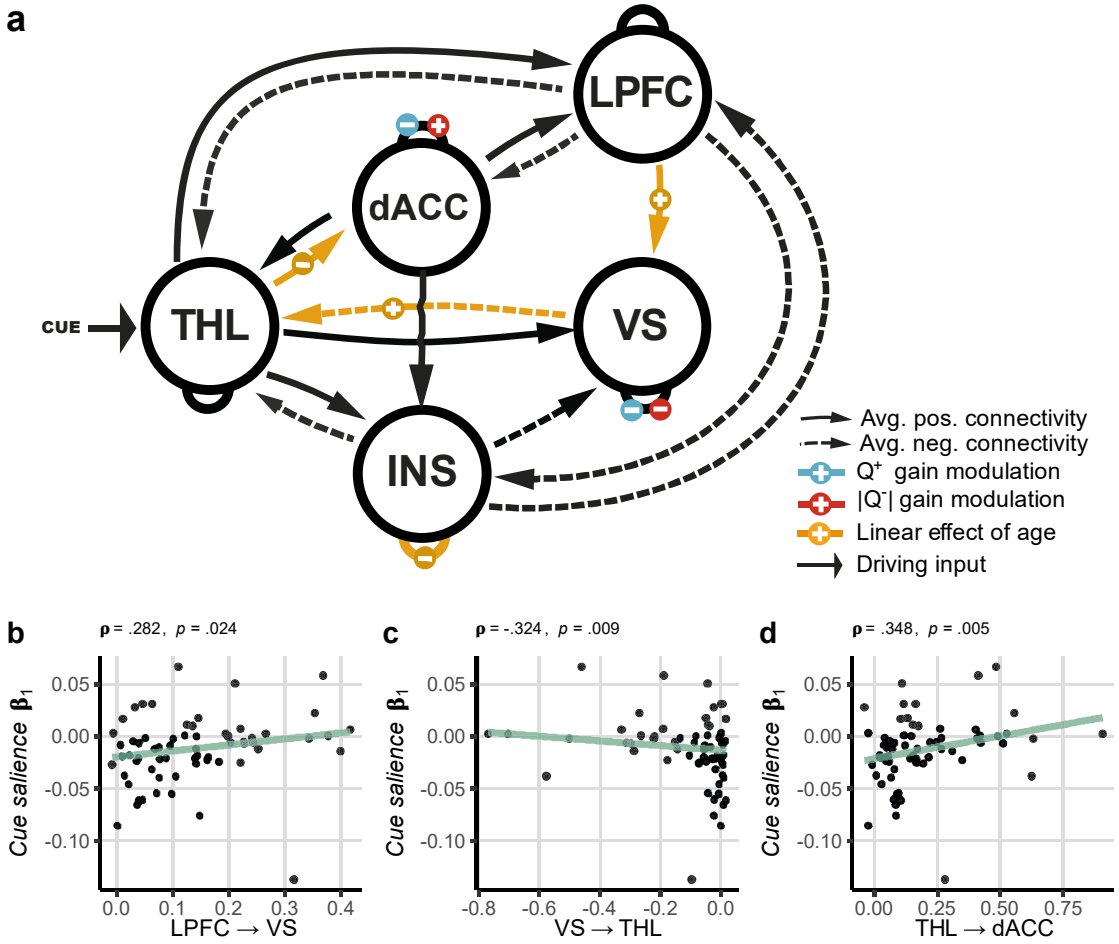

Figure 2.4. (a) The average group effective connectivity during reward and loss anticipation. The arrows reflect the posterior estimates of the second-level PEB model after Bayesian model reduction (Table 2.3). Self-connections are depicted as half-circle on each region. Solid lines indicate positive effective connectivity whereas dashed lines represent negative effective connectivity. Effective connectivity parameters between (b) LPFC → VS, (c) VS → THL, and (d) THL → dACC from our averaged DCM were significantly correlated with model parameter β_1. i.e. the logRT model component related to the cue salience. Green lines depict the fitted median regression line; Spearman's ρ and p-values are given for each correlation. Abbreviations can be found in Table 2.3.

Table 2.3. Average connectivity during anticipation phase obtained by Bayesian model averaging of PEB model parameters

Connection type	Commonalities	PP Commonalities	Age	PP Age
Endogenous parameters				
LPFC → VS	0.127	1	0.010	1
LPFC → Insula	-0.230	1	-	-
LPFC → THL	-0.293	1	-	-
LPFC → dACC	-0.261	1	0.008	0.73
VS → THL	-0.838	1	0.035	1
Insula → LPFC	-0.231	1	-0.006	0.70
Insula → VS	-0.113	1	-	-
Insula → THL	-0.142	1	0.011	0.79
THL → LPFC	0.323	1	-	-
THL → VS	0.218	1	-	-
THL → Insula	0.396	1	-	-
THL → dACC	0.846	1	-0.014	1
dACC → LPFC	0.147	1	-	-
dACC → VS	0.017	0.5	-	-
dACC → Insula	0.082	1	-	-
dACC → THL	0.552	1	-0.007	0.66
Self-inhibition parameters				
LPFC → LPFC	-0.432	1	-	-
Insula → Insula	-0.381	1	0.018	1
dACC → dACC	-0.072	0.79	-	-
VS → VS	-0.166	1	-0.012	0.75
THL → THL	-0.283	1	0.015	0.82
Modulatory parameters				
Insula → Insula, Q_t^+	-	0	0	0
Insula → Insula, $\lvert Q_t^- \rvert$	0.156	0.70	0	0
dACC → dACC, Q_t^+	0.919	1	0	0
dACC → dACC, $\lvert Q_t^- \rvert$	-0.280	1	0	0
VS → VS, Q_t^+	1.652	1	0	0
VS → VS, $\lvert Q_t^- \rvert$	1.154	1	0	0

Between-region connections are in units of Hz. Self-connections, where the source and target are the same, are the log of scaling parameters that multiply up or down the default value −0.5Hz. *n* = 66. dACC, dorsal anterior cingulate cortex; INS, insula; LPFC, lateral prefrontal cortex; THL, thalamus; VS, ventral striatum; PP. posterior probability.

We found a significant increase of connectivity with age from the LPFC to the VS and in the self-inhibition of the insula. The negative connectivity from the VS to the thalamus (THL) became less inhibitory with age. Decrease of effective connectivity with age was found in the connectivity from the thalamus to the dACC. These results indicated that the cortico-striatal-thalamic circuitry is fine-tuned with age (Figure 2.4). Using leave-one-out cross-validation (LOOCV), we assessed whether these effects were predictive for the age of an independent subject, i.e. we fitted the PEB model to all but one subject to obtain the model parameters and use the effective connectivity of the left out subject to predict their age. That is, we assessed if we could predict the age of an independent subject given only its intrinsic connectivity. As the correlation between the estimated and the actual

age was significant, $r(64) = 0.26$, $p = 0.02$, we can expect that if we included new subjects, they would exhibit the same association based on the model parameters of this cohort.

In a final step, we assessed whether the cue salience behavioral parameter β_1 was related to the connectivity parameters revealed in the PEB model (Figure 2.4). We found that β_1 was significantly correlated with the posterior mean of the connections LPFC → VS, Spearman's $\rho = 0.282$, $p = 0.02$, VS → THL, $\rho = -.324$, $p = 0.009$, and THL → dACC, $\rho = .348$, $p = 0.005$. The association between VS → THL and β_1 remained significant after removing five left sided outliers of the DCM parameter determined by Rosner's test ($\rho = -.301$, $p = 0.021$). No significant association was found for the self-inhibition of the insula and β_1, $\rho = -.199$, $p = 0.12$.

2.4 Discussion

The ability to adjust behavior is pivotal when facing ever-changing environmental demands. Here, we demonstrate that during an instrumental task the ability to specifically increase response vigor for high incentives improves from early adolescence to early adulthood and is paralleled by developmental changes of information flow within cortico-striatal-thalamic connectivity. These results suggest that rather than a simple cortical-subcortical imbalance (Casey et al., 2011; Steinberg, 2010), the network that supports incentive-guided action undergoes a fine-tuning of effective connectivity across adolescence into young adulthood. By applying a trial-by-trial reinforcement model in conjunction with dynamic causal modeling, we were able to extend previous studies of the neurobiology of instrumental vigor across development (Cho et al., 2013; Lamm et al., 2014). We found evidence that differences in adaptive responses can be linked to age-related changes in cortico-striatal-thalamic effective connectivity. These findings support that a transient maturation of cortico-striatal-thalamic circuits serve the development of efficient motivated behavior.

On the behavioral level, we demonstrated that increase in vigor is dependent on the cue salience of a given trial rather than on the valence of a cue (i.e. reward or loss) or the average reward or loss rate. This is in line with the results of previous work that similarly found faster responses during incentivized trials (Pfabigan et al., 2014; Wrase et al., 2007; Wu et al., 2014). Although the average reward rate has been linked to increased instrumental vigor (Niv et al., 2007), the average rates for reward and loss are low in a paradigm like the Monetary Incentive Delay Task (Beierholm et al., 2013).

Adult participants were able to speed up during high expected values and this ability increased linearly from late adolescence to young adulthood. This behavior occurred together with a lower learning rate, suggesting a more stable representation of values and a stronger behavioral discrimination between low and high incentives over the course of the task in adults, whereas younger participants showed lower discrimination in the beginning and acquired it over time. Selective improvement of performance for high incentives has been found previously in young adults compared to children and adolescents

(Hämmerer, Li, Müller, & Lindenberger, 2011; Störmer, Eppinger, & Li, 2014) and is consistent with theories of cognitive control that assert that action execution of adults (e.g. in response to a go-signal) can be selectively modulated by incentives (Botvinick & Braver, 2015). Moreover, out data show that with increasing age the effect of prediction errors on the subsequent expected values decreases, resulting in lower learning rates. Together, this indicates that the behavioral adaptation is less and less influenced by feedback that is not relevant to perform well in the task (due to more stable behavior) across development (Van Den Bos et al., 2012). Importantly, as our sample had an age range from as young as 11 years up to 35 years, we determined that this effect does not reach ceiling levels after adolescence but continues to increase into adulthood. It is possible that this age-dependent incentivized vigorous behavior is related to differences in subjective valuation of monetary values between adolescents and adults. However, in our study, increased vigor during high cue salience was not related to differences in valuations of monetary outcome. This suggests that age-related differences cannot be attributed to valuation per se, but likely originate from the cognitive demands of the task.

In addition to the age-dependent modulation of the salience effect on instrumental vigor, we observed that post-error slowing was decreasing with age. Our behavioral model comparison suggested that logRT is more related to misses in general, rather than to distinct prediction error signals incorporating magnitude and valence (i.e. reward omission or monetary loss). While younger participants showed increased logRT after missing a target, this effect was decreasing into adulthood. The developmental trajectory of post-error adaptation is not fully understood and currently under debate. Different conflicting accounts have been reported, with studies employing varying paradigms showing increase (Overbye et al., 2019), but also decrease (J. B. Taylor, Visser, Fueggle, Bellgrove, & Fox, 2018) in post-error slowing and post-error improvement (Van Duijvenvoorde, Zanolie, Rombouts, Raijmakers, & Crone, 2008) across adolescent development. According to Wessel (2018) post-error slowing reflects an adaptive orienting response, where a cascade of cognitive processes is initiated to eventually improve performance. In that sense, immature cognitive control of attentional orientation might yield slower post-error performance in younger individuals (Geier & Luna, 2009).

Model-based analyses of incentive anticipation demonstrated consistent activation in the insula and the dACC, the principle nodes of the salience network, as well as the lateral prefrontal cortex, the striatum and the thalamus. Importantly, we observed a significant modulation of expected value signals for both, reward and loss, in the ventral striatum, the thalamus, the insula and the dACC (implicitly compared to non-incentivized trials). Former work reported that an expected value signal in the striatum can boost instrumental vigor in reward approach and loss avoidance behavior, i.e. regardless of valence (Dayan, 2012; Rigoli et al., 2016). This is consistent with previous studies that assessed reward processing in adolescents and adults (Cao et al., 2019; Oldham et al., 2018). Crucially, we did observe a difference across age for the encoding of the expected value dur-

ing loss trials, but not to reward. We identified clusters in the anterior insula and the dorsal ACC where the $|Q_t^-|$ signal positively correlated with age. In adults, aversive processing has been repeatedly shown to be associated with activation of the dACC (Jensen et al., 2003; Pohlack, Nees, Ruttorf, Schad, & Flor, 2012). Research of loss or aversion processing from early adolescence to adulthood has been sparse, nevertheless, the few existing studies reported decreased activity in the dorsal caudate (Cho et al., 2013; Lamm et al., 2014), the ACC (Bjork et al., 2010), and the insula (Galván & McGlennen, 2013) in adolescents compared to adults. Our findings of decreased activity in caudate, insula and dACC for $|Q_t^-|$ corroborate these earlier findings and extend it by showing age-related increase of activity in lateral prefrontal cortex. This correlation was not explained by differences of valuation nor arousal ratings, thus suggesting that the observed differences are not rooted in age-dependent salience attribution. This could indicate that the integration of motivational and salient events for more proactive control in loss avoidance continues to mature into young adulthood.

Contrary to our hypothesis, we could not replicate earlier findings of general heightened reward sensitivity of the NAcc during adolescence compared to adulthood (Barkley-Levenson & Galván, 2014; Braams, Peters, Peper, Güroğlu, & Crone, 2014; Somerville et al., 2011). Our results point to a more specific age-by-sex interaction effect of reward prediction error signals in the nucleus accumbens. In particular, our results suggest that in females, nucleus accumbens activity related to reward prediction errors decreases, activity in males increases across age. Previous studies have not only shown the influence of gonadal hormones on structural brain changes during puberty (Peper, Pol, Crone, & Van Honk, 2011) but also on accumbens activity during reward processing (Forbes et al., 2010; Ladouceur et al., 2019). Because no hormonal levels were measured in the present study and given the exploratory nature of this finding, this link remains suggestive, should be interpreted with caution and addressed in future studies. Nonetheless, behavioral model comparison substantiated that prediction error signaling did not show to affect response vigor significantly across age.

Based on the findings of behavioral and whole-brain analyses, we conducted an effective connectivity analysis to assess age-related changes in the incentive processing network. Our results suggested that response vigor is closely related to the expected value of reward and loss incentives and changes across adolescence along with associated brain activity. Thus, we assessed the maturation of the functional architecture of the network comprised of regions (1) encoding expected values and (2) serving cognitive control of motivational processes (Botvinick & Braver, 2015).

First, in line with our hypothesis, we observed an increase of connectivity between the LPFC and the VS. The LPFC is well known for supporting motivated behavior by storing and updating goal-relevant information and executing regulative control (Botvinick & Braver, 2015). Nevertheless, it remains unclear how a protracted LPFC maturation (Gogtay et al., 2004) affects the orchestration of incentive-based behavioral adaptations in

concert with other, differentially developing regions. Across adolescence, cognitive control emerges transiently and is associated with task performance differences between youth and adults (Crone & Dahl, 2012). Prior work has shown elevated corticosubcortical connectivity during processing of salient stimuli predicting reward in adults (Ballard et al., 2011; Kinnison, Padmala, Choi, & Pessoa, 2012). Recently, first evidence has emerged that this adult ability to selectively exert cognitive control and improve performance to obtain high rewards is associated with the development of corticostriatal connectivity during adolescence (C. Insel et al., 2017). Hence, the observed increase of information flow from the LPFC to the VS could reflect a strengthened control signal that is necessary to retrieve cognitive resources to improve performance.

Secondly, we identified a developmental change in effective connectivity from the VS to the thalamus. This functional pathway has already been identified in previous DCM studies (Cho et al., 2013; Li et al., 2015) and its engagement seems to be particularly present during adolescence compared to adulthood (Cho et al., 2013). Our results confirm a developmental decrease of coupling from VS to the thalamus across adolescence and adulthood. The striatum projects to the ventral pallidum, which in turn projects mainly inhibitory GABAergic to the thalamus. Therefore, the VS is in a suitable position to regulate the disinhibition of the thalamus (Haber & Knutson, 2010). The striatothalamic pathway has been implicated in successful reinforcement learning, in particular learning the relationship between an action and their consequences (Dudman & Krakauer, 2016; Pessiglione et al., 2007). The thalamus shares bidirectional connections with a wide range of cortical regions (Haber & Knutson, 2010) and evidence from animal studies suggest, that thalamic lesions severely affect the ability to use rewards for goal-directed behavior (Chakraborty, Kolling, Walton, & Mitchell, 2016; Leung & Balleine, 2015). Moreover, pharmacogenetical models of thalamic hypofunction during Pavlonian conditioning are associated with failures of reward-related behavioral modulations (Parnaudeau et al., 2013). Given the importance of integrity of this pathway in reinforcement learning, this underlines that the maturation of striatothalamic connectivity supports the facilitation of salience attribution to an incentivized stimulus and thereby promoting signals indicating a need for cognitive control. According to recent proposals, the dACC integrates these signals for monitoring demand and the allocation of cognitive control to maximize outcome (Cavanagh & Frank, 2014). This computation in the dACC is thought to result in a specification of which control mechanism to execute in order to optimize behavior adaptively, which can be transmitted to other regions (e.g. the LPFC). Suggestive evidence from primate studies indicates that the dACC uses valence-specific representations of outcome uncertainty for this purpose (Monosov, 2017), and information of past outcomes is encoded in inhibitory interneurons (Kawai, Yamada, Sato, Takada, & Matsumoto, 2018; Sajad, Godlove, & Schall, 2019). In the DCM framework, a summary measure of the excitatory/inhibitory balance within a region is modeled by the self-connections (Zeidman et al., 2019). The DCM results demonstrated that the dACC did modulate its self-connections with regard to the cue valence that was processed. This supports

the evidence presented and further emphasizes the valence-encoding role of the dACC in humans.

Moreover, we found a decrease in effective connectivity from the thalamus to the dACC across adolescence. Although it is widely appreciated that excitatory thalamocortical connections critically contribute to reward-related behavior as motor planning and salience detection (Pergola et al., 2018), surprisingly little research on functional coupling during incentive processing has been performed in humans. Increased functional connectivity between thalamus and dACC has been associated with increased risk-taking behavior in adult smokers (Wei et al., 2016). In addition, different lines of evidence have shown significant remodeling of this circuit during adolescence. For instance, the levels of glutamate in the medial prefrontal cortex are elevated in adolescence and decrease across young adulthood (Marsman et al., 2013). Moreover, myelinic maturation within this circuitry has been associated with lower impulsivity (Ziegler et al., 2019). Animal research suggests that the role of the thalamus in adapting behavior as a function of incentives might be fundamentally dependent on inhibitory activity of thalamocortical neurons (Delevich, Tucciarone, Huang, & Li, 2015; Rikhye, Gilra, & Halassa, 2018). Taken together, adolescent hyperconnectivity between thalamus and dACC might reflect an immature mechanism of generating appropriate control signals for adjusting behavior. A decrease in connectivity across adolescence could therefore reflect a damping of the striatothalamic feedback to the cortex and a shift towards cortical control.

Lastly, we found that the self-inhibition of the insula during incentive anticipation increases with age. In parallel to the dACC above, this means that the input gain decreases across adolescence. The insula is a hub that shapes motivational states and attention based on the affective evaluation of sensory input and tags relevant stimuli for further processing (Gogolla, 2017). The adolescent disinhibition of the insula that might reflect a distorted weighting of ascending salience-attributed sensory signals. Immature salience attribution in this brain hub orchestrating cognitive control might have contributed to the failure of flexible behavior. Further, our results indicate a functional coupling from the valence-sensitive dACC to the insula, both being highly implicated in processing salience (Uddin, 2015). Again, a decrease in sensitivity to inputs could therefore reflect a shift from weighing bottom-up salience signals towards a mature top-down cognitive control to achieve an adjustment of behavior to salient stimuli.

These findings support the idea that appropriate attentional filtering is important to adapt ones behavior to incentivized stimuli (Parro, Dixon, & Christoff, 2018). Different brain systems like the salience network or prefrontal-striatal network work in concert to support appropriate filtering and adjustment of behavior. Maturation of the cortico-striato-thalamic system should eventually facilitate cognitive processes or motor responses via exertion of cognitive control of the LPFC. In line with this idea, a recent rodent study has shown that the prefrontal cortex is able to modulate sensory processing in the thalamus via the basal ganglia for attentional filtering of sensory signals fostering goal-directed behavior (Nakajima, Schmitt, & Halassa, 2019). This finding demonstrates the

complexity of neural circuits involved in motivated behavior and underscores the necessity to study how developmental processes manifest with appropriate network models in humans.

To summarize, our study demonstrated that the ability to adapt response vigor towards salient cues in a trial-by-trial fashion improves from early adolescence to adulthood. Furthermore, we show how classic models of reinforcement learning in conjunction with biophysics of neuronal dynamics can reveal developmental aspects of the underlying functional architecture of behavior. We corroborate previous studies that found that performance of adults improves during incentivized tasks (Chiew & Braver, 2016; Locke & Braver, 2008; Pfabigan et al., 2014; Wrase et al., 2007; Wu et al., 2014) and show that, compared to adults, adolescents have difficulties to adapt behavior for high subjective value (C. Insel et al., 2017). Our computational fMRI approach allowed us to link the overt behavioral adaptations guided by latent processes to maturational changes in activity and functional coupling. With this, we provide evidence that progressive fine-tuning of the cortico-striatal-thalamic circuit, away from a subcortical reward-seeking system and towards more cognitive regulation, takes place to facilitate motivated action. Additionally, this approach revealed a functional sex difference in the development of striatal reward prediction error signaling. Although here, this did not affect task performance critically, it is highly likely that diverging sex-specific trajectories extending into adulthood have implications in the context of decision making and risk-taking. Thus, we believe this study could have important ramifications that pertain public health and the prevention of high-risk behavior. During this important stage of development adolescents form habits that can lead to problems in later life (e.g. obesity, diabetes, or smoking) or have acute effects (e.g. substance abuse or sexually transmitted disease; Kann et al., 2018; Kühn et al., 2019). Hitherto, however, the efficacy of incentive based education and prevention programs is not well established (Bright, Felix, Kuper, & Polack, 2018; Johnston, Liberato, & Thomas, 2012; Levitt, List, Neckermann, & Sadoff, 2016). Neuroscientifically informed policies that respect age and sex-specific neural and behavioral constraints across development might be able to improve intervention approaches (Whitten, 2013). The results obtained in this study indicate that interventions using incentives might not always be sufficient to efficaciously unfold the full motivational potential in adolescents and might also differentally engage girls and boys. Hence, future programs might benefit from adjustments to suit cognitive brain trajectories and being attuned to specific needs. To eventually optimize targeted intervention programs, it is important to further characterize motivational effects of incentives in different neurodevelopmental phases.

2.5 Methods

2.5.1 Participants

We recruited a group of 67 subjects (age: M = 21.4, SD = 5.9, range 11-35y, 46 females and 21 males, 62 right and 5 left handed). Inclusion criteria comprised age 8-45 years and signed informed consent. Parents gave signed informed consent for subjects younger

than 14 years old. Exclusion criteria comprised any MRI contraindication, pregnancy, a history of brain injury, a current psychiatric disorder, other major medical illnesses, and drug abuse. Three adolescent participants had a past diagnostic work-up for ADHD but they were currently symptom-free and were not taking any medication during the study. All participants were reimbursed for participation and informed about the opportunity to additionally win up to CHF 20 during the task. This study was approved by the ethics committee of the Kanton Zürich and was conducted in accordance with the Declaration of Helsinki.

2.5.2 Experimental design

In our study, we employed the Monetary Incentive Delay (MID, Figure 2.1) task to investigate motivational states and outcome processing (Knutson et al., 2000). This task allows to investigate incentive anticipation and the ensuing feedback processing, while minimizing possible cognitive confounds due to the simple decision processes (Oldham et al., 2018). Every trial started with a cue indicating the level of magnitude (CHF 1, CHF 4) and the valence (reward, loss-avoidance, null) for a button press on time ("hit"). Participants were instructed to use the index finger of their dominant hand to press a button on a two-button fiber-optic response pad (Current Design Inc., Philadelphia, PA) as soon as the go-signal target symbol, a star, appeared. In total, each cue was presented 24 times (i.e. 120 trials in total, mean stimulus onset asynchrony = 2675ms), in two separate MRI runs. We used an adaptive algorithm that adjusted the presentation times of the target to the response time of the participant to ensure a hit rate of ~66%. The cue symbols indicating valence (square, triangle, and circle) were counterbalanced across subjects. The level of magnitude was represented by using a full symbol for high magnitude (CHF 4) and an empty symbol for low magnitude (CHF 1). All participants had a short training session outside the scanner (~2 minutes) to become familiarized with the task and we ensured that cue-outcome contingencies were understood. The task was implemented in python (pygame, https://www.pygame.org) and presented using video goggles (VisuaStimDigital, Resonance Technology, Northridge, CA) with a resolution of 800x600px.

Subjective liking and arousal for rewards and losses were assessed after the MRI scan outside the scanner. Participants were presented with the amount of money they were able to win in each condition and they were asked to rate their (1) liking ("How much did you like this outcome?") and (2) arousal ("How excited were you by the outcome?") during the feedback phase of the respective outcome on a continuous scale using a slider between 0 (strongly dislike, not aroused) and 100 (strongly like, highly aroused).

2.5.3 MRI data acquisition and preprocessing

MRI recordings were conducted on an Achieva 3T scanner (Philips Medical Systems, Best, the Netherlands) using a 32-channel head coil array. Functional images were acquired with a multi-slice echo-planar images (EPI) sequence [335 volumes per session, TR = 1.6s, TE = 35ms, 15° tilted downwards of AC-PC, 50 slices, voxel size = 2.4 × 2.4 × 2.2mm^3, matrix size = 76 × 78px, flip angle = 75°, gap = 0.35mm, SENSE-factor = 2, MB-factor =

2]. After the two task sessions, a T1-weighted anatomical image was acquired for each subject with a 3D magnetization-prepared rapid gradient-echo sequence (MP-RAGE) [time between two inversion pulses = 2484ms, inversion time TI = 900ms, inter-echo delay = 6.7ms, aligned at AC-PC, flip angle = 9°, voxel size = 1.05 × 1.05 × 1.2mm^3, field of view = 270 × 253mm^2, 170 sagittal slices]. Slice-time corrected functional data was realigned and coregistered to the T1-weighted image. The deformation fields derived from the segmentation of the T1 image were used for normalization to the Montreal Neurological Institute (MNI)-152 template space. Finally, we applied spatial smoothing with a 6mm full-width-half-maximum (FWHM) kernel to the functional data. All steps were conducted in SPM12 (7487). Motion artefacts were addressed by calculating the framewise displacement (FD) of each subject across the task (Power, Barnes, Snyder, Schlaggar, & Petersen, 2012). No subject exceeded a mean FD of 0.5mm (M = 0.17, SD = 0.08mm), however, single volumes that exceeded a FD greater than 1mm were censored in the ensuing analyses by including an additional binary regressor (% volumes censored per subject M = 0.66, SD = 1.58%).

2.5.4 Behavioral analysis of raw data

We performed the raw data analysis on log transformed RTs to achieve a more normally distributed data set. We conducted a linear mixed model analysis with random intercept and the five task conditions (high reward, low reward, neutral, low loss, and high loss) and age as fixed factors. Significant main or interaction effects were subsequently analyzed using post hoc Tukey tests. Response data that deviated more than three standard deviations from the respective mean per condition and per subject were excluded from the analysis (1.90%).

Similarly, we analyzed the subjective liking ratings of the monetary value in the feedback using a linear mixed model with condition and age as fixed factor, and participants as random factor. Extreme ratings were excluded (+/- 3 SDs, four ratings in total). We excluded eight subjects from this behavioral analysis for not completing the subjective ratings due to time constraints.

The behavioral analysis was conducted in *R* (version 3.5.3, The R Foundation for Statistical Computing, http://www.r-project.org/index.html) using the package *lme4*. The significance level for all statistical tests of the behavioral analyses was $p < 0.05$, two-tailed. In mixed models, we used the Satterthwaite approximation for the degrees of freedom.

2.5.5 Computational learning model

We adapted the Rescorla-Wagner model (1972) to compute different signals of interest across trials. After cue presentation, it has been observed that brain activity in dopaminergic brain regions correlate with an expected value Q_t (O'Doherty et al., 2003). During receipt or omission of reward or loss respectively, prediction errors are thought to be teaching signals that enable the adaptation of future behavior to optimize outcome and continue to be computed even when behavior is already highly trained (Bayer & Glimcher,

2005). To disentangle effects of loss and reward, we defined two different signals, based on the current cue. The probability of achieving a miss was $P(Hit) = 1 - P(Miss) \approx 66\%$.

$$Q_t^+ = \begin{cases} C_t \cdot v_t & C_t > 0 \\ 0 & C_t \leq 0 \end{cases} \tag{2.1}$$

$$Q_t^- = \begin{cases} C_t \cdot (1 - v_t) & C_t < 0 \\ 0 & C_t \geq 0 \end{cases} \tag{2.2}$$

Here, Q_t^+ represents an expected reward, whereas Q_t^- represents an expected loss, dependent on the subjective probability for a reward v_t and loss $(1 - v_t)$ and the possible outcome C_t.

Depending on the actual outcome in the trial, reward (δ_t^+) and loss (δ_t^-) prediction error signals were calculated as

$$\delta_t^+ = \begin{cases} R_t - Q_t^+ & C_t > 0 \\ 0 & C_t \leq 0 \end{cases} \tag{2.3}$$

$$\delta_t^- = \begin{cases} Q_t^- - R_t & C_t < 0 \\ 0 & C_t \geq 0 \end{cases} \tag{2.4}$$

The update rule for the hit probability in the subsequent trial was given by

$$v_{t+1} = \begin{cases} v_t + \alpha \cdot \dfrac{\delta_t}{|C_t|} & C_t \neq 0 \\ v_t & C_t = 0 \end{cases} \tag{2.5}$$

where α was a free parameter and corresponded to the learning rate, constrained to the boundaries 0 and 1, and δ_t represented the signed prediction error (i.e. it reflects merged δ_t^+ and $-\delta_t^-$). In addition, average reward and loss at each trial was defined as:

$$\bar{R}_t = \begin{cases} \bar{R}_{t-1} + \alpha \cdot (R_{t-1} - \bar{R}_{t-1}) & C_t > 0 \\ \bar{R}_{t-1} & C_t \leq 0 \end{cases} \tag{2.6}$$

$$\bar{L}_t = \begin{cases} \bar{L}_{t-1} + \alpha \cdot (|L_{t-1}| - \bar{L}_{t-1}) & C_t < 0 \\ \bar{L}_{t-1} & C_t \geq 0 \end{cases} \tag{2.7}$$

where R_t and L_t represent the actual and $\bar{R}_t$ and $\bar{L}_t$ the average reward or loss at trial t. The trajectories resulting from the learning model were then used to generate trial-by-trial predictions of logRTs in the response model (Figure 2.2).

2.5.6 Response model

We compared five different response models that could explain the observed response time data. All models assume that the logRT is a linear combination of individual task-related parameters and a constant term. Given the results from the raw data analysis (main effect of condition) and the results from previous work (Dudman & Krakauer, 2016), we strongly expected the values Q_t^+ and Q_t^- to modulate response vigour in our subjects. We therefore included these terms in all response models. In addition, we included a linear function g to model any drift across task duration. We created different

response models and used Bayesian model comparison for the formal assessment of additional factors affecting response vigor. First, we defined a model, where the average reward and loss rates were additional predictors for the logRT (M1, equation 2.8) as previous work has shown that average reward rate is related to tonic dopamine and could boost vigor across task trials (Beierholm et al., 2013; Niv et al., 2007).

Another possibility is that dopaminergic release by reward prediction errors affects subsequent performance (Bestmann et al., 2014). In addition, loss prediction errors might be signaled differently and could modulate vigor on the next trial through a different mechanism (Lawson et al., 2014). Therefore, we created a second response model, where reward and loss expected values and prediction errors could influence vigor independently (M2, equation 2.9). Other research has indicated that cue salience (i.e. unsigned expected value) and novelty (i.e. unsigned prediction error) can influence dopaminergic activity (Bunzeck & Düzel, 2006). We therefore created three additional response models where cue salience and novelty (M3, equation 2.10), valence-dependent expected values and novelty (M4, equation 2.11), or cue salience and reward and loss prediction errors (M5, equation 2.12) served as predictor for logRT. $PostError$ denotes a binary vector of trials after an error, Rep denotes a binary vector of successive presentation of equal cues, and ζ denotes Gaussian noise.

Response model M1:

$$log(RT)^t = \beta_0 + \beta_1 \cdot |Q_t^-| + \beta_2 \cdot Q_t^+ + \beta_3 \cdot \bar{R}_{t-1} + \beta_4 \cdot \bar{L}_{t-1} + \\ + \beta_5 \cdot PostError_t + \beta_6 \cdot Rep_t + \beta_7 \cdot g(t) + \zeta_t \quad (2.8)$$

Response model M2:

$$\log(RT)^t = \beta_0 + \beta_1 \cdot |Q_t^-| + \beta_2 \cdot Q_t^+ + \beta_3 \cdot \delta_{t-1}^- + \beta_4 \cdot \delta_{t-1}^+ + \beta_5 \cdot g(t) + \zeta_t \quad (2.9)$$

Response model M3:

$$\log(RT)^t = \beta_0 + \beta_1 \cdot |Q_t| + \beta_2 \cdot |\delta_{t-1}| + \beta_3 \cdot PostError_t + \beta_4 \cdot g(t) + \zeta_t \quad (2.10)$$

Response model M4:

$$\log(RT)^t = \beta_0 + \beta_1 \cdot |Q_t^-| + \beta_2 \cdot Q_t^+ + \beta_3 \cdot |\delta_{t-1}| + \beta_4 \cdot PostError_t + \beta_5 \cdot g(t) + \zeta_t \quad (2.11)$$

Response model M5:

$$\log(RT)^t = \beta_0 + \beta_1 \cdot |Q_t| + \beta_2 \cdot \delta_{t-1}^- + \beta_3 \cdot \delta_{t-1}^+ + \beta_4 \cdot g(t) + \zeta_t \quad (2.12)$$

2.5.7 Behavioral model fitting and model comparison

The behavioral models were fitted to the data using the TNU Algorithms for Psychiatry-Advancing Science (TAPAS, http://www.translationalneuromodeling.org/tapas) HGF Toolbox 5.3, using a quasi-Newton optimization algorithm. Priors are summarized in Table S2.1. Trials without response were excluded for the model fitting procedure. For model comparison, we used Bayesian Model Selection (*spm_BMS.m*) to choose the best-fitting model by comparing the negative free energies, an approximation to the log-model evidence. Herein, we report the exceedance probability (XP) of each model, i.e. the

probability that one model explains the data better than the other models, and the posterior probability (PP) of each model.

Subsequently, we were interested in whether the parameters of the winning behavioral model relate to age. As developmental trajectories might show a nonlinear pattern, we first compared if the individual model parameters followed a linear, quadratic or inverse-age function. For fitting the function to the behavioral model parameters, we used the MATLAB function *fitnlm*.

2.5.8 Simulation analyses

We performed posterior predictive checks to assess the reliability of the behavioral model by mirroring the raw data analysis with simulated logRT data to see if we can replicate meaningful effects in our data. Based on the estimated individual parameters from the best-fitting model, we ran 1000 simulations per parameter set obtained for each subject and averaged the simulated trial-by-trial logRT using TAPAS.

2.5.9 Model-based fMRI - GLM analysis

The goal of the fMRI analysis was to identify reward and loss related signals during anticipation and outcome processing that covary with age. In the first-level analysis, we created a general linear model (GLM) for each participant. The cue onsets were convolved with the hemodynamic response function, and the Q_t^+ and the $|Q_t^-|$ values were added as parametric modulators, representing the expected outcomes based on previous experience. Secondly, the feedback onsets convolved with the HRF were added to the model with δ_t^+ and δ_t^- serving as parametric modulators. Note, that the neutral condition was the unmodulated case and thus the reference in both anticipation and feedback case. In addition, we added the temporal and dispersion derivatives of each regressor, and the six realignment parameters and a vector for scans with $> 1mm$ FD as nuisance regressor to the model. Finally, we applied a 1/128Hz cut-off high-pass filter to eliminate low frequency drifts.

In the random effects group analysis we conducted four multiple regression analyses, where individual contrast images for $Q_t^+/|Q_t^-|$ and δ_t^+/δ_t^- served as dependent variable. These second-level models included the group mean, age, sex and the interaction term age × sex as predictors. We used t-contrasts to test the individual effects for significance. Two participants with poor behavioral model fit were included using the prior expectation of the learning model parameters. We report results from the whole-brain analysis using cluster-level family-wise error correction ($p_{\mathrm{FWEc}} < 0.05$) with a cluster-defining threshold of ($p_{\mathrm{CDT}} < 0.001$). All fMRI analyses were conducted in SPM12 (7487).

2.5.10 Dynamic causal modeling

To assess, how these age-dependent changes emerged on a network level, we conducted a dynamic causal modeling (DCM) analysis. DCM has been demonstrated to be more ca-

pable to separate age-related vascular from neural changes compared to functional connectivity measures (Tsvetanov et al., 2016) rendering it a useful tool for studying the developing brain.

In DCM studies, normally a model space is specified, in which individual models represent specific hypotheses about the functional architecture of the brain. The models within the model space can then differ in either the presence or absence of an intrinsic connection or the contextual modulation of a connection. However, in our study, the goal was not to find the best model structure. Instead, our goal was to assess how connectivity strengths (1) are altered under different contextual manipulations (i.e. processing Q_t^+ or $|Q_t^-|$) and (2) are modulated by development, i.e. how they change across age. For this, we harnessed recent methodological improvements of DCM analysis in the framework of PEB to estimate connectivity parameters in the incentive processing circuitry. In the first-level analysis, we iteratively estimated the full model of each participant within an empirical Bayesian inversion scheme that uses the group average parameter estimates as priors for the estimations in the next iteration (Zeidman et al., 2019). After the inversion of the full model for each participant, we performed a second-level analysis using a PEB model to determine the group average and the age effect for each connectivity parameter, separately for intrinsic and modulatory connections. Based on the results from the GLM analyses, we created a PEB model that included the group mean and the mean-centered age. We applied Bayesian model reduction to perform an automatic search over reduced PEB models and iteratively removed model parameters that did not contribute to the evidence. Finally, we performed Bayesian model averaging of the best PEB models by averaging their parameters weighted by the model evidence. We report the posterior probabilities of the model with as compared to the model without the respective parameter. The significance threshold for the posterior probability was set to >.95. Leave-one-out cross-validation was used to assess whether the model parameters possessed predictive validity for the age of participants.

The selection of regions in each individual was guided by findings from previous studies and the results from our GLM analyses. As our main research question pertained to developmental changes of connectivity in corticostriatal regions, we selected five regions that play a significant role in incentive processing and are hypothesized to change their connectivity patterns throughout development (Cho et al., 2013; C. Insel et al., 2017; Van Den Bos et al., 2015). Thus, we chose one striatal, one thalamic and three cortical regions that spanned the network of interest.

For the VS, we used an anatomical mask derived from the Harvard–Oxford atlases (http://www.cma.mgh.harvard.edu/). For the cortical (LPFC [-33, 44, 22] MNI [mm]; insula [-39, 14, 0]; dACC [-1, 8, 46] MNI [mm]) and thalamic ([11, -14, 6] MNI [mm]) regions, we created search spheres with a radius of 8mm around the maximum group activation. In each region, we extracted the first eigenvariate of the time course of all voxels surpassing a threshold of 0.05 and within a radius of 3*mm* from the individual peak activation and adjusted it for the effects of interest. One subject was excluded from the DCM

analysis, as we did not find any active voxels surpassing our threshold in the thalamus. For the DCM analysis, the two scan sessions were concatenated with SPM. We added an additional nuisance regressor to the concatenated model that modelled the volumes at session transition. All stimulus cues were entered as driving input in the thalamus. The full model comprised a fully connected corticothalamic network that projected unidirectionally to the VS. In turn, the VS had one main output to the thalamus, modeling the principle anatomy of cortico-striatal-thalamic loops Haber & Knutson, 2010. Q_t^+ and $|Q_t^-|$ were allowed to modulate the self-connections of the VS, the insula, and the dACC. Limiting task modulation to the self-connections allows a straightforward biological interpretation of the modulatory parameter estimates, namely the change in synaptic gain for a given task context (Zeidman et al., 2019).

2.6 Supplementary information

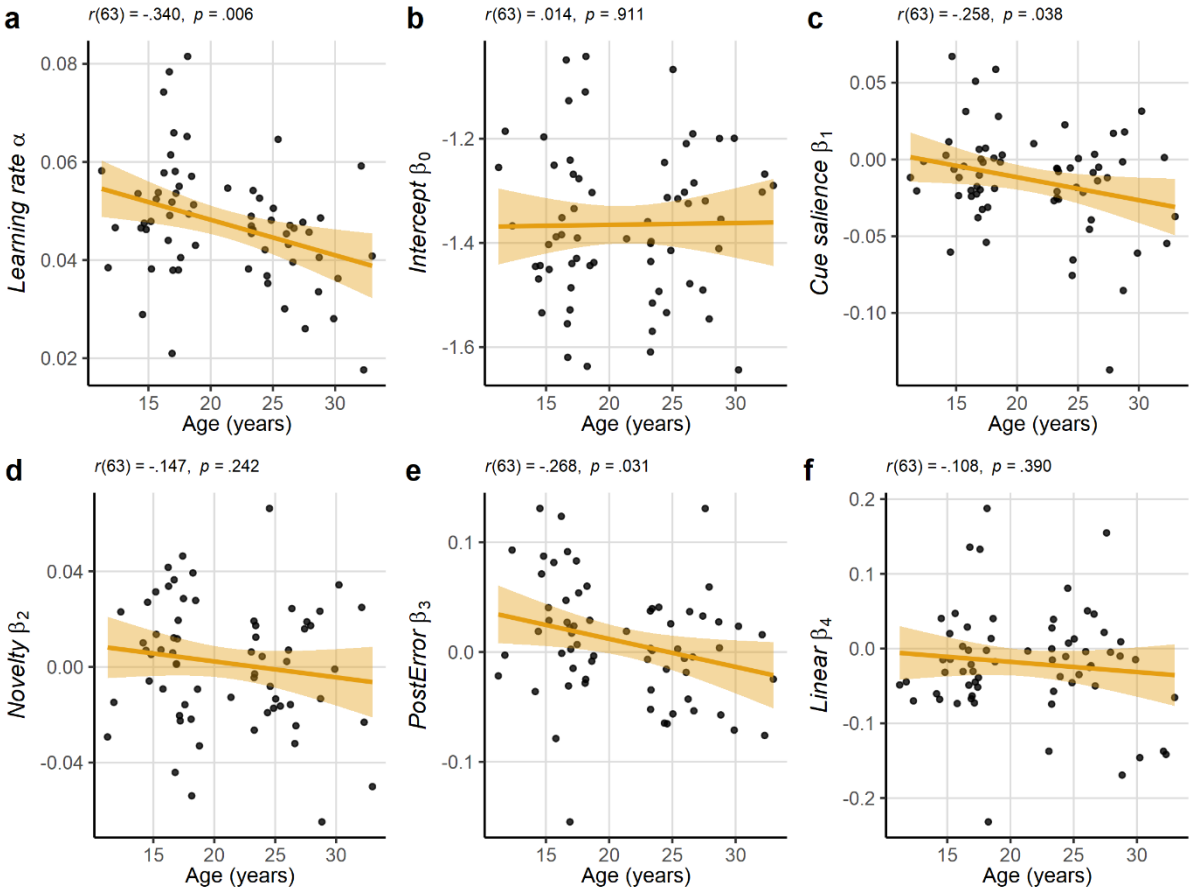

Figure S2.1. Pearson correlations of all behavioral parameters with age. As presented in the article, learning rate, cue salience and PostError parameter showed a significant correlation with age. *n* = 65.

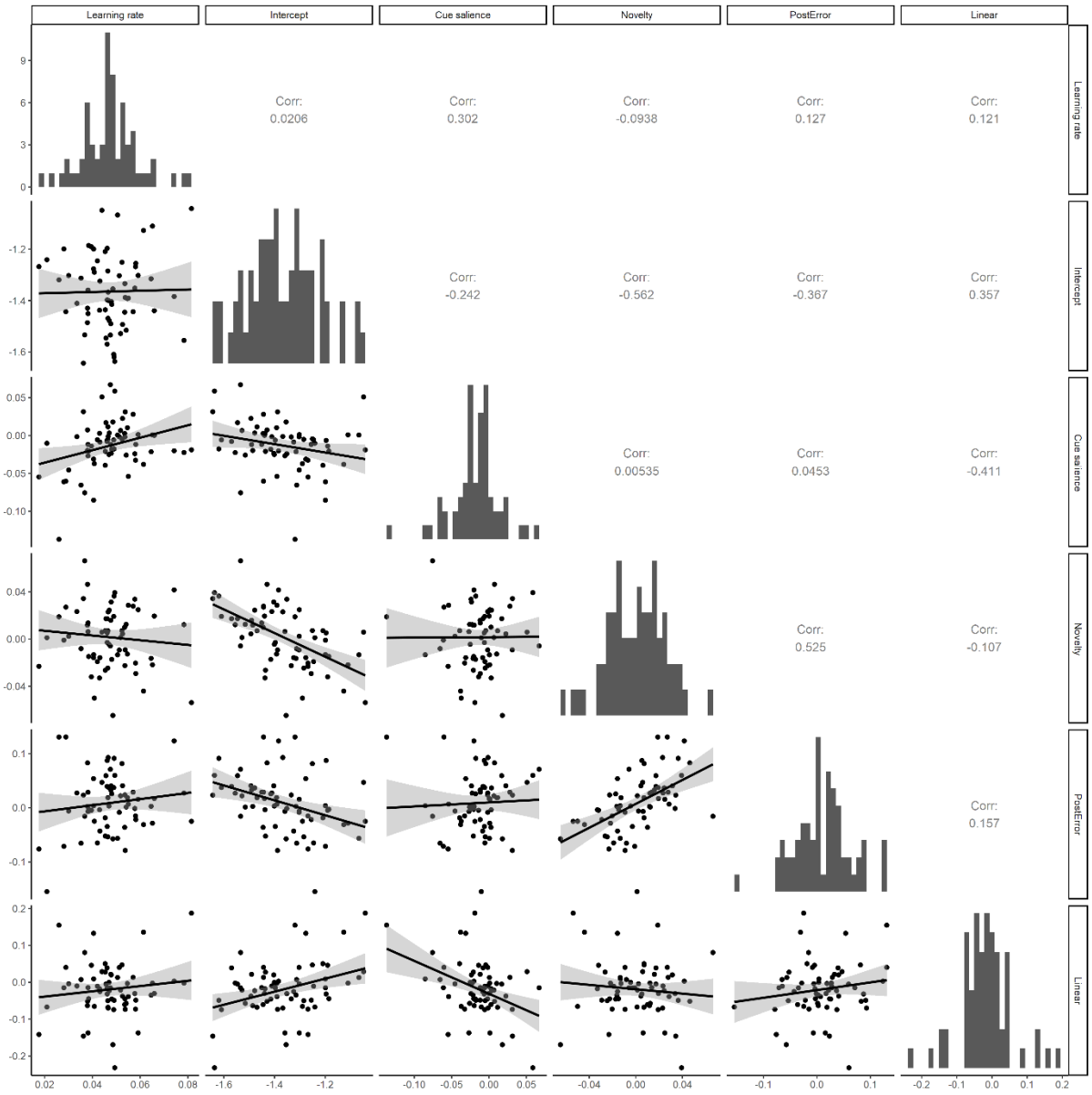

Figure S2.2. Intercorrelations between the parameters of the winning behavioral model. The mere moderate correlations between the parameters indicate that the effect captured by our behavioral model could be disentangled well. Data points represent the raw beta values of each subject. Corr. values represent Pearson's correlation coefficient. n = 65.

Table S2.1. Parameter prior means (variance) of the reinforcement learning model and the response models.

Reinforcement learning model					
α (logit-space)	0.05 (1)				
ν_0 (logit-space)	0.66 (0.5)				
Response models	**M1**	**M2**	**M3**	**M4**	**M5**
β_0	0 (4)	0 (4)	0 (4)	0 (4)	0 (4)
β_1	0 (4)	0 (4)	0 (4)	0 (4)	0 (4)
β_2	0 (4)	0 (4)	0 (4)	0 (4)	0 (4)
β_3	0 (4)	0 (4)	0 (4)	0 (4)	0 (4)
β_4	0 (4)	0 (4)	0 (4)	0 (4)	0 (4)
β_5	0 (4)			0 (4)	
β_6	0 (4)				
β_7	0 (4)				
ξ	log(3) (log(2))	log(3) (log(2))	log(3) (log(2))	log(3) (log(2))	log(3) (log(2))

Table S2.2. Results of the group analysis (*N* = 67). Significant clusters on whole-brain level in the second-level analyses for the main effect of anticipation and feedback and the group effects of the parametric modulations of Q^+, $|Q^-|$, δ^+, δ^-.

	MNI coordinates [mm]			Significant activation		Peak
Brain region	x	y	z	$p_{FWE(c)}$	k	Z
Average positive effect of anticipation[a]						
L Supplementary motor area	-1	8	46	<.001	26877	>8
L Precentral	-35	-24	56	<.001		>8
R Middle cingulate gyrus	7	10	36	<.001		>8
R Inferior occipital gyrus	35	-86	-2	<.001		>8
L Middle occipital gyrus	-25	-88	6	<.001		>8
L IFG pars opercularis	-59	8	12	<.001		>8
R IFG pars opercularis	47	16	0	<.001		>8
L Insula	-39	14	0	<.001		>8
R Insula	37	16	4	<.001		>8
R Putamen	21	12	4	<.001		>8
L Putamen	-23	10	4	<.001		>8
R Thalamus	11	-14	6	<.001		>8
L Thalamus	-13	-16	6	<.001		>8
L Supplementary motor area	-9	-6	60	<.001		>8
R Cerebellum	29	-64	-26	<.001	13580	>8
L Cerebellum	-33	-56	-32	<.001		>8
R Fusiform gyrus	37	-68	-14	<.001		>8
L Middle frontal gyrus	-33	44	22	<.001	894	7.45
R Middle frontal gyrus	33	42	26	<.001	496	7.00
R Supramarginal gyrus	55	-34	-44	<.001	1461	6.98
Brain stem	-3	-38	-40	<.001	80	6.98
L Middle occipital gyrus	-25	-74	28	<.001	73	6.09
R Inferior parietal	29	-54	54	<.001	135	5.94

R Rolandic operculum	29	-54	54	.001	88	5.68
R Precuneus	13	-70	42	.001	79	5.68
Average positive effect of feedback reception						
R Posterior cingulate gyrus	37	-34	-8	<.001	533	6.32
L Hippocampus	-25	-42	8	<.001	390	6.28
L Caudate nucleus	-1	10	14	<.001	415	6.16
R Fusiform gyrus	51	-68	-26	0.005	171	5.89
R Angular gyrus	39	-62	56	<.001	532	5.2
Effect of Q^+						
Average effect of Q^+						
R Middle temporal gyrus	41	-74	40	<.001	966	6.9
L IFG pars opercularis	-19	34	48	<.001	2066	5.84
R Cerebellum	-3	-82	-4	<.001	1546	5.59
L Middle temporal gyrus	-35	-72	38	<.001	1129	5.58
R Posterior cingulate gyrus	3	-22	6	<.001	366	5.42
R Middle cingulate gyrus	-5	-38	34	<.001	383	5.3
R Middle frontal gyrus	31	16	54	<.001	403	5.3
L Superior parietal lobule	9	-4	68	<.001	1723	4.98
R Putamen	13	18	-10	0.022	109	4.8
R Lingual gyrus	27	-52	6	0.001	198	4.79
L Inferior temporal gyrus	-57	-50	-14	<.001	259	4.72
L Parahippocampal gyrus	-37	-48	-2	<.001	273	4.66
L IFG pars triangularis	-49	40	6	0.003	159	4.59
R Superior parietal lobule	29	-38	60	<.001	250	4.48
L Middle frontal gyrus	-21	68	12	<.001	281	4.28
R Superior frontal gyrus, medial orbital	-3	52	-10	0.002	172	4.09
Effect of Age						
-						
Effect of Sex						
-						
Effect of Age X Sex						
-						
Effect of $\lvert Q^- \rvert$						
Average effect of $\lvert Q^- \rvert$						
R Middle temporal gyrus	41	-74	40	<.001	966	6.9
L IFG pars opercularis	-19	34	48	<.001	2066	5.84
R Cerebellum	-3	-82	-4	<.001	1546	5.59
L Middle temporal gyrus	-35	-72	38	<.001	1129	5.58
R Posterior cingulate gyrus	3	-22	6	<.001	366	5.42
R Middle cingulate gyrus	-5	-38	34	<.001	383	5.3
R Middle frontal gyrus	31	16	54	<.001	403	5.3
L Superior parietal lobule	9	-4	68	<.001	1723	4.98
R Putamen	13	18	-10	0.022	109	4.8
R Lingual gyrus	27	-52	6	0.001	198	4.79
L Inferior temporal gyrus	-57	-50	-14	<.001	259	4.72
L Parahippocampal gyrus	-37	-48	-2	<.001	273	4.66

L IFG pars triangularis	-49	40	6	0.003	159	4.59
R Superior parietal lobule	29	-38	60	<.001	250	4.48
L Middle frontal gyrus	-21	68	12	<.001	281	4.28
R Superior frontal gyrus, medial orbital	-3	52	-10	0.002	172	4.09
L Superior temporal pole	-35	18	-18	<.001	721	7.04
R Superior temporal pole	33	18	-16	<.001	901	6.66
R Superior frontal gyrus, medial	3	16	60	<.001	2650	6.35
L IFG pars triangularis	-31	58	8	<.001	354	5.52
L IFG pars opercularis	-43	36	8	<.001	596	5.16
L Parahippocampal gyrus	-21	-16	-26	0.003	163	5.04
L Middle temporal gyrus	-45	-26	-18	0.005	151	4.66
Effect of Age						
R Middle cingulate gyrus	3	-34	20	<.001	2617	5.16
L IFG pars triangularis	-33	32	34	<.001	705	5.01
R Superior occipital gyrus	33	-52	-24	<.001	305	4.91
L Inferior parietal lobule	-69	-24	28	<.001	1327	4.86
R Lingual gyrus	23	-90	-14	<.001	751	4.83
L Superior occipital gyrus	-35	-90	-14	<.001	483	4.76
R Anterior cingulate gyrus	27	46	28	<.001	323	4.67
L Caudate nucleus	-17	-12	16	0.001	99	4.59
R Superior temporal gyrus	61	-38	38	<.001	788	4.53
R Middle frontal gyrus	21	56	-8	<.001	155	4.45
L Superior temporal gyrus	-39	22	6	<.001	480	4.42
R Superior parietal lobule	7	-62	68	0.049	91	4.39
R Caudate nucleus	17	-10	18	<.001	265	4.19
R Supplementary motor area	27	-18	66	0.012	123	4.14
R Hippocampus	31	16	2	<.001	232	4.06
R Middle frontal gyrus	41	44	14	0.033	100	3.93
Effect of Sex						
-						
Effect of Age X Sex						
-						
Effect of δ^+						
Average effect of δ^+						
R Superior frontal gyrus	-5	44	-14	<.001	3018	6.04
L Olfactory cortex	-17	6	-16	<.001	3002	5.89
R Cerebellum	25	-88	0	<.001	976	5.69
R Paracentral lobule	-19	-28	62	<.001	2285	5.12
R Caudate nucleus	17	4	20	<.001	464	4.98
R Superior temporal pole	55	-12	-2	<.001	233	4.97
R Angular gyrus	27	-50	24	0.004	143	4.82
L Cerebellum	-17	-52	-22	0.002	166	4.77
L Posterior cingulate gyrus	-13	-36	6	<.001	331	4.67
L IFG pars triangularis	-55	22	20	<.001	206	4.65
R Insula	55	0	6	0.039	93	4.61
R Middle cingulate gyrus	5	34	44	0.015	114	4.55

L Precentral gyrus	-31	-12	38	0.029	99	4.55
R Middle frontal gyrus	31	54	10	0.034	96	4.33
R Insula	33	26	4	<.001	517	4.24
L Inferior occipital gyrus	-19	-84	6	<.001	277	4.09
Effect of Age						
L Cerebellum	-7	-58	-16	<.001	576	4.64
R Fusiform gyrus	41	-56	-20	0.010	117	4.21
L Superior temporal gyrus	-43	-48	6	0.013	102	4.17
Effect of Sex						
-						
Effect of Age X Sex						
R Superior temporal gyrus	61	-2	-10	0.032	97	4.45
R Caudate	13	10	-14	0.021	106	4.41
Post hoc t-test of negative Age X Sex interaction						
R Caudate	13	10	-14	0.007	155	4.60
R Superior temporal gyrus	61	-2	-10	0.032	97	4.58
R Cerebellum	47	-68	-26	0.045	106	4.56
L Putamen	-23	8	-4	0.022	106	4.43
L Middle cingulate gyrus	-13	-38	50	0.008	151	3.88
Effect of δ^-						
Average effect of δ^-						
L Precentral gyrus	7	-30	60	<.001	37170	7.19
L Superior temporal pole	-35	18	-18	<.001	721	7.04
R Superior temporal pole	33	18	-16	<.001	901	6.66
R Superior frontal gyrus, medial	3	16	60	<.001	2650	6.35
L IFG pars triangularis	-31	58	8	<.001	354	5.52
L IFG pars opercularis	-43	36	8	<.001	596	5.16
L Parahippocampal gyrus	-21	-16	-26	0.003	163	5.04
L Middle temporal gyrus	-45	-26	-18	0.005	151	4.66
Effect of Age						
-						
Effect of Sex						
R Postcentral gyrus	47	-32	58	0.029	108	4.23
Effect of Age X Sex						
-						

Significance level at whole-brain cluster-level threshold $p_{FWEc} < 0.05$, cluster-defining threshold at $p_{CDT} < 0.001$.

[a]As the positive effect of anticipation resulted in a huge cluster of several 10.000 voxels, we applied a voxel-wise FWE corrected threshold of $p_{FWE} < 0.05$ for this contrast and report cluster peaks with k > 50. Labels of brain regions were determined using the SPM Anatomy Toolbox.

k, cluster size; R, right; L, left.

3 Maladaptive learning from aversive outcomes in the orbitofrontal cortex in adolescents with major depression

David Willinger[a,b], Iliana I. Karipidis[a,b,c], Selina Neuer[a], Sophie Emery[a], Carolina Rauch[a], Isabelle Häberling[a], Gregor Berger[a], Susanne Walitza[a,b], Silvia Brem[a,b]

3.1 Overview

Understanding the mechanisms in the brain's incentive network that give rise to symtpoms of MDD during adolescence provides new perspectives to address MDD in early stages of development. This study determines whether instrumental vigor and brain responses to appetitive and aversive monetary incentives are altered in adolescent MDD and associated with symptom severity. In this functional magnetic resonance imaging (fMRI) study, adolescents with moderate to severe MDD (n = 30), and healthy controls (n = 33) matched for age, sex, and IQ performed a monetary incentive delay task. During outcome presentation, prediction error signals were used to study the response and functional coupling of the incentive network during learning of cue-outcome associations. A computational reinforcement model was used to assess adaptation of instrumental response vigor. Brain responses and effective connectivity to model-derived prediction errors were assessed and related to depression severity and anhedonia levels. Participants with MDD behaved according to a more simplistic learning model and exhibited slower learning. Neuroimaging data showed impaired loss error processing in the orbitofrontal cortex associated with aberrant gain-control revealed by an effective connectivity analysis. Anhedonia scores correlated with loss related error signals in the posterior insula and habenula. Adolescent MDD is selectively related to impaired processing of error signals during loss, but not reward, in the orbitofrontal cortex. Aberrant evaluation of loss outcomes might reflect an early mechanism of how negative bias and helplessness manifest in the brain. This approach sheds light on pathomechanisms in MDD and may improve early diagnosis and treatment selection.

3.2 Introduction

Major depressive disorder (MDD) is among the most prevalent mental health problems in adolescents worldwide (Lopez, Mathers, Ezzati, Jamison, & Murray, 2006) with an estimated 12 months prevalence of 7.5% in mid to late adolescence (Avenevoli et al., 2015). Adolescent MDD increases the risk for substance misuse, can severely impair success in school, social life, and cognitive functions (Kieling et al., 2019), and is a major risk factor for suicide, which is among the leading causes of death at this age (Cha et al., 2018). Despite these adverse outcomes, relatively little is known about brain mechanisms related

[a] Department of Child and Adolescent Psychiatry and Psychotherapy, University Hospital of Psychiatry Zurich, University of Zurich, Switzerland
[b] Neuroscience Center Zurich, University of Zurich and ETH Zurich, Switzerland
[c] Center for Interdisciplinary Brain Sciences Research, Stanford University School of Medicine, Stanford, CA, USA

We thank Plamina Dimanova, Nada Frei, Noemi Baumgartner, Mona Albermann, Kristin Nalani, Paola Keller, Desiree Thommen, Luana Signer, and Philipp Stämpfli for technical and clinical assistance.

to MDD with early onset. Recent evidence suggest that disrupted prediction error (PE) signaling constitutes a potential brain mechanism that promotes the persistence of negative beliefs and anhedonia (C. Chen, Takahashi, Nakagawa, Inoue, & Kusumi, 2015).

It is widely established that the dopaminergic system is fundamental in encoding reward and loss PEs (Kumar et al., 2018), which are crucial in reinforcement learning and decision making. Influential computational models (Rescorla & Wagner, 1972; Sutton, 1988) suggest that during the anticipation of an incentive an expected value (Q) signal is generated, which is the product of a learned probability and the magnitude of the incentive. During outcome receipt, the difference between the expected value signal and the actual outcome is signaled as PE to update predictions. While the ventral striatum primarily encodes PEs in reward contexts, the anterior insula does so in avoidance or loss contexts (Palminteri et al., 2012). In addition, brain regions sensitive to errors in reward and loss contexts are the orbitofrontal cortex (OFC) and the dorsal anterior cingulate cortex (dACC; Knutson, Fong, Bennett, Adams, & Hommer, 2003; S. F. Taylor et al., 2006).

Previous studies have demonstrated that PE signaling is deficient in adult MDD in reward (Gradin et al., 2011; Kumar et al., 2008) and loss contexts (Kumar et al., 2018). Therefore, it is crucial to establish whether this aberrant signaling is already present in early onset MDD or whether deviations in reward and loss processing are a downstream effect of chronicity and burden (Kumar et al., 2018). Studies in young cohorts suggest that blunted reward sensitivity in the ventral striatum predicts symptom deterioration (B. D. Nelson, Perlman, Klein, Kotov, & Hajcak, 2016; Stringaris et al., 2015) and is present in individuals at high familial risk for depression (Luking, Pagliaccio, Luby, & Barch, 2016b; Sharp et al., 2014). In addition, there is emerging evidence of impaired loss sensitivity in the incentive network in high risk groups (Luking, Pagliaccio, Luby, & Barch, 2016a) that predicts future depressive symptoms (Jin et al., 2017). However, it remains unclear whether atypical learning signals are linked to deficient adaptive, motivated behavior. Previous studies showed mixed results when applying computational models to behavioral data in adult MDD, with learning rates depending on the task used and the specific learning process probed (Scholl & Klein-Flügge, 2018). This clearly indicates that more work is necessary to identify brain mechanisms that give rise to aberrant incentive processing in depression, particularly during development.

In this functional magnetic resonance imaging (fMRI) study, we hypothesized that the encoding of reinforcement learning signals is impaired in adolescent MDD. We employed an monetary incentive delay task (Knutson et al., 2003; Knutson et al., 2000) with varying magnitude (low, high) and valence (reward, loss) to probe the neural circuits supporting PE and expected value processing. On a neural level, we hypothesized (a) decreased reward PE signaling within the striatum in MDD (C. Chen et al., 2015; Stringaris et al., 2015), (b) a negative association between anhedonia scores and blunted responses to rewards in the striatum and the OFC (Der-Avakian & Markou, 2012), and (c) reduced reactivity of the OFC during loss events (Jin et al., 2017). Behaviorally, we tested whether instrumental

vigor was differentially modulated in MDD, and whether there are differences in the update of value representations in the instrumental learning task.

3.3 Methods

3.3.1 Participants

Thirty MDD patients and 33 healthy individuals matched for age, IQ, gender, and handedness participated in this study (Table 3.1). Participants with MDD were recruited through clinical services. All participants underwent a semistructured clinical interview (K-SADS-PL, Kaufman et al., 1997, or MINI-KID, Sheehan et al., 2010).

Table 3.1. Clinical and demographic characteristics of study participants

	Controls	MDD	Test statistic	*p* value[a]
Age (years), mean (SD)	16.2 (1.9)	16.1 (1.4)	$U = 553.5$	.425
Sex (males), No. (%)	10 (30%)	10 (33%)	$^2(1) = 0.07$	.796
Handedness (right), No. (%)	32 (97%)	28 (93%)	$^2(1) = 0.46$	.500
In-scanner movement (FD, *mm*)	0.18 (0.09)	0.18 (0.08)	$t(61) = 0.09$	.930
CD-RISC, mean (SD)	72.9 (10.1)	38.6 (15.6)	$t(58) = 10.16$	< .001
CDI, mean (SD)	8.4 (6.6)	29.6 (9.3)	$U = 38.0$	< .001
Anhedonia	2.3 (2.2)	10.5 (2.8)	$U = 13.5$	< .001
Negative mood	2.2 (2.0)	6.4 (2.4)	$U = 88.0$	< .001
Negative self-esteem	1.0 (1.2)	5.0 (1.7)	$U = 42.0$	< .001
Ineffectiveness	1.2 (1.2)	5.0 (1.9)	$U = 54.5$	< .001
Interpersonal problems	1.1 (1.2)	3.7 (1.5)	$U = 74.5$	< .001
Stomach	0.6 (0.6)	1.1 (0.8)	$U = 301.5$	.018
RIAS IQ, mean (SD)	104.5 (6.9)	108.0 (8.7)	$t(60) = -1.75$	.079
PSS, mean (SD)	22.4 (6.6)	28.8 (7.7)	$t(57) = -3.44$	.001
SDQ, mean (SD)	8.8 (5.3)	16.3 (5.6)	$t(56) = -5.26$	< .001
WISC-IV Digitspan (forward), mean (SD)	8.9 (2.1)	8.8 (2.0)	$t(60) = 0.32$	.747
WISC-IV Digitspan (backward), mean (SD)	8.6 (1.6)	9.4 (2.0)	$t(60) = -1.70$	.094
WISC-IV Mosaic, mean (SD)	57.0 (5.7)	59.0 (6.2)	$t(56) = -1.27$	.208
Current Medication, No. (%)				
No medication	NA	10 (33%)	NA	NA
SSRI	NA	18 (60%)	NA	NA
Dual-action antidepressant[b]	NA	2 (7%)	NA	NA
NERI	NA	2 (7%)	NA	NA
Antipsychotic[c]	NA	2 (7%)	NA	NA
Methylphenidate	NA	2 (7%)	NA	NA

Abbreviations: CD-RISC, Connor-Davidson Resilience Scale; CDI, Children Depression Inventory; FD, framewise displacement; RIAS, Reynolds Intellectual Assessment Scales; PSS, Perceived Stress Scale; SDQ-K, Strength and Difficulty Questionnaire for Children; WISC, Wechsler Intelligence Scale for Children.

[a] Uncorrected *p* values for between-group comparisons; significance threshold $p < .05$.

[b] Serotonin-noradrenalin reuptake inhibitor

[c] Used for behavioral control

Participants with MDD fulfilled a diagnosis according to the DSM-IV (codes 296.20-296.23, 296.30-296.33). Past and present comorbid diagnoses in patients comprised anxiety disorders ($n = 7$), obsessive-compulsive disorder ($n = 1$), and attention-deficit

hyperactivity disorder (n = 2). Moreover, we assessed a battery of self-report questionnaires, IQ and working memory of all participants (Table 3.1). We included total scores and scores from the anhedonia subscale from the German version of the Child Depression Inventory (CDI; Stiensmeier-Pelster, Braune-Krickau, Schürmann, & Duda, 2014) in our neuroimaging analyses. Healthy 8-18-year old controls (HC) were recruited through schools and volunteer websites. For controls, exclusion criteria comprised any current psychiatric disorder, other major medical illnesses, drug abuse, any MRI contraindication, pregnancy, and a history of brain injury. Three control participants had a past diagnostic work-up for ADHD but they were currently symptom-free and were not taking any medication during the study. This study was approved by the ethics committee of the Kanton Zürich and was conducted in accordance with the Declaration of Helsinki. All participants gave their written informed consent, parents or legal guardians gave signed informed consent for children under the age of 14 years. They were reimbursed for participation and informed about the opportunity to additionally win up to CHF 20 during the task.

3.3.2 Experimental task

Participants performed a monetary incentive delay task (Knutson et al., 2003), in which they had to respond quickly to a visual go-symbol with a button press to either gain or avoid losing money. Valence (potential reward or loss) and magnitude (high or low) of the incentive was indicated by a cue in the beginning of each trial. For a detailed task description, please see Chapter 2. After the scanning session, participants were asked to rate their arousal and valence for each trial type (Table S3.1, Figure S3.2).

3.3.3 Image acquisition and preprocessing

For a detailed description of the data acquisition, please refer to Chapter 2. The preprocessing of functional data comprised slice-timing correction, realignment and subsequent coregistration to the anatomical (T1-weighted) scan. The data was transformed into MNI-152 space with deformation field normalization and smoothed with $6mm$ FWHM kernel. To account for motion artefacts during the scan, we calculated the framewise displacement (FD) across volumes (Power et al., 2012). No subject exceeded a mean FD of $0.5mm$ (HC: M = 0.18, SD = $0.09mm$; MDD: M = 0.18, SD = $0.08mm$), however, single volumes that exceeded a FD greater than $1mm$ were censored in the ensuing analyses by including an additional binary regressor (% volumes censored per participant M = 0.92, SD = 1.90%).

3.3.4 Behavioral raw data analysis

Response data were log-transformed to achieve a more normally distributed data set for data analysis. Data that deviated more than three standard deviations from the respective mean per condition and per subject were excluded from the analysis (2%). For a conventional analysis, we averaged the logRTs for each subject and each condition and performed an ANOVA to assess effects of condition, group, and their interaction.

The post-scan ratings were mean-centered within subject and rotated to obtain measures of positive and negative arousal for each trial type (high/low reward, high/low loss, neutral) as they increase interpretability in terms of approach and avoidance behavior (Knutson, Katovich, & Suri, 2014). To this end, we calculated negative and positive arousal for each outcome according to the equations:

$$\text{Positive Arousal} = \frac{(\text{Arousal} + \text{Valence})}{\sqrt{2}} \tag{3.1}$$

$$\text{Negative Arousal} = \frac{(\text{Arousal} - \text{Valence})}{\sqrt{2}} \tag{3.2}$$

We had to exclude four controls from this behavioral analysis as they did not complete the ratings due to time constraints. We performed an ANOVA with condition and group as fixed factor. The significance level for all statistical tests of the behavioral analyses was $p < 0.05$, two-tailed.

3.3.5 Computational Modeling

The task employed here allows to assess mechanisms that determine behavior (i.e. instrumental response vigor). To assess these quantities we constructed several competing generative behavioral models that predicted trial-by-trial reaction times for each participant. This allowed us to identify parameters with mechanistic meaning for observed response vigor, the latent representation of value, and the participant-specific learning rate.

The computational model was similar to the one used in Chapter 2 to model response vigor. However, as behavioral adaptation might occur with different rates in controls and patients in reward and loss contexts, respectively, we defined two different reinforcement learning models: one model with a single learning rate for both updates, one with separate learning rates for reward and loss updates. The update rules for the hit probability in the subsequent trial in the respective models was therefore:

Rescorla-Wagner Model 1 (rw1)

$$v_{t+1} = \begin{cases} v_t + \alpha \cdot \varphi_t & C_t \neq 0 \\ v_t & C_t = 0 \end{cases} \tag{3.3}$$

Rescorla-Wagner Model 2 (rw2)

$$v_{t+1} = \begin{cases} v_t + \alpha^+ \cdot \varphi_t^+ & C_t > 0 \\ v_t & C_t = 0 \\ v_t + \alpha^- \cdot \varphi_t^- & C_t < 0 \end{cases} \tag{3.4}$$

where φ_t denotes the overall prediction error in both contexts.

As in previous work (Cao et al., 2019), the signal for updating future predictions φ was scaled by the magnitude of the experienced outcome, such that the update was independent of the experienced reward/loss value:

$$\varphi_t^+ = \begin{cases} \frac{R_t}{C_t} - v_t & C_t \neq 0 \\ 0 & C_t = 0 \end{cases} \tag{3.5}$$

$$\varphi_t^- = \begin{cases} (1 - v_t) - \frac{R_t}{C_t} & C_t \neq 0 \\ 0 & C_t = 0 \end{cases} \tag{3.6}$$

By this definition, outcome errors φ_t^+ and φ_t^- reflect "surprise" signals that can be used for learning stimulus-outcome associations in either reward or loss context, whose size depends solely on the reinforcement history and not value.

The expected reward Q_t^+ and the expected loss Q_t^- dependent on the subjective probability for a reward v_t and loss $(1 - v_t)$ and the possible outcome C_t.

$$Q_t^+ = C_t \cdot v_t \tag{3.7}$$

$$Q_t^- = C_t \cdot (1 - v_t) \tag{3.8}$$

Conventional (magnitude-dependent) reward (δ_t^+) and loss (δ_t^-) prediction error signals were calculated as the difference between expected value and outcome

$$\delta_t^+ = \begin{cases} R_t - Q_t^+ & C_t > 0 \\ 0 & C_t \leq 0 \end{cases} \tag{2.3}$$

$$\delta_t^- = \begin{cases} Q_t^- - R_t & C_t < 0 \\ 0 & C_t \geq 0 \end{cases} \tag{2.4}$$

In addition, average reward and loss at each trial was defined as:

$$\bar{R}_t = \begin{cases} \bar{R}_{t-1} + \alpha \cdot (R_{t-1} - \bar{R}_{t-1}) & C_t > 0 \\ \bar{R}_{t-1} & C_t \leq 0 \end{cases} \tag{2.6}$$

$$\bar{L}_t = \begin{cases} \bar{L}_{t-1} + \alpha \cdot (|L_{t-1}| - \bar{L}_{t-1}) & C_t < 0 \\ \bar{L}_{t-1} & C_t \geq 0 \end{cases} \tag{2.7}$$

where R_t and L_t represent the actual and $\bar{R}_t$ and $\bar{L}_t$ the average reward or loss at trial t.

For details on the response models please refer to Chapter 2, here, the definition is repeated for completeness:

Response model M1:

$$\log(RT)^t = \beta_0 + \beta_1 \cdot |Q_t^-| + \beta_2 \cdot Q_t^+ + \beta_3 \cdot \bar{R}_{t-1} + \beta_4 \cdot \bar{L}_{t-1} + \\ + \beta_5 \cdot PostError_t + \beta_6 \cdot Rep_t + \beta_7 \cdot g(t) + \zeta_t \tag{2.8}$$

Response model M2:

$$\log(RT)^t = \beta_0 + \beta_1 \cdot |Q_t^-| + \beta_2 \cdot Q_t^+ + \beta_3 \cdot \delta_{t-1}^- + \beta_4 \cdot \delta_{t-1}^+ + \beta_5 \cdot g(t) + \zeta_t \tag{2.9}$$

Response model M3:

$$\log(RT)^t = \beta_0 + \beta_1 \cdot |Q_t| + \beta_2 \cdot |\delta_{t-1}| + \beta_3 \cdot PostError_t + \beta_4 \cdot g(t) + \zeta_t \tag{2.10}$$

Response model M4:

$$\log(RT)^t = \beta_0 + \beta_1 \cdot |Q_t^-| + \beta_2 \cdot Q_t^+ + \beta_3 \cdot |\delta_{t-1}| + \beta_4 \cdot PostError_t + \beta_5 \cdot g(t) + \zeta_t \quad (2.11)$$

Response model M5:

$$\log(RT)^t = \beta_0 + \beta_1 \cdot |Q_t| + \beta_2 \cdot \delta_{t-1}^- + \beta_3 \cdot \delta_{t-1}^+ + \beta_4 \cdot g(t) + \zeta_t \quad (2.12)$$

In these models, $PostError$ reflects a binary vector of trials after a miss, Rep is a binary vector of successive presentation of the same cues, and ζ denotes Gaussian noise.

All behavioral models were fitted using the TNU Algorithms for Psychiatry-Advancing Science (TAPAS, http://www.translationalneuromodeling.org/tapas) HGF Toolbox 5.3. Trials without response were omitted during model fitting; priors are reported in Table S3.2. We used random-effects Bayesian Model Selection (Stephan, Penny, Daunizeau, Moran, & Friston, 2009) to choose the best-fitting model by comparing the negative free energies. Besides the posterior probabilities, the exceedance probability (XP) of each model is reported. In a last step, we assessed group differences of response parameters. Note, that for the learning rate parameter we detected an outlier after visual inspection and a significant Grubb's test. Thus, for this parameter, we performed a group comparison with and without this subject. Intercorrelations between parameters were assessed after excluding this subject for this analysis. However, as this participant was a patient, and the higher learning rate might reflect a pathological process, we included this subject in further analyses. For group comparisons, we used either two-sample t-tests or Mann-Whitney U-tests, in case the Shapiro-Wilk test indicated violation of normality assumptions in at least one of the groups. Finally, posterior predictive checks were conducted to assess the reliability of the behavioral model. For this, we averaged the logRTs of 1000 simulations with the individual parameters of the best-fitting model for each subject in TAPAS and mirrored the raw data analysis with the synthetic data.

3.3.6 Functional MRI analysis

To investigate the trial-by-trial effect of the computational variables derived from the computation model, we used the variables of the winning model across participants. We modelled the first-level with separate GLMs, where we entered the expected values during cue onset and reward-/loss prediction error during feedback onset as parametric modulators for reward and loss separately. The effect of the scaled update signal, i.e. outcome error, was assessed in a model using one parametric modulator for outcome magnitude and one for the magnitude-independent error signals for reward and loss trials separately. The modulator for the outcome error was orthogonalized with respect to the outcome magnitude regressor, such that any shared variance between the correlating regressors was assigned to the latter. This outcome error regressor captures the deviation from the expected outcome (hit or miss) independent of the magnitude context. With this approach, we not only investigated brain regions encoding the effect of conventional reward/loss prediction errors with the multiplicative term of incentive probability × out-

come, but also reveal brain regions, that code deviations from expected reward/loss outcomes across trials irrespective of magnitude. All first-level models included six realignment parameters and a binary vector for scans with > 1*mm* FD as nuisance regressor to the model. Neutral trials were modelled in a separate regressor. Finally, we applied a 1/128Hz cut-off high-pass filter to eliminate low frequency drifts. Group effects were assessed with two-sample *t*-tests, where we entered the contrast images for the expected values and the prediction errors for healthy controls and MDD patients. The cluster-level significance threshold for the whole-brain group analyses was set to $p_{FWEc} < 0.05$ with a cluster-defining threshold of $p_{CDT} < 0.001$. All fMRI analyses were conducted in SPM12 (7487), labels for brain regions are based on the Automated Anatomical Atlas (Rolls, Huang, Lin, Feng, & Joliot, 2020).

To reveal the functional coupling between regions of the incentive network in participants with and without MDD, we performed a dynamic causal modeling (DCM; Friston et al., 2003; Zeidman et al., 2019) analysis. The regions for this analysis were selected based on (a) previously published findings of incentive processing (Hauser et al., 2015; Knutson et al., 2003; Palminteri et al., 2012), (b) findings of studies in participants with a history and at risk for MDD (Jin et al., 2017; Luking et al., 2016a) and (c) our second-level general linear model (GLM) analyses (Figure 3.1A, Table S3.4-S3.8, Figure S3.3-S3.5). The aforementioned studies have provided compelling evidence that the insula and the dACC play a significant role in loss-avoidance learning, a finding we corroborate across groups during reward and loss processing. Furthermore, we found a significant group effect in the OFC that suggested aberrant network dynamics in MDD.

We localized the effects of hits and misses across trials, by performing whole-brain contrasts and using the CDI as covariate (Figure S3.5, Table S3.7). We extracted the timeseries for each subject from activated voxels ($p < 0.05$) within a 12*mm* spherical search volume around the group maxima from the miss-hit (insula [35, 18,-18]; dACC [7, 36, 30] MNI [mm], Table S3.7) and the all-events (IOG [29, -80, -16] MNI [mm], Table S3.8) contrast. The search volume for the OFC ([-7, 46, -16] MNI [mm]) comprised the voxels in the active OFC cluster of the second-level hit-miss contrast (Table S3.7, cluster-extent threshold $p_{FWEc} < 0.05$). If a participant's maximum within the search volume differed from the group maximum, we centered the sphere around the participant's maximum. The first eigenvariate of the time course of all active voxels ($p < 0.05$) was then extracted and adjusted for any motion effects. One patient and two controls had to be excluded from this analysis, as they did not show any activation in the dACC for the defined threshold.

The feedback regressor was the driving input for the visual region. The model comprised direct forward connections from the visual area to all other fully interconnected regions. Although our main interest was to study network effects during the loss condition, the reward condition was also included to fit the DCM to the timecourses. We included contextual modulation of prediction errors, magnitude (-1 for low and +1 for high), and their interaction term, i.e. magnitude–sensitive PEs on the self-connections of the regions. In

this model, the self-connections embody the change in synaptic gain for a given task context (Zeidman et al., 2019). Here, our goal was to identify the network dynamics that give rise to the lower error signal in the OFC in MDD patients during loss processing. For this, we set up a DCM analysis within the Parametric Empirical Bayes (PEB) framework.

On the first-level, the full model of each participant was estimated iteratively in an empirical Bayesian inversion scheme (Zeidman et al., 2019). The individual DCM parameters from the first-level were then entered in the second-level PEB model to determine (1) the differences between the MDD group and controls and (2) the group mean. This analysis was carried out separately for intrinsic and modulatory connections. We performed Bayesian model reduction to iteratively discard those model parameters not contributing to the model evidence. Then, we averaged the parameters of the best PEB models weighted by the posterior probability of the respective model. Leave-one-out cross-validation was used to assess whether the predicted and actual group effect showed an independent out-of-sample correlation.

3.4 Results

3.4.1 Altered learning of cue-outcome associations in MDD

Bayesian model comparison revealed that the response model including *cue salience* and *novelty* terms using a single learning rate fitted the response data best across groups. Nevertheless, we found that a more complex model with separate learning rates for rewards and losses fitted data better in controls only (exceedance probability, XP = 51.6%), whereas for patients the simpler model with a single learning rate performed better (XP = 57.7%, Table 3.2).

Table 3.2. Bayesian model comparison. Results showed that the dual learning rate model fitted the response data best in controls, whereas for MDD patients a simpler model with one learning rate performed better. Across all subjects, the simpler model provided the best model fit.

Model	**MDD**		**Controls**		**All subjects**	
	PP	XP	PP	XP	PP	XP
M1-rw1	1.0	2.5	1.0	2.3	1.0	1.4
M2-rw1	1.0	2.5	1.0	2.3	1.0	1.4
M3-rw1	23.1	57.7*	11.1	25.9	40.5	55.4*
M4-rw1	1.0	2.6	1.0	2.4	1.0	1.4
M5-rw1	1.5	3.8	1.8	4.1	1.7	2.4
M1-rw2	1.0	2.5	1.0	2.3	1.0	1.4
M2-rw2	1.0	2.5	1.0	2.3	1.0	1.4
M3-rw2	7.8	19.6	22.1	51.6*	22.9	31.4
M4-rw2	1.0	2.6	1.0	2.4	1.0	1.4
M5-rw2	1.5	3.8	1.9	4.4	1.8	2.4

Asterisks indicate the winning model. rw1: Rescorla-Wagner with a single learning rate; rw2: Rescorla-Wagner with a dual learning rate; PP: expected posterior probability; XP: exceedance probability.

A between-group comparison of parameters of the best-fitting model across all participants showed that the learning rate was marginally lower in MDD (α: MDD, M = 0.050,

SD = 0.021; controls, *M* = 0.052, *SD* =0.012; *U*=617, *p*=.095). This difference was significant after removing one outlier (Grubb's test: *G*=6.107, *U*=0.389, $p < 10^{-11}$; α: MDD, *M* = 0.046, *SD* = 0.009; controls, *M* = 052, *SD* = 0.012; t(60)=2.04, *p*=.046). The response model parameters did not differ significantly between groups (all $p > .10$, Table S3.3). These results demonstrate a non-discriminable value update mechanism underlying adolescent MDD for both reward and loss conditions, whereas the controls' response vigor was best described by a more flexible dual update model for both valences. In addition, the comparison of the learning parameter shows that MDD participants changed their value expectations slower across the task.

3.4.2 OFC gain control explains atypical aversive outcome signaling in MDD

When processing loss feedback, participants with MDD showed a significantly lower response to the outcome error signal φ_t^- in the OFC (Figure 3.1, Table 3.3). While this region reflected a signal encoding the difference between subjective belief of the outcome and the actual outcome in reward and loss in controls (Table S3.4), patients only expressed this signal during rewarding and not loss-avoidance trials. To further scrutinize the origin of this effect, we performed a DCM analysis (Figure 3.2, Table 3.5). Bayesian model averaging showed that the effect of loss-magnitude on the self-inhibition parameter of the OFC was significantly more negative in MDD, i.e. the region was more disinhibited during processing the outcome of high compared to low loss.

Table 3.3. Differences between MDD and HC groups for outcome error φ_t^- processing

	MNI coordinates [mm]			Significant activation		Peak
Brain region	x	y	z	p_{FWEc}	k	Z
Controls > MDD						
L Precuneus	-3	-46	10	.022	128	4.67
R Superior frontal gyrus, dorsolateral	17	28	44	.046	108	4.48
L Middle temporal gyrus	-67	-50	-6	.000	269	4.24
L Angular gyrus	-55	-64	24	.006	166	4.08
R Middle temporal gyrus	65	-10	-16	.034	116	4.05
L Medial orbital frontal gyrus	-3	60	-8	.019	132	3.93
MDD > Controls						
NS						

Significance level at whole-brain cluster-level $p_{FWEc} < .05$, cluster-defining threshold $p_{CDT} < .001$. Abbreviations: k, cluster size; R, right; L, left; MNI, Montreal Neurological Institute.

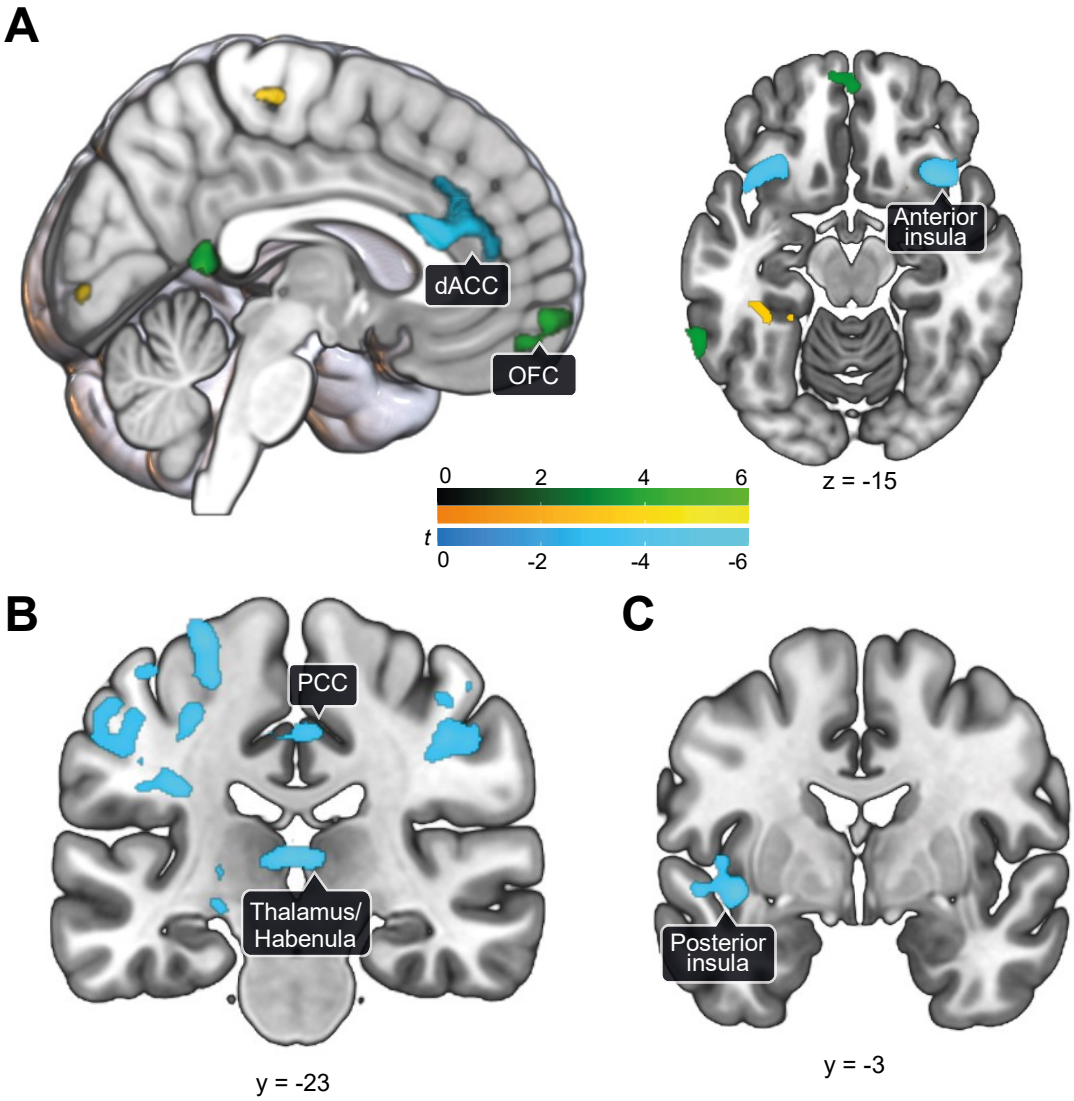

Figure 3.1. (A) Effect of loss outcome error $\boldsymbol{\varphi_t^-}$ during feedback presentation across all participants (N = 63). Patients showed significantly reduced responses related to errors during loss compared to controls in the orbitofrontal cortex (OFC, green). Consistent effects across groups (positive effect in yellow, negative effect in blue) of outcome error were found in the dorsal anterior cingulate cortex (dACC) and the anterior insula. Details are reported in Table 3.3 and main effects in Table S3.4. p_{FWEc}<.05, p_{CDT} < .001. (B) Assessment of the effect of anhedonia within patients (n = 29). A negative relationship between magnitude-modulated loss prediction error $\boldsymbol{\delta_t^-}$-related actitivity and anhedonia scores was observed in a cluster in the medial thalamus / habenula and the posterior cingulate cortex (PCC). (C) A significant negative association between anhedonia scores and loss outcome error $\boldsymbol{\varphi_t^-}$-related activity was found in the posterior insula.

Table 3.4. Correlation with CDI and anhedonia-subscale scores within patients (n = 29).

	MNI coordinates [mm]			Significant activation		Peak
Brain region	x	y	z	p_{FWEc}	k	Z
Negative correlation CDI with φ_t^-						
L Insula	-45	0	-4	.027	107	4.80
L Postcentral gyrus	-67	-20	24	.023	111	4.45
Negative correlation between anhedonia CDI-subscale with φ_t^-						
L Precentral gyrus	-27	-14	54	.012	127	4.77
L Insula	-37	-10	-2	<.001	289	4.73
R Supramarginal gyrus	63	-28	32	.019	115	4.17
L Inferior parietal lobule	-51	-30	52	.015	104	3.90
Negative correlation CDI with δ_t^-						
L Fusiform gyrus	-35	-44	-22	.001	230	4.75
L Middle temporal gyrus	-55	-74	6	.007	166	4.45
R Inferior temporal gyrus	49	-58	-6	.045	112	4.39
L Fusiform gyrus	-35	-10	-32	.006	169	4.27
R Inferior parietal lobule	25	-48	50	.003	188	4.22
Negative correlation between anhedonia CDI-subscale with δ_t^-						
L Inferior temporal gyrus	-37	-12	-34	.001	214	5.60
R Postcentral gyrus	29	-32	36	<.001	866	4.94
R Fusiform gyrus	37	-38	-14	.005	175	4.81
R Thalamus MDm	9	-28	6	<.001	268	4.54
L Postcentral gyrus	-33	-36	56	<.001	1492	4.51
R Inferior temporal gyrus	49	-58	-6	<.001	272	4.43
L Middle temporal gyrus	-51	-78	2	<.001	389	4.31
L Cerebelum 4 5	-23	-38	-22	<.001	263	4.23
R Superior frontal gyrus, dorsolateral	31	-6	66	.001	238	4.02
R Cerebelum 6	21	-50	-24	.004	177	4.01
Positive correlation CDI with φ_t^-						
NS						
Positive correlation between anhedonia CDI-subscale with φ_t^-						
NS						
Positive correlation CDI with δ_t^-						
NS						
Positive correlation between anhedonia CDI-subscale with δ_t^-						
NS						

Significance level at whole-brain cluster-level p_{FWEc} < 0.05, cluster-defining threshold p_{CDT} < .001. Abbreviations: CDI, Child Depression Inventory; k, cluster size; R, right; L, left.

These results indicate that MDD is related to aberrant gain control in the OFC, specifically in high loss contexts. Moreover, the self-inhibition of the dACC was significantly lower across task conditions in the MDD group. The posterior mean of the group effect on the self-inhibition of the OFC was significantly related to the learning rate across all participants (Spearman's ρ = .295, p = .023, n = 59), but not in the dACC (ρ =-.135, p = .301, n = 59). A leave-one-out cross-validation using the loss-magnitude dependent difference in self-connection strength in the OFC explained a significant amount of the inter-subject variability between MDD and controls, showing an independent out-of-sample correlation of $r(58) = 0.38$, $p = 0.001$.

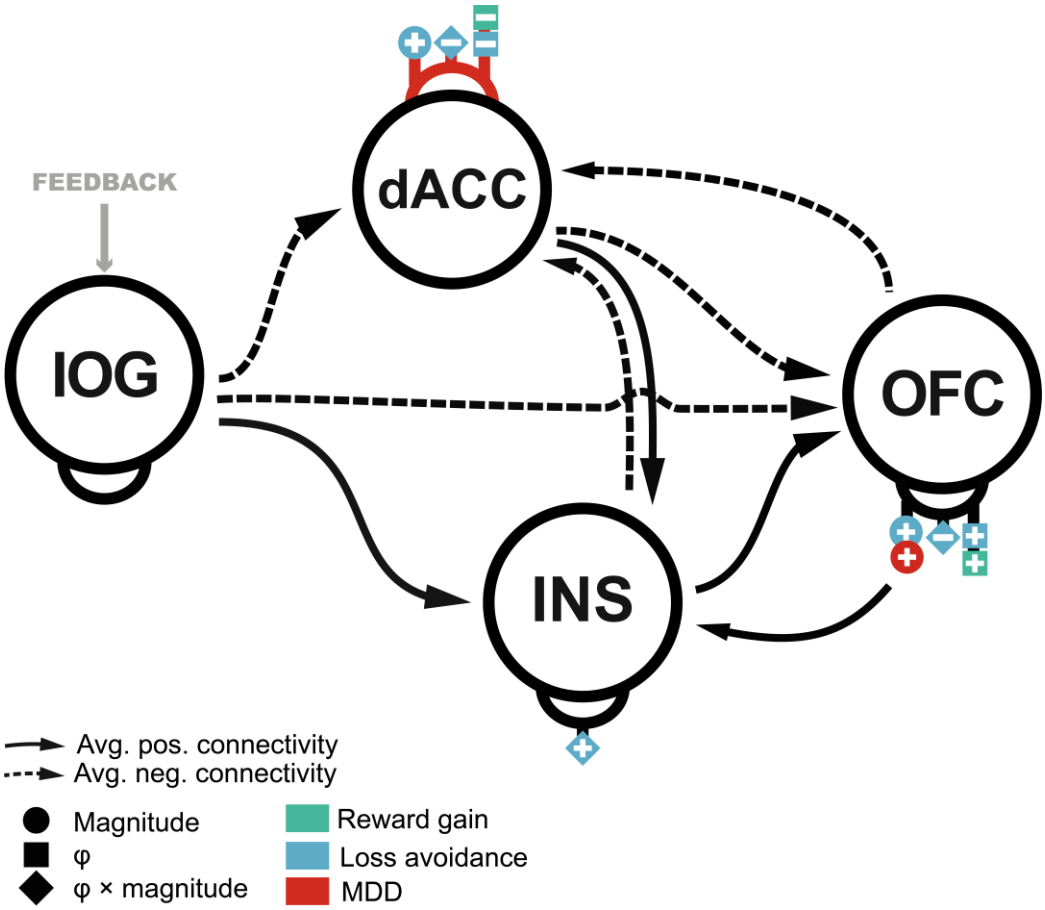

Figure 3.2. Effective connectivity during reward and loss feedback processing. There was a significant group effect of the factor loss magnitude on the self-inhibition parameter in the OFC, indicating aberrant input sensitivity during feedback in loss avoidance contexts in adolescent MDD. Cross-validation showed that this effective connectivity parameter was able to predict the group variable indicated by a significant out-of-sample correlation. Importantly, this parameter was associated with the learning rate of participants, which was lower in adolescents with MDD. The arrows reflect the posterior estimates of the second-level PEB model after Bayesian model reduction (Table 3.5). Self-connections are depicted as half-circle on each region. Solid lines indicate positive effective connectivity whereas dashed lines represent negative effective connectivity. Abbreviations: IOG, inferior occipital gyrus; dACC, dorsal anterior cingulate cortex; INS, anterior insula; MDD, major depressive disorder; OFC, orbitofrontal cortex; φ, outcome error.

Table 3.5. Average connectivity during feedback phase obtained by Bayesian model averaging of PEB model parameters

Connection type	Commonalities	PP Commonalities	Depression	PP Depression
Endogenous parameters				
OFC → Insula	0	0	0	0
OFC → dACC	-0.195	1	0.04	0.66
dACC → Insula	0.812	1	0	0
dACC → OFC	-0.345	1	0	0
Insula → OFC	0.284	1	0	0
Insula → dACC	-0.084	1	-0.049	0.85
IOG → Insula	0.144	1	0	0
IOG → dACC	-0.228	1	0.017	0.53
IOG → OFC	-0.108	1	-0.013	0.58
Self-inhibition parameters				
OFC → OFC	-0.467	1		
Insula → Insula	0.311	1		
dACC → dACC	-0.288	1	-0.139	1
IOG → IOG	1.740	1		
Modulatory parameters				
Insula → Insula, φ_t^+	0	0	0	0
Insula → Insula, M^+	0	0	0	0
Insula → Insula, $\varphi_t^+ \times \mathrm{M}^+$	0	0	0	0
Insula → Insula, φ_t^-	0	0	0	0
Insula → Insula, M^-	0	0	0	0
Insula → Insula, $\varphi_t^- \times \mathrm{M}^-$	-0.482	1	0	0
dACC → dACC, φ_t^+	1.506	1	0	0
dACC → dACC, M^+	-0.103	0.55	0	0
dACC → dACC, $\varphi_t^+ \times \mathrm{M}^+$	0	0	0.194	0.89
dACC → dACC, φ_t^-	1.986	1	0	0
dACC → dACC, M^-	-0.628	1	0	0
dACC → dACC, $\varphi_t^- \times \mathrm{M}^-$	-0.308	0.99	0	0
OFC → OFC, φ_t^+	1.506	1	0.351	0.90
OFC → OFC, M^+	0.148	0.54	0	0
OFC → OFC, $\varphi_t^+ \times \mathrm{M}^+$	0	0	-0.140	0.51
OFC → OFC, φ_t^-	-0.748	1	0	0
OFC → OFC, M^-	-0.495	1	-0.663	1
OFC → OFC, $\varphi_t^- \times \mathrm{M}^-$	0.781	1	0	0
Input parameter				
Feedback → IOG	4.287	1	-0.208	0.64

Between-region connections are in units of Hz. Self-connections, where the source and target are the same, are the log of scaling parameters that multiply up or down the default value −0.5Hz. *N* = 60. dACC, dorsal anterior cingulate cortex; INS, insula; IOG, inferior occipital gyrus; LPFC, lateral prefrontal cortex; M, magnitude; φ, error signal; THL, thalamus; VS, ventral striatum; PP. posterior probability.

3.4.3 Neural correlates of depression severity and anhedonia

Regression analysis of anhedonia scores within patients revealed significant associations in brain signaling of learning variables for loss. Particularly, we found that magnitude-modulated loss prediction error signaling δ_t^- was associated negatively with anhedonia scores in the medial thalamus/habenula, the posterior cingulate cortex, the postcentral gyrus, and the fusiform gyrus (Table S3.5, Figure 3.1B). The loss-related outcome error signal φ_t^- in the posterior insula was associated negatively with anhedonia scores (Table S3.5, Figure 3.1C). No within-patients associations were observed for reward-related signals.

3.4.4 Neural correlates of effects of SSRI within participants with MDD

We conducted additional analyses comparing effects of the learning parameters $\boldsymbol{\delta}$, $\boldsymbol{\varphi}$ and Q to see if there were systematic differences of brain activity related to SSRI medication. We compared groups of 18 medicated with SSRI to 10 other patients (unmedicated or no SSRI). However, GLM analyses of expected values and prediction error did not reveal any significant group differences. No clusters survived ($p_{\text{FWEc}} < .05$) a threshold of $p_{\text{CDT}} < .001$, nor a more lenient cluster-defining threshold of $p_{\text{CDT}} < .005$. Moreover, DCM parameter values of the OFC and the dACC were assessed for SSRI effects. To this end, we extracted the individual posterior means for each patient and split them into two groups (SSRI, other or no medication). However, we did not find a significant difference between the patients taking SSRIs and others in the OFC, $t(26) = -1.44$, $p = 0.162$, nor the dACC, $t(26) = -0.08$, $p = 0.936$. Note that one patient was excluded from this analysis as their medication history was not disclosed.

3.5 Discussion

In this study, we used a combination of computational modeling, fMRI and connectivity analysis to study reward and loss processing in adolescent MDD. We demonstrated that (1) adolescent MDD is associated with slower learning in the instrumental learning task and that this in turn is linked to (2) aberrant gain control in the OFC during feedback processing, (3) anhedonia-related reduction of representation of loss outcomes in the posterior insula and habenula, and (4) intact reward prediction error processing in the ventral striatum and the medial prefrontal cortex. Thus, the present work provides novel insights into the neurobiological foundation of altered learning mechanisms in loss avoidance in early onset MDD.

Our computational modeling approach revealed differences in behavioral adaptation that were reflected in the learning rate to update one's belief about prospective receipt of reward and loss. By testing a series of behavioral models, we showed that (1) participants with MDD adapted their instrumental vigor according to a simpler learning model with a single learning rate for reward and loss and (2) that they also updated the expected values slower than controls. While this does not indicate a different learning mechanism in MDD per se, it suggests that instrumental behavior does not rely on differential update rates

to adapt behavior for approaching reward and avoiding loss. The speed of learning depends on the perceived volatility of the environment (Behrens, Woolrich, Walton, & Rushworth, 2007), which has been found to be affected in participants suffering from anxiety (Browning, Behrens, Jocham, O'reilly, & Bishop, 2015), a highly prevalent comorbidity in depression (Häberling et al., 2019). Reduced learning and updating of one's belief system and hence an inability to update and hold an appropriate structure of possible aversive outcomes might provide the basis for a biased evaluation of the environment.

On the neural level, we linked these differences in value representation updates derived from the computational model to neural *feedback* processing. While in controls OFC activity was related to an outcome error signal across task conditions, this was absent in participants with MDD during loss processing. Effective connectivity analysis further showed that this effect was primarily driven by aberrant gain control in the OFC, specifically the sensitivity tuning in varying magnitude contexts. The gain in the OFC is modulated by various neurotransmitters (Robbins & Arnsten, 2009) but there is evidence that dopamine plays a role in learning approach and avoidance behavior (Palminteri et al., 2012) supported by dense dopaminergic projections to the OFC (Kahnt & Tobler, 2017). This gain control might be critical to modulate the activity of the OFC during error processing in reward (Ramnani, Elliott, Athwal, & Passingham, 2004) and loss (S. F. Taylor et al., 2006) contexts, updating neural representations of value (Sul, Kim, Huh, Lee, & Jung, 2010) and maintaining a representation of the task structure (Wilson, Takahashi, Schoenbaum, & Niv, 2014). The latter entails updating an estimate of certainty for a specific outcome that seems to fail in adolescent MDD when evaluating unexpected outcomes in avoidance learning. Strikingly, we found a significant negative association between the learning rate and the gain control of the OFC during loss processing.

In adolescents at familial risk for MDD (Jin et al., 2017) impaired functional connectivity between the posterior insula and the OFC during loss processing in adolescents was predictive for future depressive symptomatology after nine months. Concordantly, in the present study decreased loss error signaling was related to anhedonia in the posterior insula. While the encoding of magnitude-modulated loss PE signals did not significantly differ between groups, a within-patients analysis revealed that BOLD responses related to loss PE in a cluster comprising the medial thalamus and habenula were significantly negatively associated with anhedonia. Previous work has implicated impaired habenula function and morphology in depression and anhedonia (Kumar et al., 2018; Lawson et al., 2017). Thus, disrupted loss processing could reflect an important factor that contributes to increased susceptibility to adolescent MDD.

Recent computational accounts on depression suggest that it is related to an aberrant cognitive prior that underlies negative bias in evaluation of the state of the environment (Clark, Watson, & Friston, 2018). An overgeneralization of one's own states might eventually lead to helplessness behavior, where negative outcomes are associated with poor performance and failure of oneself, and positive outcomes are regarded as mere random

events. Our results could indicate that a negative prior about the outcome is not updated due to dysfunction in the OFC, and this might contribute to maintaining a negative bias and a feeling of loss of control over outcomes. The latter is consistent with the differences of negative arousal derived from the postscan ratings, where participants with MDD expressed more relief (i.e. more deactivation) in rewarding outcomes and more fear (i.e. more activation) during loss outcomes. However, unlike in adult depression (Steele, Kumar, & Ebmeier, 2007), computational modeling did not indicate that this higher range of negative arousal significantly affected response vigor in MDD.

Contrary to our hypothesis, we did not find any evidence of impact of depression on PE processing in rewarding contexts as previous studies in adult MDD (Gradin et al., 2011; Kumar et al., 2018; Kumar et al., 2008). Here, we postulate that two factors could have led to this null finding. First, in our MID task participants did not have to learn anything to perform well. This design was employed to minimize confounds of (a) brain maturation and development within participant groups (Nussenbaum & Hartley, 2019) and (b) diagnosis (Snyder, 2013) on learning performance, which could be difficult to disentangle in more complex learning paradigms. Nevertheless, our results are in concordance with previous findings of intact reward PE signaling in a non-learning task in adult MDD (Rutledge et al., 2017). Second, there is evidence that impairments of reward PE signaling are related to the number of depressive episodes across life-time (Kumar et al., 2018). This might explain the results in participants with an early onset as in our study and could indicate that previous reports of impaired reward PE signaling errors are related to the chronicity of the disorder.

It has to be considered that the majority of participants with depression were receiving antidepressive medication (Table 3.1). It is possible that intake of selective serotonin reuptake inhibitors (SSRIs) might have affected error signaling and learning of cue-outcome associations (McCabe, Mishor, Cowen, & Harmer, 2010). Based on this assumption, we would expect blunting of reward responses due to the administration of SSRIs. However, additional control analyses of brain activity and connectivity comparing participants with MDD with (n = 18) and without (n = 10) SSRI-intake revealed no significant effect. Although we cannot fully rule out that medication had an effect based on these rather small subsamples, we consider it unlikely that this was the case in this study.

In conclusion, this is the first study to show that adolescent MDD is associated with specific impairments of error processing in loss avoidance contexts, whereas reward sensitivity is intact. Given the critical role of evaluating an action that led to an unexpected aversive outcome, this deficit could be directly related to severe difficulties in decision making and in social life and by contributing to the development and persistence of a negative bias in depression. Our study provides a first important step towards identifying computational mechanisms in adolescent MDD and paves the way for establishing computational assays (Stephan & Mathys, 2014) that will facilitate the translation into clinical practice.

3.6 Supplementary information

3.6.1 Behavioral raw data analysis

A group-by-condition ANOVA for mean logRTs did not reveal a significant effect of condition, $F(4, 305) = 0.95$, $p = 0.434$, nor the interaction term, $F(4, 305) = 0.23$, $p = 0.921$, but a trend in the factor group, $F(1, 305) = 3.532$, $p = 0.061$. The number of response omissions did not differ between groups (HC: $M = 5.4$. $SD = 2.9$; MDD: $M = 6.4$, $SD = 5.1$; $U = 460$, $p = 0.633$), yielding comparable hit rates across conditions (Table S3.1). This suggests that the MID task was well balanced for both groups, showing no behavioral group differences in terms of reaction times and response omissions.

Table S3.1. Behavioral data of the monetary incentive delay task. Experimental manipulation ensured that a hit rate of around 66% was achieved.

	High loss		Low loss		Neutral		Low Gain		High Gain	
	HC	MDD	HC	MDD	HC	MDD	HC	MDD	HC	MDD
Hit rate	61 (9)%	64 (10)%	58 (12)%	57 (12)%	-	-	60 (11)%	63 (13)%	64 (10)%	62 (12)%
Response time	268 (11) ms	259 (10) ms	276 (16) ms	270 (11) ms	269 (10) ms	265 (10) ms	270 (10) ms	261 (12) ms	265 (9) ms	254 (8) ms
Arousal rating	59.4 (30.8)	65.3 (30.7)	46.1 (28.5)	58.8 (25.2)	33.7 (24.1)	29.2 (21.5)	51.6 (22.4)	39.8 (26.6)	64.9 (27.6)	56.8 (33.8)
Valence rating	13.1 (20.3)	5.7 (9.7)	27.4 (23.3)	20.3 (23.4)	44.4 (19.4)	47.4 (21.9)	67.7 (17.6)	71.0 (19.3)	86.8 (21.1)	91.6 (12.3)
Negative arousal	30.4 (18.7)	40.1 (17.5)	10.9 (15.7)	25.2 (14.9)	-9.9 (15.2)	-14.8 (18.9)	-13.7 (14.3)	-24.0 (11.6)	-17.8 (15.4)	-26.6 (15.1)
Positive arousal	-18.8 (16.9)	-18.5 (16.9)	-18.0 (18.1)	-12.8 (23.9)	-14.8 (18.1)	-14.6 (18.3)	14.4 (12.9)	9.6 (18.0)	37.2 (21.6)	36.2 (20.5)

All values are means (SD). Rating range was 0-100. Negative and positive arousal values are ranging from -113 to 113 and are constrained to a maximum within-subject range of 141 for extreme values (but are typically much lower).

3.6.2 Control analysis and model simulation

Posterior means of the behavioral model showed only a moderate correlation among each other (all $|r| < 0.53$ (Figure S3.1). We simulated behavioral data from the model parameters obtained for each participant and repeated the raw data analysis of synthetic logRTs. Note that in the raw data analysis we found a trend of faster response in the patients group. This trend was replicated in the analysis of variance with the simulated data: main effect of group, $F(1,305) = 3.145$, $p = .078$, main effect of condition, $F(4,305) = 0.980$, $p = .418$, group-by-condition interaction, $F(4,305) = 0.251$, $p = .909$. In addition, observed and simulated mean logRTs showed a strong correlation, Spearman's $\rho = 0.828$, $p < 10^{-15}$. Hence, this suggests that the simulated data was comparable to the empirical data in both groups.

Table S3.2. Parameter prior means (variance) of the reinforcement learning model and the response models.

Reinforcement learning model	**rw1**			**rw2**	
v_0 (logit-space)	0.66 (0.5)			0.66 (0.5)	
α (logit-space)	0.05 (1)			-	
α^- (logit-space)	-			0.05 (1)	
α^+ (logit-space)	-			0.05 (1)	
Response models	**M1**	**M2**	**M3**	**M4**	**M5**
β_0	0 (4)	0 (4)	0 (4)	0 (4)	0 (4)
β_1	0 (4)	0 (4)	0 (4)	0 (4)	0 (4)
β_2	0 (4)	0 (4)	0 (4)	0 (4)	0 (4)
β_3	0 (4)	0 (4)	0 (4)	0 (4)	0 (4)
β_4	0 (4)	0 (4)	0 (4)	0 (4)	0 (4)
β_5	0 (4)	-	-	0 (4)	-
β_6	0 (4)	-	-	-	-
β_7	0 (4)	-	-	-	-
ξ	log(3) (log(2))	log(3) (log(2))	log(3) (log(2))	log(3) (log(2))	log(3) (log(2))

Table S3.3. Averaged parameter estimates and parameter comparison between groups

	HC	MDD	Statistic	*p* value
α	0.052 (0.012)	0.046 (0.009)	t (60) = 2.04[a]	0.046
β_0	-1.356 (0.156)	-1.362 (0.096)	t (61) = 0.184	0.854
β_1	-0.006 (0.030)	-0.023 (0.029)	U = 608[b]	0.122
β_2	0.004 (0.026)	0.015 (0.030)	t (61) = -1.490	0.141
β_3	0.026 (0.050)	0.005 (0.054)	t (61) = 1.638	0.106
β_4	-0.015 (0.074)	0.001 (0.061)	t (61) = -0.903	0.370

Mean (SD) within each group.

[a]We removed one outlier in the MDD group, determined by visual inspection and a significant Grubb's test

[b]Mann-Whitney *U*-tests were used for group comparison, because Shapiro-Wilk tests were significant in the MDD group.

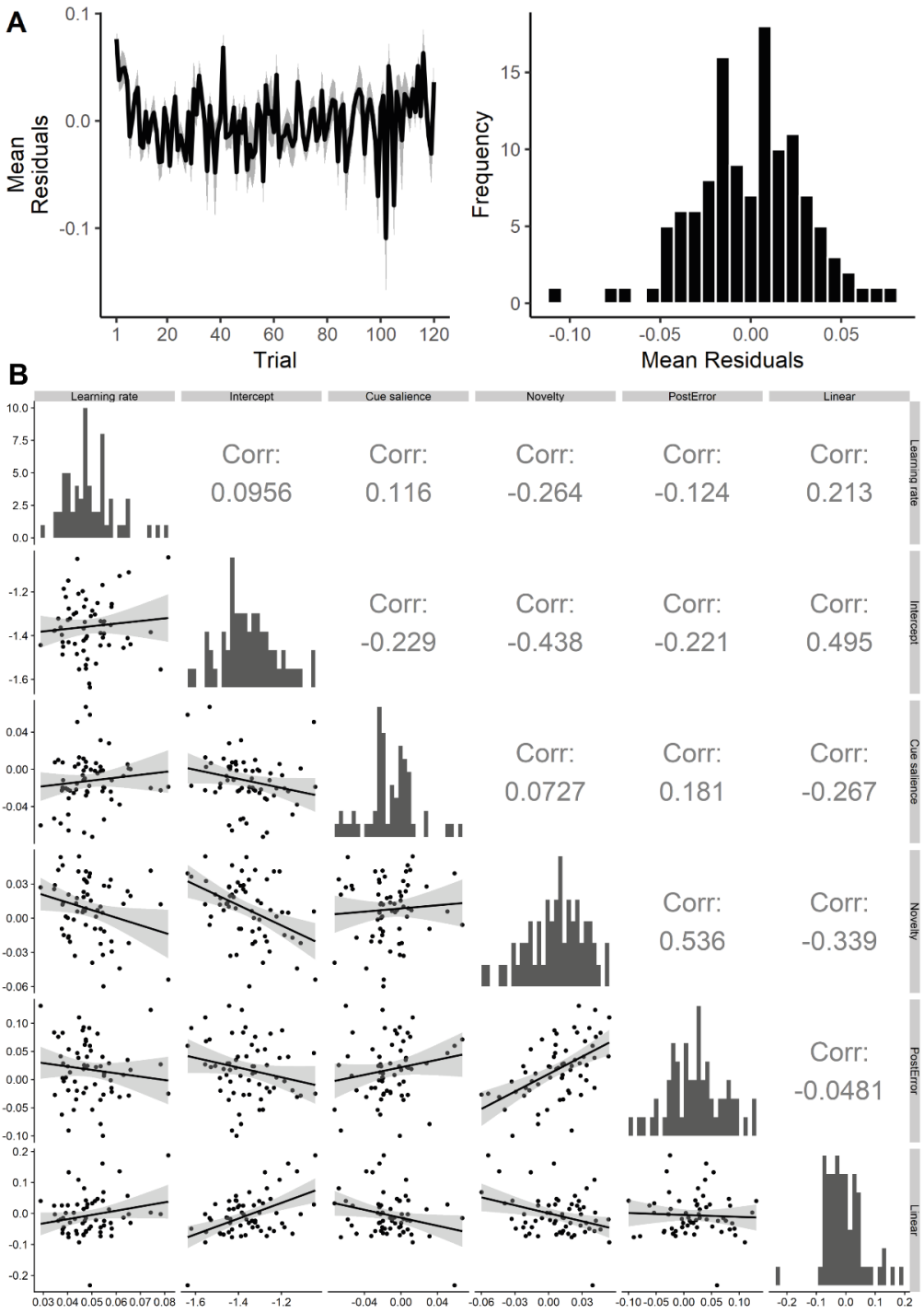

Figure S3.1. Assessment of the best-fitting behavioral model. (A) The distribution of the model residuals suggested that the model was able to capture the response patterns in the empirical data across participants. Shaded area indicates the SEM. (B) Intercorrelations between the parameters of the winning behavioral model. We found only small to moderate correlations between the parameters of the behavioral model. This suggests that the effects predicting the response vigour could be disentangled well in the model. Corr values are Pearson correlation coefficients. Data points represent the beta values. *n* = 62.

3.6.3 Increased negative arousal of HC and MDD in post-scan ratings

The analysis of ratings of negative arousal revealed a significant main effect of condition $F(4, 295) = 156.04$, $p < 10^{-15}$, and a group-by-condition interaction, $F(4, 295) = 7.73$, $p < 10^{-5}$, specifically patients showed higher negative arousal in loss (-4CHF: $p = .018$; -1 CHF: $p < .001$) and lower negative arousal in reward (+4CHF: $p = .031$; +1CHF: $p = .011$) than controls (Figure S3.2). Positive arousal ratings were comparable between groups and across conditions, group-by-condition: $F(4,295) = 0.56$, $p = .693$.

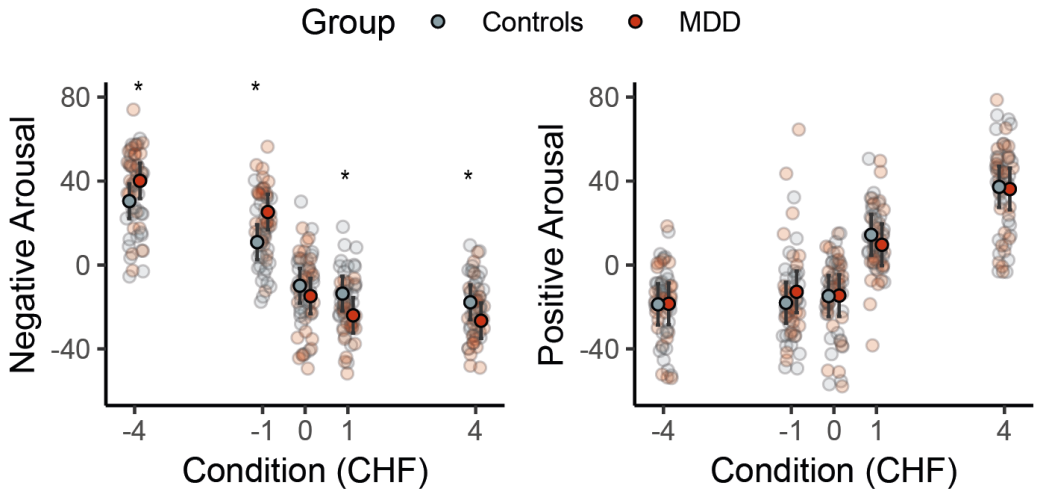

Figure S3.2. Subjective negative and positive arousal of outcomes. Post-scan ratings of subjective liking and arousal for each trial type was centered and rotated to obtain estimates of negative and positive arousal. Patients differed significantly on the negative arousal scale, by rating loss higher and reward lower than controls. No group difference was observed in the positive arousal scale.

3.6.4 Neural correlates of outcome error (φ) processing: main effects

We located increasing activity encoding of reward outcome errors ($\boldsymbol{\varphi_t^+}$) in the putamen, caudate, orbitofrontal cortex (OFC), lateral prefrontal cortex (PFC), temporal lobe, whereas activity in the insula, dorsal anterior cingulate gyrus (dACC), and ventrolateral PFC decreased (Table S3.4).

When processing loss (i.e. negative $\boldsymbol{\varphi_t^-}$), $\boldsymbol{\varphi_t^-}$ was negatively associated with activity in the anterior insula, dorsal ACC, ventrolateral PFC and the supramarginal gyrus. Avoiding loss (i.e. positive $\boldsymbol{\varphi_t^-}$) was associated with clusters in the caudate and putamen, dorsolateral PFC, superior temporal gyrus, paracentral lobule, lingual gyrus and occipital lobe (Table S3.4, Figure 3.1).

Table S3.4. Results of fMRI analyses: main effects of φ_t^+ and φ_t^- during feedback processing (N = 63).

	MNI coordinates [mm]			Significant activation		Peak
Brain region	x	y	z	p_{FWEc}	k	Z
Positive effect of φ_t^+						
R Putamen	21	10	-10	<.001	18272	6.52
L Medial orbital frontal gyrus	-7	50	-12			6.28
R Caudate	19	14	18			5.26
L Putamen	-15	18	-4	<.001	3072	6.35
L Superior frontal gyrus, dorsolateral	-25	24	46	<.001	1499	5.68
L Inferior Temporal gyrus	-47	-6	-26	<.001	494	5.14
L Middle occipital gyrus	-39	-74	36	<.001	594	5.07
L Inferior temporal gyrus	-57	-46	-16	.03	120	4.42
L Superior temporal gyrus	-63	-16	4	.01	150	4.24
Negative effect of φ_t^+						
R Inferior frontal gyrus, triangular	49	18	2	<.001	1970	6.54
R Insula	35	22	-8			6.42
R Inferior frontal gyrus, orbital	45	44	-8			6.42
R Superior frontal gyrus, medial	3	34	46	<.001	1045	6.07
R Superior frontal gyrus, medial	3	40	40			4.91
R Middle cingulate	3	34	46			4.85
L Insula	-29	26	-10	.001	215	4.77
Positive effect of φ_t^-						
R Caudate	21	8	16	<.001	506	5.87
L Paracentral lobule	-13	-28	60	<.001	386	4.89
R Calcarine	9	-84	2	<.001	367	4.65
L Caudate	-23	-20	26	<.001	283	4.50
R Superior temporal gyrus	53	-12	0	<.001	265	4.49
L Middle occipital gyrus	-37	-68	32	.019	132	4.41
R Putamen	25	14	-2	.016	137	4.35
L Fusiform	-33	-40	-20	.013	142	4.20
L Middle occipital gyrus	-45	-66	0	.038	113	4.06
L Superior frontal gyrus, dorsolateral	-15	42	48	.001	231	3.96
Negative effect of φ_t^-						
R Insula	35	16	-14	<.001	425	5.35
L Supramarginal gyrus	-69	-36	32	.011	148	4.49
L Insula	-29	20	-18	.002	205	4.41
R Superior anterior cingulate	3	22	20	<.001	446	4.25

Significance level at whole-brain cluster-level $p_{FWEc} < 0.05$, cluster-defining threshold $p_{CDT} <$ .001. Abbreviations: k, cluster size; R, right; L, left.

3.6.5 Neural correlates of prediction error (δ) processing: main effects

The difference between reward magnitude and expected value (δ_t^+) during the feedback phase was positively associated with BOLD changes in the ventral and dorsal striatum, the ventromedial PFC, OFC, postcentral gyrus, temporal lobes, and the occipital lobe for increasing δ_t^+. A network containing the anterior insula, dorsomedial PFC, and ventrolateral PFC was negatively associated with δ_t^+ (Table S3.5). During loss processing, δ_t^- was positively associated with activation in the caudate, putamen, middle temporal, superior frontal, middle frontal, and postcentral cortex and superior parietal lobe. In addition, decreasing $\boldsymbol{\delta_t^-}$ was associated with higher activation within the dorsal ACC, dorsomedial PFC, anterior insula, middle temporal gyrus, ventrolateral PFC, supramarginal gyrus, and midbrain (Table S3.5). We did not find any differences between patients and controls.

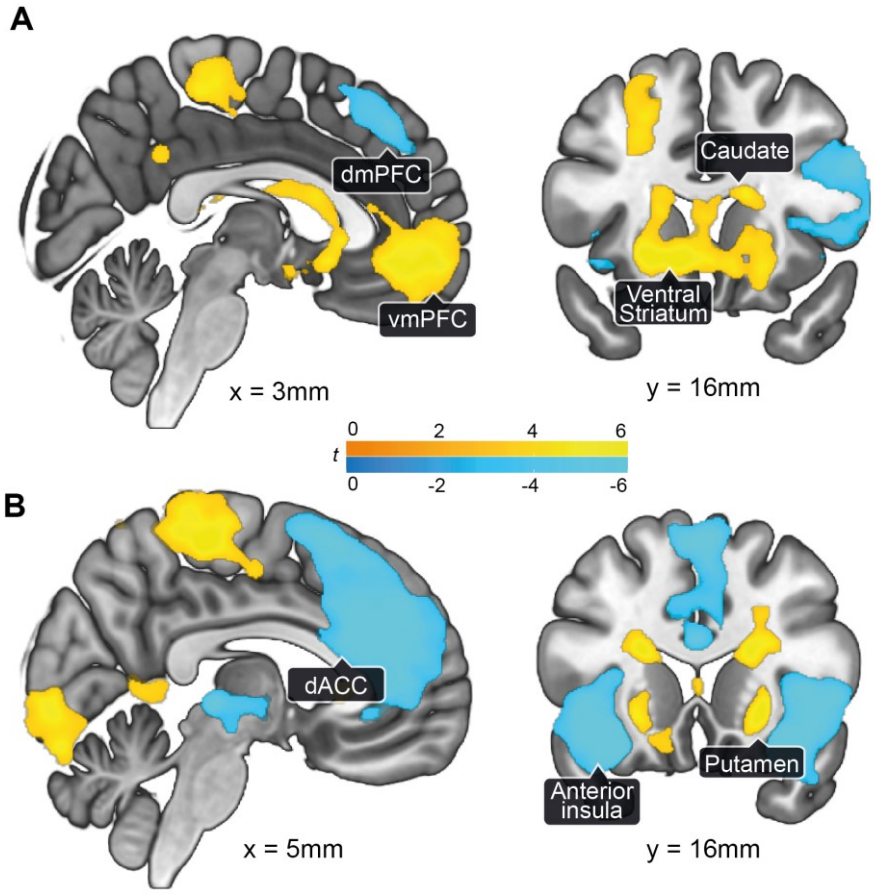

Figure S3.3. Activity associated with magnitude-related PE across groups. (A) The network encoding increasing (yellow) magnitude-related PE $\boldsymbol{\delta_t^+}$ was related to activity in the ventral striatum, caudate, and the ventromedial prefrontal cortext (vmPFC). Decreasing (blue) $\boldsymbol{\delta_t^+}$ modulated activity in the lateral prefrontal cortex and the dorsomedial PFC (dmPFPC). (B) Activity in the putamen and caudate was related to increasing (yellow) $\boldsymbol{\delta_t^-}$, whereas activity in the insula and the medial prefrontal cortex, especially the dorsal ACC, was associated with decreasing (blue) $\boldsymbol{\delta_t^-}$. $p_{FWEc} < .05$, $p_{CDT} < .001$, $N = 63$.

Table S3.5. Results of fMRI analyses: main effects of magnitude-modulated prediction errors δ_t^+ and δ_t^- during feedback processing (N = 63).

	MNI coordinates [mm]			Significant activation		Peak
Brain region	x	y	z	p_{FWEc}	k	Z
Positive effect of δ_t^+						
R Ventral striatum	9	10	-14	<.001	9071	6.67
L Medial orbital frontal gyrus	-5	48	-14			6.65
L Superior frontal gyrus, dorsolateral	-17	34	46	<.001	1708	6.50
R Inferior occipital gyrus	27	-90	-2	<.001	1392	5.92
L Postcentral gyrus	-17	-30	60	<.001	2783	5.46
L Middle occipital gyrus	-19	-92	4	<.001	320	4.92
L Middle temporal gyrus	-57	-10	-22	<.001	260	4.79
R Middle temporal gyrus	57	-10	-20	.014	138	4.69
L Posterior cingulate	-1	-46	32	<.001	218	4.39
L Middle occipital gyrus	-43	-74	32	.013	140	4.32
Negative effect of δ_t^+						
R Inferior frontal gyrus, triangular	55	20	28	<.001	1633	5.83
R Insula	33	24	2			6.48
L Superior frontal gyrus, medial	3	40	42	<.001	424	4.76
L Insula	-29	22	8	.001	228	4.31
L Inferior frontal gyrus, orbital	49	44	-4	.002	200	4.27
Positive effect of δ_t^-						
L Caudate	-19	2	20	<.001	24740	7.18
R Caudate	19	-8	24			6.96
L Putamen	-13	12	-14	<.001	340	4.92
L Middle temporal gyrus	-57	-32	6	<.001	396	4.30
L Superior frontal gyrus, dorsolateral	-19	26	52	.005	189	4.20
L Middle frontal gyrus	-39	46	-18	.002	230	4.18
Negative effect of δ_t^-						
L Insula	-31	18	-18	<.001	1559	>8
R Insula	33	18	-16	<.001	2297	7.63
L Superior frontal gyrus, medial	7	32	26	<.001	4963	6.64
L Red nucleus	-3	-24	26	<.001	538	6.59
R Supramarginal gyrus	59	-46	46	<.001	761	5.60
L Supramarginal gyrus	-65	-46	38	<.001	380	5.29
R Middle temporal gyrus	53	-32	-6	<.001	343	5.17
R Superior frontal gyrus, dorsolateral	19	54	24	.013	157	3.99

Significance level at whole-brain cluster-level p_{FWEc} < 0.05, cluster-defining threshold p_{CDT} < .001. Abbreviations: k, cluster size; R, right; L, left.

3.6.6 Neural correlates of expected value (Q) processing: main effects

We analyzed the main effect of expected value during the anticipation phase of the task, and found a network that was positively associated with increasing Q^+ comprising the ventromedial PFC, lateral PFC, ventral striatum, anterior insula, midbrain (Table S3.6, Figure S3.4). No cluster was found that was negatively associated with Q^+. Furthermore, we did not find any cluster that was associated during processing the expected loss value Q^- across participants. No differences between patients and controls were observed.

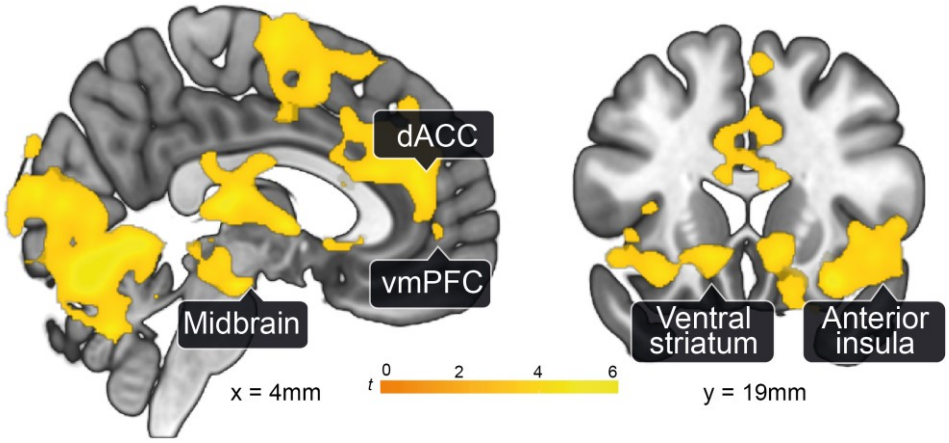

Figure S3.4. Activity associated with expected reward values Q^+ across groups. Across both groups, an increase in expected reward values Q^+ was associated with increased activation in the midbrain, the ventral striatum, the anterior insula, the dorsal anterior cingulate cortex (dACC), and the ventromedial prefrontal cortex (vmPFC). $p_{FWEc} < .05$, $p_{CDT} < .001$, $N = 63$.

Table S3.6. Expected value (Q)-signaling across both groups.

	MNI coordinates [mm]			Significant activation		Peak
Brain region	x	y	z	p_{FWEc}	k	Z
Positive effect of Q^+						
R Insula	39	28	-4	< .001	21473	6.07
Raphe	-5	-30	-14			5.87
Vermis 6	7	-66	-10			5.66
R Insula	29	18	-16			5.24
Unknown	7	-30	-14			5.21
R Pallidum	11	4	-4			5.21
R Thalamus PuM	3	-24	10			5.20
L Pregenual anterior cingulate	1	42	20			5.02
R Supramarginal gyrus	53	-42	30	.003	198	3.90
L Postcentral gyrus	-49	-14	50	.005	183	3.90
R Precentral gyrus	39	-10	48	.002	209	3.86
Positive effect of Q^-						
NS						

Significance level at whole-brain cluster-level $p_{FWEc} < 0.05$, cluster-defining threshold $p_{CDT} < .001$. Abbreviations:k, cluster size; R, right; L, left.

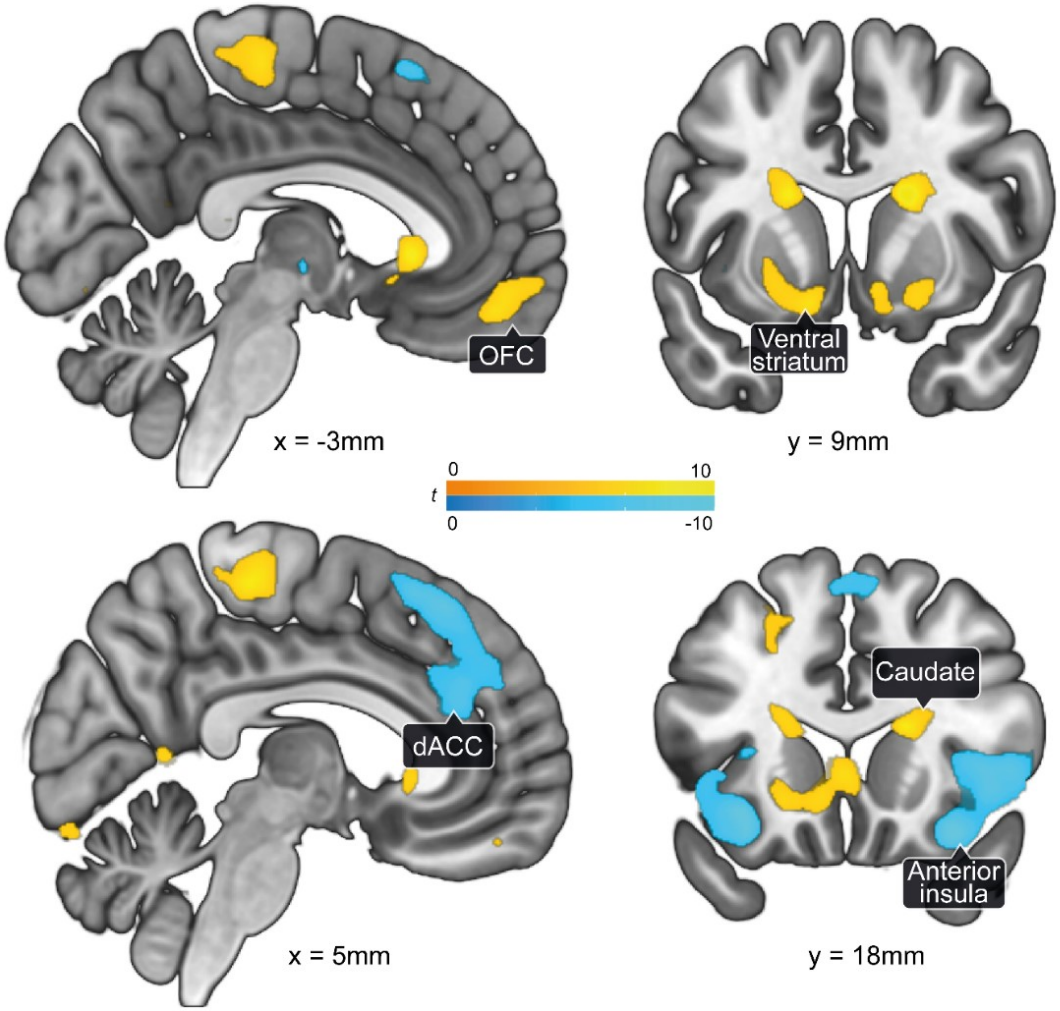

Figure S3.5. Activity associated with hits and misses across groups. Distinct brain activation in networks of hit (reward and loss avoidance, yellow) and miss (reward omission and loss, blue). $p_{FWE} < .05$, $N = 63$.

Table S3.7. Task activation across groups (n = 60).

	MNI coordinates [mm]			Significant activation		Peak
Brain region	x	y	z	p_{FWEc}	k	Z
Hit-miss						
R Caudate	19	8	20	< .001	2752	7.75
L Caudate	-19	-10	24	< .001	714	7.66
L Putamen	-21	12	-8	< .001	791	7.15
L Paracentral lobule	-7	-36	62	< .001	1951	6.96
L Cuneus	-23	-52	20	< .001	167	6.63
L Calcarine	-21	-86	6	< .001	348	6.59
R Hippocampus	43	-22	-18	< .001	82	6.43
R Superior frontal gyrus, dorsolateral	-21	32	46	< .001	369	6.31
L Medial orbital frontal gyrus	-7	46	-16	< .001	250	6.24
R Superior temporal gyrus	67	-8	-2	< .001	186	5.95
L Precuneus	-11	-54	8	< .001	123	5.90
R Calcarine	15	-12	46	< .001	51	5.72
Miss-hit						
R Superior frontal gyrus, medial	5	28	54	< .001	784	7.49
R Superior frontal gyrus, medial	7	36	30	< .001		6.87
R Insula	35	18	-18	< .001	1195	7.43
L Insula	-31	20	-14	< .001	579	7.38
R Middle temporal gyrus	57	-30	-8	< .001	56	6.21
R Supramarginal gyrus	63	-48	36	< .001	67	5.40

Significant clusters on whole-brain level in the second-level analyses for the contrasts reward-miss and miss-reward. This analysis aimed at revealing the task-relevant network across groups and guide the selection of ROIs for further analysis, thus we included the CDI score as covariate.
Significance level at whole-brain voxel-level p_{FWE} < 0.05, k > 50.
Abbreviations: CDI, Child Depression Inventory; k, cluster size; R, right; L, left.

Table S3.8. Positive effect of all feedback events.

	MNI coordinates [mm]			Significant activation		Peak
Brain region	x	y	z	p_{FWEc}	k	Z
R Occipital inferior gyrus	29	-80	-16	< .001	26989	> 8
R Superior frontal gyrus, medial	5	28	44	< .001	1931	> 8
L Hippocampus	-25	-28	-6	< .001	139	7.47
L Inferior temporal gyrus, orbital	-37	20	-6	< .001	402	7.36
L Inferior parietal lobule	-49	-44	54	< .001	1553	7.14
Vermis 4 5	1	-36	-2	< .001	142	6.73
R Caudate	9	24	0	< .001	145	6.66
L Middle frontal gyrus	-39	60	-4	< .001	1071	6.43
R Cerebellum 10	23	-40	-44	< .001	115	6.34
L Postcentral	-65	-10	30	< .001	133	5.91

This contrast was used to identify activity in the visual areas for the DCM analysis.
Significance level at whole-brain whole-brain p_{FWE} < 0.05, minimum cluster size k > 100.
Abbreviations: DCM, dynamic causal modeling; k, cluster size; R, right; L, left.

4 Valence-dependent coupling of prefrontal-amygdala effective connectivity during facial affect processing

David Willinger[a,c], Iliana I. Karipidis[a,c,d], Selina Beltrani[a], Sarah V. Di Pietro[a,c], Ronald Sladky[b,e], Susanne Walitza[a,c], Philipp Stämpfli[a,b], Silvia Brem[a,c]

4.1 Overview

Despite the importance of the prefrontal-amygdala network for emotion processing, valence-dependent coupling within this network remains elusive. In this study, we assessed the effect of emotional valence on brain activity and effective connectivity. We tested which functional pathways within the prefrontal-amygdala network are specifically engaged during the processing of emotional valence. Thirty-three healthy adults were examined with functional magnetic resonance imaging while performing a dynamic faces and dynamic shapes matching task. The valence of the facial expressions varied systematically between positive, negative, and neutral across the task. Functional contrasts determined core areas of the emotion processing circuitry, comprising the medial prefrontal cortex (MPFC), the right lateral prefrontal cortex, the amygdala and the right fusiform face area. Dynamic causal modeling demonstrated that the bidirectional coupling within the prefrontal-amygdala circuitry is modulated by emotional valence. Additionally, Bayesian model averaging showed significant bottom-up connectivity from the amygdala to the MPFC during negative and neutral, but not positive, valence. Thus, our study provides strong evidence for alterations of bottom-up coupling within the prefrontal-amygdala network as a function of emotional valence. Thereby our results not only advance the understanding of the human prefrontal-amygdala circuitry in varying valence context, but, moreover, provide a model to examine mechanisms of valence-sensitive emotional dysregulation in neuropsychiatric disorders.

4.2 Significance statement

Recent neuroimaging studies have emphasized the importance of valence-sensitivity within the prefrontal-amygdala network during emotion processing. Yet, it remains elusive which specific pathways are involved in processing affective information, and how this information is integrated in the brain's network. In particular, the amygdala's role

[a] Department of Child and Adolescent Psychiatry and Psychotherapy, University Hospital of Psychiatry Zurich, University of Zurich, Switzerland
[b] Department of Psychiatry, Psychotherapy and Psychosomatics, University Hospital of Psychiatry Zurich, University of Zurich, Zurich, Switzerland
[c] Neuroscience Center Zurich, University of Zurich and ETH Zurich, Switzerland
[d] Center for Interdisciplinary Brain Sciences Research, Stanford University School of Medicine, Stanford, CA, USA
[e] Social, Cognitive and Affective Neuroscience Unit, Department of Basic Psychological Research and Research Methods, Faculty of Psychology, University of Vienna, Vienna, Austria

The study was funded by the Department of Children and Adolescent Psychiatry and Psychotherapy, University Hospital of Psychiatry Zurich, University of Zurich. We thank Nathalie Holz for her suggestions on the task design.

A similar version of this manuscript has been published:
Willinger, D., Karipidis, I. I., Beltrani, S., Di Pietro, S. V., Sladky, R., Walitza, S., Stämpfli, P., & Brem, S. (2019). Valence-dependent coupling of prefrontal-amygdala effective connectivity during facial affect processing. *Eneuro*, *6*(4).

in signaling valence information to the cortex is subject to ongoing discussions. Moreover, as aberrant brain function has been found in the amygdala and the prefrontal cortex in various debilitating psychiatric disorders, understanding the mechanisms of processing emotional stimuli with different valence (positive, negative, neutral) is particularly relevant for the field. Our findings indicate changes in coupling strength as a function of emotional valence within the prefrontal-amygdala network.

4.3 Introduction

The prefrontal-amygdala network plays a pivotal role in adapting human behavior to constantly changing environmental demands. Previous neuroimaging research has emphasized the importance of interactions between the prefrontal cortex and the amygdala (AMY) during affective processing (Ochsner et al., 2009; Phillips, Ladouceur, & Drevets, 2008) and has tried to disentangle bottom-up from top-down mechanisms of emotion processes (Comte et al., 2016; Ochsner et al., 2009; Pessoa, 2017; Whalen et al., 2013). Emotional salience related to the perceptual properties of a stimulus, as mediated by emotional faces, is thought to be propagated from the amygdala to the prefrontal cortex via *bottom-up* connections (McRae, Misra, Prasad, Pereira, & Gross, 2011). It has long been recognized that the amygdala plays a crucial role in immediate, automatic processing of emotional information and the modulation of attention (A. K. Anderson, Christoff, Panitz, De Rosa, & Gabrieli, 2003; Ochsner et al., 2009; Phelps, 2006). Conversely, *top-down* signaling during emotion processing has been attributed to different forms of emotion regulation, where the lateral prefrontal cortex (LPFC) supports top-down evaluation of contextual significance and altering of the affective response by exerting cognitive control over limbic regions (Dima, Stephan, Roiser, Friston, & Frangou, 2011; Ochsner & Gross, 2005; Quirk & Beer, 2006), even without explicit instruction (Drabant, McRae, Manuck, Hariri, & Gross, 2009).

This coupling between the LPFC and the amygdala is central to theoretical models of emotion processing. Nevertheless, emotion processing involves complex interactions between amygdala driven bottom-up salience processing, and top-down contextualization and evaluation of stimuli, supported by the LPFC, whose strength and directions can differ substantially depending on context, e.g. emotional valence or task demands (H. Kim et al., 2004; Pessoa, 2017). Lately, this has led to new conceptions, where emotion processing is strongly interwoven with other mental entities that constitute cognition (e.g. memory or attention), and relies on dynamic, context-sensitive interactions of top-down and bottom-up processes (Pessoa, 2017).

Given that structural connections between the LPFC and the amygdala are sparse (Ray & Zald, 2012), regulatory signals from the LPFC are likely mediated to the amygdala via the medial prefrontal cortex (MPFC). The MPFC is situated perfectly to pass on top-down appraisal and regulation signals to limbic structures as it shares rich bidirectional connections with the LPFC and the amygdala (Price, 2005; Ray & Zald, 2012). As such, the MPFC has not only been implicated in regulation of emotional responses, in particular to

aversive stimuli, but also in integrating affective and contextual information, i.e. bottom-up and top-down signals, to support generation of affective meaning (Comte et al., 2016; Delgado, Nearing, LeDoux, & Phelps, 2008; Etkin, Buchel, & Gross, 2015; Lindquist, Satpute, Wager, Weber, & Barrett, 2015; Ochsner et al., 2009; Roy, Shohamy, & Wager, 2012; Silvers, Shu, Hubbard, Weber, & Ochsner, 2015). However, the valence-dependent coupling between regions comprising the emotion processing circuitry is only poorly understood. Particularly, the role of the amygdala in encoding valence is still debated (Ball et al., 2009; Jin, Zelano, Gottfried, & Mohanty, 2015), and thus far, valence-dependent alterations of directed coupling between the amygdala and the prefrontal cortex during emotion processing has not been investigated, despite it being strongly implicated in psychopathology (Dichter, Felder, & Smoski, 2009; Disner et al., 2011; Sladky, Hoflich, et al., 2015).

In this study, we used a novel dynamic face- and shape-matching task to investigate the effect of valence of facial expressions on effective connectivity within the prefrontal-amygdala circuitry in 33 healthy adults. Dynamic faces have a higher ecological validity than traditionally used static faces and have been shown to elicit strong responses in brain networks of interest in several fMRI paradigms (Arsalidou, Morris, & Taylor, 2011; Furl, Henson, Friston, & Calder, 2014; H. Kessler et al., 2011). Negatively, neutrally, and positively valenced facial expressions were used to examine the effect of valence on the prefrontal-amygdala network, shapes served as a control condition.

In agreement with previous findings using static face processing tasks (Gläscher, Tüscher, Weiller, & Büchel, 2004; Vuilleumier, Richardson, Armony, Driver, & Dolan, 2004) or affective pictures (Urry et al., 2006), we expected an increased activation for negative valence in bilateral amygdalae and the MPFC compared to the neutral and positive valence conditions of our dynamic paradigm. Moreover, dynamic causal modeling was used to clarify the contextual influence of valence on the functional architecture of the emotion-processing network. Thus, we investigated whether valence of facial affect modulates effective connectivity within the hierarchical network architecture in a bottom-up, a top-down, or, as recently suggested (Pessoa, 2017), a bidirectional manner. Given the role of the MPFC in integrating context and salience to shape emotional responses (Etkin et al., 2015; Roy et al., 2012), we hypothesized that affective information would modulate bidirectional connections between MPFC and amygdala, as well as between MPFC and LPFC.

4.4 Materials & Methods

4.4.1 Participants

A group of 33 healthy volunteers (age in years, M = 27.4, SD = 5.2, 24 females and 9 males, 30 right and 3 left handed) was recruited for this study. Inclusion criteria were age of 18-45 years and signed informed consent. Exclusion criteria included any MRI contraindication, pregnancy, a history of brain injury, psychiatric disorders, other major medical illnesses, and drug abuse. No subject reported any past or current psychiatric disorder.

During scanning, none of the subjects exceeded our motion threshold of a mean frame-wise displacement (Power et al., 2012) of 0.5mm (M = 0.14, SD = 0.09mm). This study was approved by the ethics committee of the Kanton Zurich and was conducted in accordance with the Declaration of Helsinki.

4.4.2 Experimental design

All participants completed a 6 minutes fMRI dynamic face- and shape-matching task (Figure 4.1A), which is based on the static task used by Hariri, Tessitore, Mattay, Fera, and Weinberger (2002).

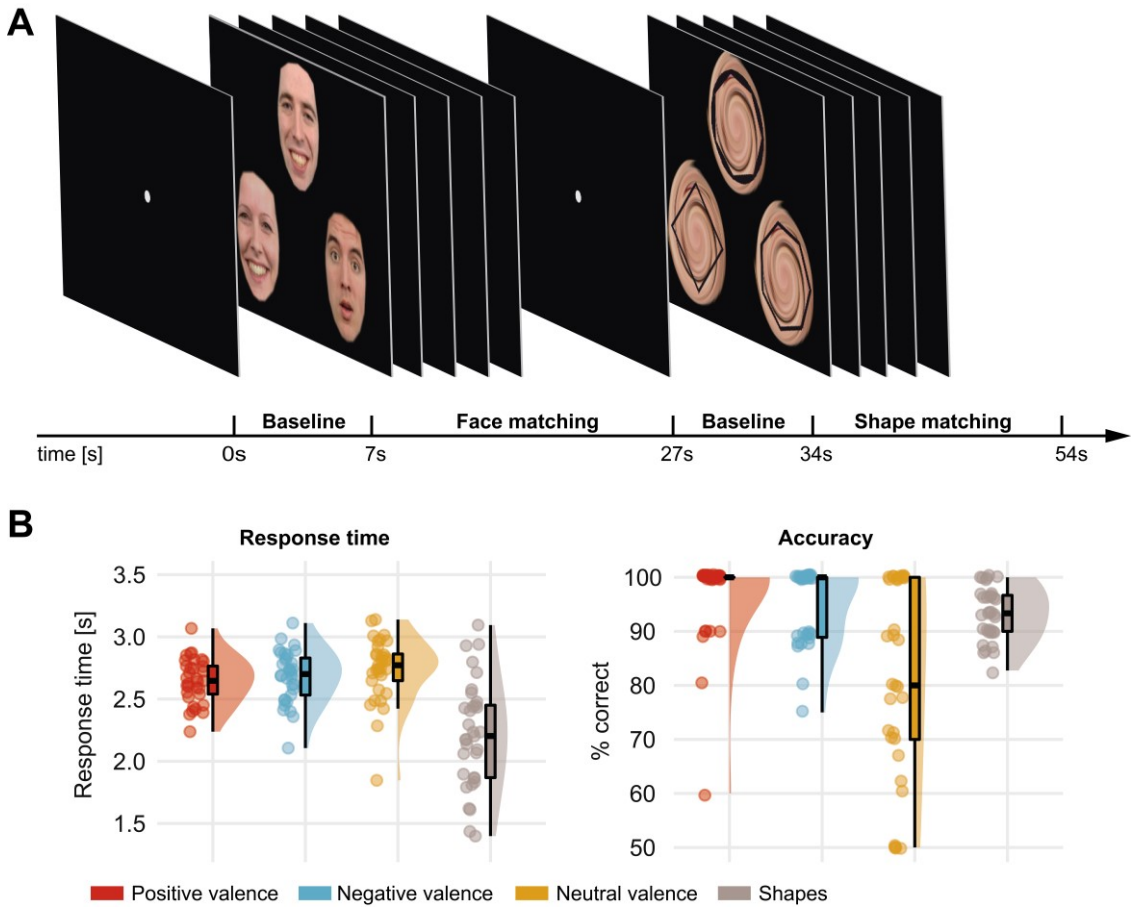

Figure 4.1. (A) Experimental design of the study. All participants were presented with triplets of emotional faces (blocks of positive, negative and neutral valence) and shapes (polygons). The task comprised matching the dynamic target image on top with one of the two static probe images at the bottom by the emotion (faces, with distinct emotional expressions for the static probe images) or number of vertices (shapes). (B) Behavioral results of the dynamic face-matching and shape-matching paradigm. Response times were comparable across different valence conditions. Response accuracy in trials with neutral faces was lower than in trials with positive or negative faces.

Face-matching and shape-matching blocks had a length of 20 seconds and were presented alternatingly. Each of the 12 blocks (6 face and 6 shape blocks) consisted of five

trials with a length of 4 seconds. In each trial, participants were presented with a dynamic target item and two static probe images below it, one of which matched with the target item with regard to shape or facial expression. Subjects were instructed to match either the left or the right static item at the bottom with the dynamic item on top and to press either the left or the right button with their dominant hand on a response-pad to indicate their choice as soon as they recognized which probe matched.

We used face images from the Radboud face database (Langner et al., 2010), including the faces of 38 Caucasian actors (19 females). In total, we presented six face-matching blocks (30 trials, 10 trials per valence), two positively valenced, including happy, surprised and neutral faces, two negatively valenced, sad and disgusted faces, and two neutrally valenced, neutral and surprised faces (see Sander & Scherer, 2014, and Soriano Salinas, Fontaine, & Scherer, 2015, for a view on surprise as neutrally valenced). In the negatively valenced condition, we used sad and disgusted faces instead of widely used fearful faces to capture negative emotion processing not related to arousal (Langner et al., 2010; Remington, Fabrigar, & Visser, 2000; Trautmann, Fehr, & Herrmann, 2009). To construct the stimuli for the positive condition, we used the inherently positive valence "happy", and included faces with neutrally valenced expressions, surprised and neutral, for the face-matching task. This formed overall positively valenced stimulus triplets in all trials, as ambiguous faces (i.e. surprised or neutral) have been shown to be perceived more positive when being presented within a positive context (Neta, Davis, & Whalen, 2011). Importantly, in each trial of the positive condition, subjects were presented with at least one happy face, either as target or probe.

Prior to the study, we established that a positive condition comprising neutral and positive faces only had lower task difficulty than the other two conditions (neutral and negative blocks). The selection of both, neutral and surprised faces, in the positive condition rendered the task difficulty across conditions comparable. Based on the face scores determined by Langner et al. (2010), the average valence of the faces used was M =3.55 (SD = 0.08) for the positive condition, M = 2.94 (SD = 0.04) for the neutral condition, and M = 2.01 (SD = 0.04) for the negative condition, where 1 represents the most negative and 5 the most positive possible valence rating. Shapes were superimposed to a whirled face in six shape-matching blocks.

We adapted the original task used by Hariri et al. (2002) using a dynamic video sequence of the target emotion or shape to make our task ecologically more valid. Target faces on top were morphed from 0% (neutral faces) to 100% (emotional faces) within the trial time of 4 seconds. *Neutral* dynamic target stimuli were morphed to *surprised* emotion from 0% to 30% intensity and back to 0% intensity to introduce neutrally valenced facial motion. Similarly, during the shape-matching task, the target shape was morphed from a round circle into a polygon with three to eight vertices. During face-matching and shape-matching, probe images on the bottom always remained static. After the button press, the video sequence continued until the end of the morphing sequence of the target

face or shape (100% morphed, 4 seconds). We ensured correct understanding of the task by familiarizing the subjects with the task outside the scanner.

4.4.3 Data acquisition and preprocessing

All MRI recordings were performed on a Philips Achieva 3 Tesla scanner (Philips Medical Systems, Best, the Netherlands) using a 32-channel head coil. Functional images were acquired with a multiband echo-planar images (EPIs) sequence [175 volumes, repetition time $TR = 2s$, $TE = 35ms$, 15° tilted downwards of AC-P C, 54slices, voxel size = 2.0 × 2.0 × $2.5mm^3$, matrix size = 96 × 94, flip angle = 80°, no gap, SENSE-factor = 2, MB-factor = 2]. Before the actual data acquisition, we acquired five dummy scans to establish steady-state conditions. After performing the task, we acquired a T1-weighted anatomical image for each subject that was used for coregistration and normalization of functional data with a 3D magnetization-prepared rapid gradient-echo sequence (MP-RAGE) [time between two inversion pulses = $2484ms$, inversion time TI = 900ms, inter-echo delay = $6.7ms$, aligned at AC-PC, flip angle = 9°, voxel size = 1.05 × 1.05 × $1.2mm^3$, field of view = 270 × $253mm^2$, 170 sagittal slices].

Preprocessing of the images included slice-timing correction, realignment, coregistration and segmentation. Normalization to the Montreal Neurological Institute (MNI)-152 template space was performed using the deformations derived from the segmentation step. In addition, preprocessing included resampling to $2mm$ isometric voxels, and smoothing with a $6mm$ full-width-half-maximum (FWHM) Gaussian kernel. All steps were performed using SPM12 (7219) software.

4.4.4 Behavioral analysis

To analyze the behavioral data, we conducted a repeated measures ANOVA to test for effects between valence conditions and paired t-tests to test for any difference in performance between face- and shape-matching. Trials without response or a response time < $100ms$ were excluded from the behavioral analysis (3.9% of all trials).

Whole brain analysis

The first-level analysis was conducted by building a general linear model using the individual onset and length of each trial (4 seconds) for face-matching and shape-matching convolved with the canonical hemodynamic response function as implemented in SPM12. To model the valence of faces, we added three regressors for each of the respective conditions. The final GLM for the whole brain analysis included five regressors of interest: a regressor for all faces, three parametric modulation regressors for each valence, that is for positive, negative, and neutral faces, and one regressor for shapes. The regressor "all faces" included 30 events, while parametric modulation regressors modeling positive, negative and neutral conditions comprised 10 events each. The regressor "shapes" included 30 events. In addition, we added the six realignment parameters derived from preprocessing as nuisance regressors.

The main effect of our task (face-matching > shape-matching) was investigated with a one-sample *t*-test using the respective contrast files of each subject. To examine the effect of valence, we performed an *F*-test in a second-level repeated measures ANOVA design across the positive, negative and neutral valence conditions. For both analyses the cluster-based family-wise error corrected significance threshold was set to $p_{FWEc} = 0.05$, the uncorrected voxel-wise cluster-defining threshold was set to $p_{CDT} = 0.001$.

Dynamic causal modeling

Dynamic causal modeling (DCM) is a hypothesis-driven Bayesian model comparison procedure for inferring effective connectivity between brain regions (Friston et al., 2003). DCM allows for the creation of different models to investigate the directed interactions of specific brain regions under experimentally controlled perturbations. These interactions are modelled at the neuronal level and related to the observable measurement via a hemodynamic forward model (Buxton et al., 1998). Importantly, it allows for estimation of endogenous coupling and context-specific, modulatory coupling (Friston et al., 2003; Penny, Stephan, Mechelli, & Friston, 2004). The neural model is given by the neural state equation

$$\frac{dz}{dt} = \left(A + \sum u_j B^j \right) z + Cu$$

in which the vector z represents the time series of the neural signal in a given region of interest and u represents the experimental inputs $(1 \dots j)$. Intrinsic (endogenous) coupling parameters between regions are stored in matrix A, modulatory parameters for a stimulus u_j are stored in matrix B, and direct driving inputs for regions are described in matrix C.

Regional time series extraction

In our study, we focused on the analysis of an emotion processing network model comprising four regions, whose adequacy has been demonstrated in previous studies (Almeida et al., 2011; Almeida, Mechelli, et al., 2009; Sladky, Hoflich, et al., 2015). In particular, we included 1) the ventrolateral part of the LPFC that is associated with emotion regulation (Hariri, Mattay, Tessitore, Fera, & Weinberger, 2003; Morris, Sparks, Mitchell, Weickert, & Green, 2012; Wagner & Heatherton, 2012), 2) the MPFC that is involved in integrating affective and contextual information, valence processing (Roy et al., 2012), and autonomous emotion regulation (Phillips et al., 2008), 3) the amygdala for its role in salience detection and facial emotion processing (Phelps, 2006), and 4) the fusiform face area (FFA) as part of the visual system, that is sensitive to faces (Kanwisher, McDermott, & Chun, 1997).

The selection and functional localization of our volumes of interest (VOI) in the amygdala-prefrontal network was guided by the results of the second-level group analyses (Table 4.2), similar to previous work (Hauser et al., 2014; Sladky, Hoflich, et al., 2015). For the amygdala, the LPFC, and the FFA we specified a spherical search volume at the peak of the face-matching > shape-matching contrast (amygdala [21, -10, -14], the right

LPFC [47, 30, 8], the right FFA [41, -44, -22] MNI [mm]). In addition, we defined a search volume for the MPFC at the peak of the main effect of valence in the second-level ANOVA comparing positive, negative and neutral valence conditions ([3, 50, -2] MNI [mm]). The individual VOI center coordinates were restricted to not differ more than 12*mm* (corresponding to twice the FWHM of the smoothing kernel) from the group maximum to ensure comparability between subjects.

Subjects' individual spherical VOIs were centered at the individual peaks (r = 6*mm*, $p < 0.05$, uncorrected) in the respective contrast and the first eigenvariate was extracted as summary statistic for all active voxels within the VOI. One subject was excluded from the DCM analysis, because we did not find any active voxels in the LPFC for the chosen threshold. We restricted our analysis to the right hemisphere, as previous studies suggested that it preferentially engages in processing of nonverbal emotional cues, such as emotional faces (A. K. Anderson et al., 2003; Fairhall & Ishai, 2006; Kanwisher et al., 1997; Ochsner et al., 2004; Puce, Allison, Asgari, Gore, & McCarthy, 1996; Sladky, Hoflich, et al., 2015).

Model space

We assumed bidirectional connection between MPFC and amygdala, and MPFC and LPFC. Although there is evidence that direct connections between LPFC and amygdala are only very sparse (Ray & Zald, 2012), it is possible that they exert influence via indirect pathways over each other. Hence, we included models with all possible intrinsic connectivity patterns between the LPFC and the AMY in the model space. In addition, we specified bidirectional intrinsic connections between the FFA and the LPFC and the FFA and the AMY, respectively. Modulation by valence was varied systematically across connections between MPFC and LPFC, and MPFC and amygdala in all possible modulation patterns, spanning a model space of 256 models.

For the DCM analysis we specified a second GLM that included five regressors of interest (all stimuli, all faces, positive faces, negative faces, and shapes) and the six realignment parameters as nuisance regressors. The "all stimuli" regressor included 60 trials, "all faces" included all 30 face events, regressors modeling positive, negative comprised 10 events each, and the shape regressor included all 30 shape-matching trials. The "all faces" regressor served as driving input of the FFA.

We performed random-effects family-wise Bayesian model selection (BMS) (Penny et al., 2010) as implemented in SPM12 to compute the expected posterior probabilities and the exceedance probabilities of model families within our sample. For model comparison, the BMS procedure uses the free energy that is a lower-bound approximation to the log-model evidence that accounts for both model accuracy and model complexity (Penny, 2012; Penny et al., 2004).

To test different functional architectures of contextual modulation, we created four different families of models (Figure 4.2). These model families differed in terms of connections on which emotional valence modulated effective connectivity. We created families

with no contextual (valence) modulations (one model), bottom-up modulations (15 models), top-down modulations (15 models), and bidirectional modulations (225 models).

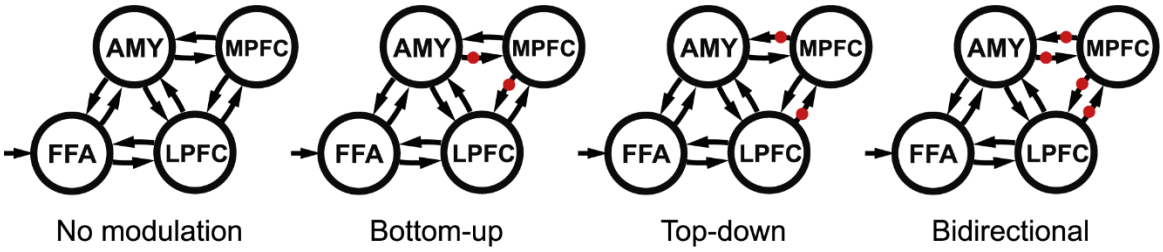

Figure 4.2. DCM model specification. We specified 256 models and grouped them into four families, depending on the location of the modulations of positive and negative valence. The modulations are depicted as red dots on the connections. In each family, all possible combinations of modulations were grouped together yielding one model with no modulation, 15 models with bottom-up modulations, 15 models with top-down modulations, and 225 models with bidirectional modulations. All faces were the driving input to the FFA.

We used Bayesian model averaging (BMA) across models to make further inferences on the significance of connections and modulation by valence (Penny et al., 2010). BMA allows for averaging the parameters while they are weighted by the posterior probability of the model and thereby accounting for the uncertainty of individual models (Stephan et al., 2010).

Subsequent one-sample *t*-tests of averaged parameter estimates were carried out in MATLAB (MATLAB and Statistics Toolbox Release 2017a, The MathWorks, Inc., Natick, Massachusetts, United States). We accounted for multiple *t*-tests of the connectivity parameters by using the procedure of Benjamini and Hochberg (1995) to control the false discovery rate (FDR, adjusted $p_{FDRc} < 0.05$).

4.5 Results

4.5.1 Behavioral analysis

The behavioral analysis of the response time (RT) and the accuracy across different conditions is summarized in Table 4.1 and depicted in Figure 4.1B.

Table 4.1. Results of the analysis of the behavioral data.

	Positive valence	Negative valence	Neutral valence	Shapes
Accuracy	96.6 ± 8.2%	94.9 ± 7.0%	81.0 ± 18.4%	93.5 ± 6.1%
Response time	2.65 ± 0.18s	2.68 ± 0.21s	2.74 ± 0.26s	2.22 ± 0.12s

Mean ± SD across all subjects (N = 33)

Responses during shape-matching were significantly faster than during face-matching, $t(32) = 5.97$, $p < 10^{-5}$. Accuracy (% correct) did not differ significantly between face-matching and shape-matching $t(32) = -2.01$, $p = 0.053$. There was a main effect of valence on accuracy for the three valence conditions, $F(2, 32) = 24.02$, $p < 10^{-7}$. Pairwise comparisons indicated that accuracy during trials with neutral faces was lower than trials with positive, $t(32) = 5.56$, $p < 10^{-5}$, and negative faces, $t(32) = 4.74$, $p < 0.0001$, suggesting a

higher difficulty in matching faces of the neutral condition. Positive and negative face conditions did not differ in accuracy, $t(32) = 1.28$, $p = 0.21$. A similar accuracy pattern was reported in previous work (Aybek et al., 2015). One-sample t-tests across positive, $t(32) = 32.69$, $p < 10^{-25}$, negative, $t(32) = 36.92$, $p < 10^{-27}$ and neutral valence, $t(32) = 9.69$, $p < 10^{-10}$, showed that accuracies were well beyond chance level (50%). Importantly, in a repeated measures ANOVA we did not find any significant differences in RTs across valence conditions, $F(2, 32) = 1.79$, $p = 0.175$.

4.5.2 Whole brain results

The dynamic face- and shape-matching task showed a significant effect of task (face-matching > shape-matching) in brain regions commonly recruited during face processing (Fusar-Poli et al., 2009), including the amygdala, the fusiform gyrus, the LPFC, the middle and superior temporal gyrus (Figure 4.3A, Table 4.2). A repeated measures ANOVA (Figure 4.3B) across valence conditions using the respective contrast images revealed a main effect of valence in the MPFC, $F(2, 32) = 17.54$, $p_{FWEc} = 10^{-7}$, the right medial temporal lobe, $F(2, 32) = 18.14$, $p_{FWEc} = 0.043$, the superior temporal gyrus, $F(2, 32) = 16.13$, $p = 0.002$, the left medial temporal lobe, $F(2, 32) = 15.34$, $p_{FWEc} = 10^{-5}$, the left cerebellum, $F(2, 32) = 18.14$, $p_{FWEc} = 0.041$, the left amygdala, $F(2, 32) = 14.72$, $p_{FWEc} = 0.009$, [$F_{(2,32)} = 14.72, p_{FWEc} = 0.009$], and the right parahippocampal gyrus, $F(2, 32) = 14.70$, $p_{FWEc} = 0.001$. Post-hoc t-tests showed that the effect in the MPFC was driven by negative valence (Table 4.2).

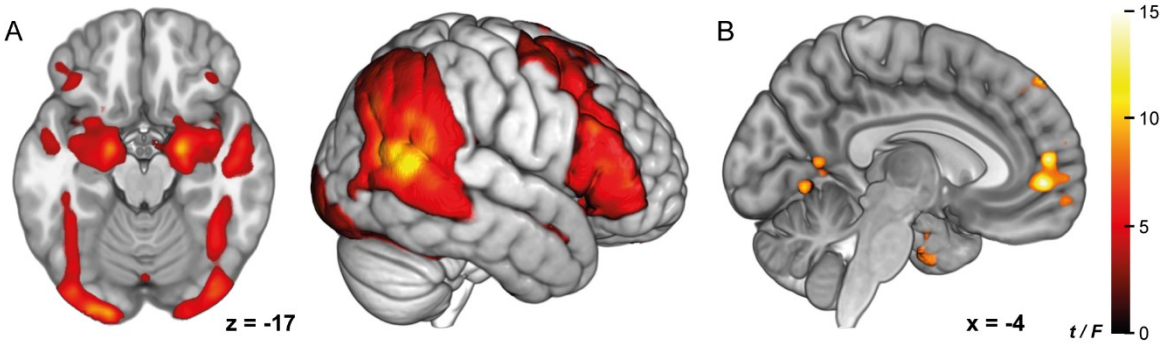

Figure 4.3. Whole brain group analyses. (A) Main effect of task face-matching > shape-matching. (B) Main effect of valence in the MPFC. Both images thresholded at $p_{FWEc} < 0.05$, with a voxel-wise cluster-defining threshold of $p_{CDT} < 0.001$, $N = 33$. Color is mapped to t-values (A) and F-values (B).

Table 4.2. Results of the group analysis (N = 33). Significant clusters on whole-brain level in the second-level contrast face- vs shape-matching, the ANOVA across valence conditions, and post-hoc *t*-tests.

	MNI coordinates [mm]			Cluster level		Peak
Brain region	x	y	z	p_{FWEc}	k	Z
Face-matching > shape-matching						
R Middle Temporal Gyrus	55	-42	6	< 0.0001	8605	7.54
R Amygdala	19	-8	-16			6.87
R Inferior Occipital Gyrus	25	-94	-4	< 0.0001	2339	7.37
R Fusiform Gyrus	41	-44	-22			6.54
L Lingual Gyrus	-21	-96	-14	< 0.0001	2143	7.30
L Fusiform Gyrus	-41	-50	-22			6.61
R Inferior Frontal Gyrus	47	30	8	< 0.0001	5966	7.24
L Middle Temporal Gyrus	-53	-60	10	< 0.0001	5096	6.91
L Inferior Frontal Gyrus	-45	34	2	< 0.0001	9479	6.63
L Amygdala	-19	-8	-14	< 0.0001	1110	6.52
R Precuneus	9	-58	40	< 0.0001	861	6.50
R Inferior Temporal Gyrus	43	-12	-42	< 0.0001	569	6.18
L Inferior Temporal Gyrus	-43	-16	-44	< 0.0001	478	6.12
L Cerebellum	-17	-74	-34	< 0.0001	604	5.78
R Middle Frontal Gyrus	27	50	6	0.0018	229	4.20
Effect of valence (ANOVA)						
R Medial Temporal Pole	45	10	-36	0.043	103	5.06
L Anterior Cingulate Cortex	-3	50	-2	< 0.0001	615	4.98
R Superior Temporal Gyrus	49	-6	-4	0.002	185	4.77
L Medial Temporal Pole	-43	14	-34	< 0.0001	350	4.65
L Lingual Gyrus	-17	-66	-4	0.041	104	4.56
L Amygdala	-19	-6	-24	0.009	142	4.55
R Parahippocampal Gyrus	23	-16	-20	0.001	196	4.55
R Amygdala	21	-2	-26			4.15
Post-hoc *t*-tests of valence conditions						
Negative faces > neutral faces						
R Medial Temporal Pole	45	10	-36	< 0.0001	511	5.53
L Anterior Cingulate Cortex	-3	50	-2	< 0.0001	1254	5.40
R Superior Temporal Gyrus	49	-6	-4	< 0.0001	393	5.26
L Temporal Pole	-35	20	-22	< 0.0001	1124	5.07
L Amygdala	-19	-6	-24			5.05
R Parahippocampal Gyrus	27	-20	-22	0.0003	325	4.99
R Amygdala	21	0	-28			4.47
L Fusiform Gyrus	-21	-52	-16	< 0.0001	791	4.67
R Paracentral Lobule	9	-32	58	0.015	173	4.55
R Paracentral Lobule	11	-44	66	0.04	138	4.26
L Inferior Frontal Gyrus	-39	32	2	0.037	141	4.24
L Middle Frontal Gyrus	-27	16	52	0.037	141	4.14
L Superior Temporal Gyrus	-55	-10	-4	0.028	150	3.95
L Superior Medial Gyrus	-9	38	50	0.036	142	3.78
Positive faces > neutral faces						
L Lingual Gyrus	-19	-66	-6	0.043	136	4.59
Negative faces > positive faces						
L Inferior Temporal Gyrus	-45	2	-34	0.048	132	4.42

Significance level at whole-brain cluster-level threshold $p_{FWEc} < 0.05$, cluster-defining threshold at $p_{CDT} < 0.001$. Labels of brain regions were determined using the SPM Anatomy Toolbox (Eickhoff et al., 2007). k, cluster size; R, right; L, left.

4.5.3 Dynamic causal modeling

Family-wise model comparison

In a first step, we compared different model families (Figure 4.4). The model family with bidirectional modulations of connections outperformed all other families with an expected posterior probability of 42% and an exceedance probability of 82%. As the model space incorporated a wide range of plausible models, we subsequently performed Bayesian model averaging to infer on the model parameters of the winning family.

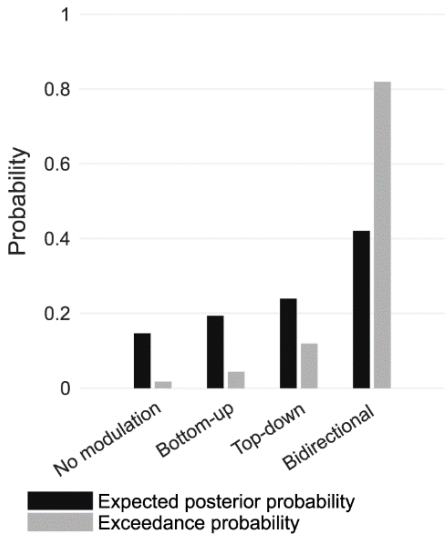

Figure 4.4. Family-wise Bayesian model comparison. Expected posterior probabilities and exceedance probabilities for the four specified model families. Asterisks (*) indicate the winning model family.

Bayesian model averaging

The results from Bayesian model averaging (BMA, Table 4.3, Figure 4.5) emphasize the relevance of connections between the amygdala and the MPFC during processing of emotional faces. One-sample *t*-tests for consistency across subjects showed that the average endogenous connectivity is significant between those regions. In addition, we found significant modulation of connectivity by valence on these connections. While the effective connectivity from the MPFC to the AMY was modulated by positive and negative valence, the modulation of effective connectivity of the bottom-up connection from the AMY to the MPFC was significant for the positive but not the negative condition.

Average intrinsic connectivity between the MPFC and LPFC differed significantly from zero across subjects. In addition, the connection between the LPFC and the MPFC showed a significant modulation effect of positive and negative valence, suggesting a specific role of this connection during processing of emotional stimuli.

After averaging the intrinsic and modulatory connectivity parameters, we correlated each subject's individual connectivity parameters with the behavioral data from the task (mean accuracy and the mean response times for the different valence conditions). However, none of the correlations remained significant after correction for multiple comparisons.

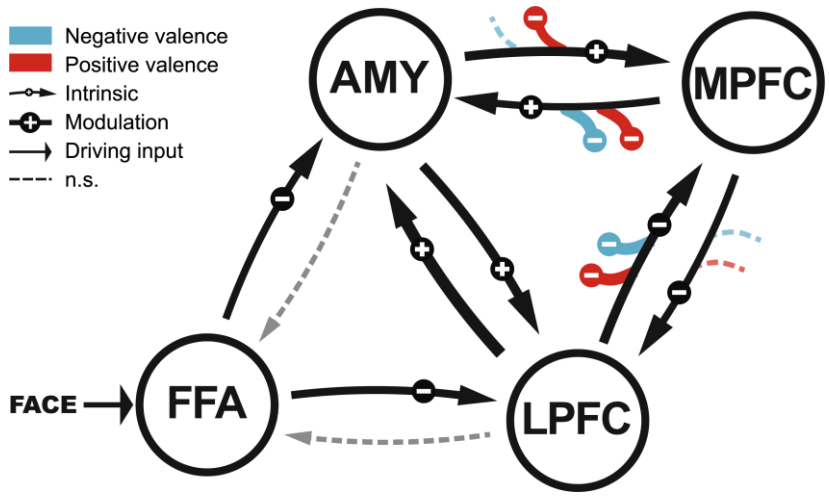

Figure 4.5. Effective connectivity during face-matching and its modulation by positive and negative valence. Parameters have been averaged with Bayesian model averaging, across all subjects and models of the winning model family. We found significant dampening of effective connectivity from the AMY to the MPFC during processing of positively valenced faces. Arrow thickness indicate effective connectivity values: thick > .15, medium > .10, thin > .05, dashed: not significant.

Table 4.3. Mean and standard deviation of endogenous and modulatory parameter estimates for all connections across all subjects and across the models of the winning bidirectional family, and the respective *p*-value resulting from a one-sample *t*-test (corrected for multiple comparisons).

Connection type	Mean	SD	p_{FDRc}
Endogenous parameters			
MPFC → AMY	0.0727	0.0193	0.0241*
MPFC → LPFC	-0.0823	0.0181	0.0109*
LPFC → MPFC	-0.1471	0.0174	< 0.0001*
AMY → MPFC	0.1122	0.0178	0.001*
AMY → LPFC	0.1207	0.0168	0.001*
LPFC → AMY	0.1702	0.0177	< 0.0001*
AMY → FFA	0.0579	0.0182	0.0505
LPFC → FFA	0.0018	0.0181	0.4786
FFA → AMY	-0.0561	0.0135	0.0325*
FFA → LPFC	-0.1076	0.0126	0.0002*
Modulatory parameters			
MPFC → AMY, positive	-0.1799	0.0698	0.01*
MPFC → LPFC, positive	0.0826	0.0607	0.0975
LPFC → MPFC, positive	-0.1575	0.0507	0.0021*
AMY → MPFC, positive	-0.1434	0.0576	0.0109*
MPFC → AMY, negative	-0.2732	0.0664	0.0001*
MPFC → LPFC, negative	0.0149	0.0625	0.4295
LPFC → MPFC, negative	-0.2174	0.0665	0.0014*
AMY → MPFC, negative	-0.1051	0.0724	0.0881

*significant (FDR adjusted $p_{FDRc} < 0.05$, df = 31). AMY, amygdala; FFA, fusiform face area; MPFC, medial prefrontal cortex; LPFC, lateral prefrontal cortex.

4.6 Discussion

Our study examined the valence-dependent functional architecture of the prefrontal-amygdala network during emotion processing using statistical parametric mapping and dynamic causal modeling. We used a dynamic face- and shape-matching paradigm in healthy subjects to assess activity and connectivity in regions supporting emotion processing and, subsequently, whether emotional valence modulates effective connectivity of bottom-up (salience signals), top-down (evaluation and regulation signals) or bidirectional connections. The results of our study suggest three main conclusions.

First, we corroborated earlier studies by showing that the MPFC as a core region of emotional response regulation is especially sensitive to negative affect (Ochsner, Silvers, & Buhle, 2012). Our data suggest that during processing of negative valence the MPFC and the right amygdala are more strongly activated than during processing of neutral valence. Second, we directly demonstrated that activity in key regions of the prefrontal-amygdala network during emotion processing is best explained by bidirectional contextual modulation of effective connectivity by valence. Accordingly, processing emotional valence directly induces changes of coupling strengths within the prefrontal-amygdala circuitry. In particular, model averaging showed that the bidirectional coupling between MPFC and AMY and unidirectional coupling between LPFC and MPFC were modulated by affective cues. This suggests that the MPFC not only serves the integration of bottom-up and top-down signals, but also continuously exerts influence on the AMY during face-matching. Third, we found evidence for a differential effect of valence on coupling between regions. On the one hand, effective connectivity from the MPFC to the amygdala was modulated by both positive and negative valence, while on the other hand effective connectivity from the amygdala to the MPFC was only significantly modulated by positive valence. Additionally, the connectivity from the LPFC to the MPFC was augmented during positive and negative valence processing. Previous studies have highlighted the role of the MPFC during emotion processing and they have underlined the role of the MPFC in processing of valence (Kawasaki et al., 2001; Winecoff et al., 2013). Studies using explicit emotion regulation paradigms have repeatedly shown that the activation of the MPFC is increased during the reappraisal of negative emotion (Delgado et al., 2008; Urry et al., 2006). Thus, it has been suggested that the MPFC supports the control of emotional responses. Moreover, recent work proposed that the involvement of the MPFC during emotion processing is related to the encoding of an integrated affective value of a stimulus (Smith et al., 2010; Winecoff et al., 2013). Importantly, this integrated affective value encoded in the MPFC might be crucially dependent on the confidence in the aggregated information (Lebreton, Abitbol, Daunizeau, & Pessiglione, 2015), which might modulate the BOLD signal in the MPFC following a U-shape pattern (Barron, Garvert, & Behrens, 2015). Our data show significantly increased activity in the MPFC during the processing of emotional stimuli with negative valence. Our findings may therefore reflect the encoding of biological significance of negatively valenced faces and provide further evidence for the encoding of stimulus valence in the MPFC. Accordingly, the lower responses

in the MPFC in the neutral condition could reflect either lower relevance or lower confidence in the nature of stimuli. This would emphasize the role of the MPFC in the integration of affective information within a valence-sensitive network, computing a value for biological significance for a given stimulus.

This is supported by our modeling results showing that the MPFC integrates affective information from multiple routes. Model averaging demonstrated that the bidirectional coupling between the amygdala and MPFC and the coupling from the LPFC to the MPFC are valence-sensitive. This is not only in line with recent theories of distributed processing of emotional stimuli along multiple parallel pathways (Pessoa & Adolphs, 2010), but also provides direct evidence for the idea that the prefrontal-amygdala circuitry can change its functional state to support appropriate mental functions for a given context (Pessoa, 2017), which potentially requires action.

Bayesian model averaging showed that positive valence significantly dampened the intrinsic connectivity between the amygdala and the MPFC, whereas negatively valenced faces did not. As task difficulty between negative and positive conditions was comparable, the observed difference in connectivity strongly suggests valence-sensitivity of the coupling between these regions. That said, the similar effective connectivity of the bottom-up connectivity during neutral and negative blocks was particularly interesting to us, since the pathway from the amygdala to the MPFC is thought to be specifically sensitive to negative valence (e.g. fear conditioning, see M. J. Kim et al., 2011, for review). Notably, this is the first study that investigated valence-dependent effective connectivity within the prefrontal-amygdala network using dynamic faces. The similar connectivity pattern of the neutral and negative condition might stem from the amygdala's role as a significance detector. The amygdala has been extensively studied and there is a broad consensus on the relevance of this brain structure in face processing (Adolphs, 2002), and more generally, the immediate detection of biological significance (Sander, Grafman, & Zalla, 2003), or resolving uncertainty (Whalen, 1998). Neutrally (i.e. ambiguously) and negatively valenced faces might induce increased predictive uncertainty compared to positively valenced faces (Whalen et al., 2013). A plausible brain response to react to predictive uncertainty would be to relocate cognitive resources to resolve it (Bubic, Von Cramon, & Schubotz, 2010). The coupling between amygdala and MPFC during processing neutral and negative facial expressions might therefore reflect a signal that translates into a need for action to increase precision and, hence, regain confidence in a volatile environment. In line with this, we found significant effective connectivity from the amygdala to MPFC during processing of neutral and negative valence, reflected in a significant intrinsic connectivity and its non-significant modulation during negative blocks, which might reflect a bottom-up confidence signal from the amygdala.

Furthermore, we found positive intrinsic connectivity from the MPFC to the amygdala during the dynamic face processing that was modulated negatively during positive and negative blocks. This is in line with previous work that observed negative effective con-

nectivity between these regions in healthy subjects using a similar task with static emotional faces (Sladky, Hoflich, et al., 2015) and might reflect a downregulation mechanism of automatic dampening the emotional response of the AMY by the MPFC to negative emotional cues (Ochsner et al., 2012).

Corticocortical effective connectivity between LPFC and MPFC was significant during face processing. A general valence-independent face-sensitive coupling between LPFC and MPFC in our dynamic task could reflect a cognitive attenuation of significance of emotional stimuli, which would eventually yield an adaption of emotional responses mediated by the MPFC. In this regard, the LPFC has been implicated in emotion regulation strategies such as repression (M. C. Anderson et al., 2004) or (spontaneous) reappraisal (Drabant et al., 2009). Based on our modeling results, we therefore propose that the valence-sensitive recruitment of the MPFC originates from the integration of affective information stemming from valence-dependent coupling within the prefrontal-amygdala network. Our findings for the afferent connections of the LPFC are in agreement with the results reported in a previous study (Sladky, Hoflich, et al., 2015), that found an up-modulating effect of the AMY on the LPFC and a down-modulating effect of medial prefrontal regions over lateral ones. On the one hand, enhanced activation of the LPFC via the AMY could reflect the allocation of attentional resources towards emotionally salient stimuli of high biological significance, on the other hand, the downmodulating signal from the MPFC might support the continuous release of these resources (Bishop, Duncan, Brett, & Lawrence, 2004; Sladky, Hoflich, et al., 2015). Interestingly, we did not observe any evidence of valence sensitivity of the connection from the MPFC to the LPFC, suggesting a general downmodulating role of this connection during face-matching.

Forward connections from the FFA showed significant intrinsic connectivity to the AMY and the LPFC, while the backward connections were not. This feed-forward functional architecture during face processing has been described previously (Fairhall & Ishai, 2006), and our results confirm these findings.

One limitation regarding the interpretation of our findings are the emotional categories of the faces used in our paradigm. Unlike many previous studies (Fusar-Poli et al., 2009; Zinchenko, Yaple, & Arsalidou, 2018), we did not use angry or fearful faces for the negative condition, but sad and disgusted faces. This choice was made deliberately to reduce the effects of arousal (Remington et al., 2000; Trautmann et al., 2009). Thus, our findings should be only interpreted in regard to the emotional expressions used in our paradigm. Despite this limitation, our results are in concordance with previous findings in the literature and provide further evidence that the state of the prefrontal-amygdala network is sensitive to valence. The goal of future investigations should be to assess whether our results are generalizable to other negative emotional expressions and to negatively valenced stimuli, other than faces, in general.

To conclude, using DCM analysis we showed valence-dependent coupling changes within the emotion processing circuitry during a dynamic face-matching paradigm. Our

findings are in agreement with recent theories of affect processing that stress the highly dynamic nature of network interactions. It has been suggested that these interactions do not only depend on task-demands, but, as our empirical data suggest, on the emotional valence of a stimulus (H. Kim et al., 2004; Pessoa, 2017). Understanding mechanisms of dynamic integration of affective value in the emotion processing network might be pivotal for explaining psychopathologies. A dysregulation of the prefrontal-amygdala network has been found in various psychiatric disorders. A disruption of neural circuitry underlying successful emotion regulation is a hallmark of various psychiatric conditions such as mood and anxiety disorders in adults (Almeida, Versace, et al., 2009; Demenescu et al., 2013; Etkin, Prater, Hoeft, Menon, & Schatzberg, 2010; Johnstone, van Reekum, Urry, Kalin, & Davidson, 2007; Liao et al., 2010; Minkova et al., 2017; Sladky, Hoflich, et al., 2015) and adolescents (Keshavan, Giedd, Lau, Lewis, & Paus, 2014; Monk, Telzer, et al., 2008; Perlman et al., 2012) and dysfunctional valence-dependent coupling might underlie the attention and processing bias in mood disorders (Clark et al., 2018; Disner et al., 2011; Groenewold, Opmeer, de Jonge, Aleman, & Costafreda, 2013). Our study provides strong evidence for alterations of coupling as a function of valence within the prefrontal-amygdala network. Based on our results, such a dynamic face-matching task thus may aid future studies to probe and disentangle mechanisms of attentional bias and valence-sensitive emotional dysregulation in neuropsychiatric disorders.

5 Inefficient evidence accumulation of social cues in adolescent major depressive disorder

David Willinger[a,b], Iliana I. Karipidis[a,b,c], Isabelle Häberling[a], Gregor Berger[a,b], Susanne Walitza[a,b], Silvia Brem[a,b]

5.1 Overview

Major depressive disorder (MDD) is often associated with a negative cognitive bias in the evaluation of interpersonal situations. Especially during adolescence, peer relationships become more complex and may contribute to an increased vulnerability in this developmental phase. The goal of this study was to examine whether adolescents with MDD exhibit altered processing of social cues. Thirty adolescents diagnosed with MDD and 33 healthy controls underwent an fMRI measurement while performing a face- and shape-matching task. The behavioral data were modelled with the linear ballistic accumulator (LBA) model to study differences in evidence accumulation and response caution. During matching of neutrally/ambiguously valenced faces, processing efficiency reflected by the drift rate parameter of the LBA was reduced in the MDD group. Critically, this reduction in processing efficiency was related to hypoactivity in the subgenual anterior cingulate cortex. Our results suggest that the negative bias in cognitive processing is related to inefficient evidence accumulation and slower sampling of nuanced social cues, which might enhance cognitive vulnerability in adolescent MDD.

5.2 Introduction

There is ample of evidence for an attentional negativity bias associated with maladaptive evaluation and behavior in individuals with major depressive disorder (MDD; Beck, 1967; Disner et al., 2011; Foland-Ross & Gotlib, 2012; Ingram, Miranda, & Segal, 1998). Especially in youth at familial risk, this cognitive bias is most likely an important vulnerability factor that facilitates the development of MDD (see Platt, Waters, Schulte-Koerne, Engelmann, & Salemink, 2017, for a review). While in adult MDD it has been shown that a stronger negativity bias increases response to behavioral activation therapy (Gollan et al., 2016), evidence on the predictability of treatment outcome in youth is limited.

Neuroscientific work suggests that such a bias might be related to impairments of brain function in the prefrontal-amygdala network (Ho et al., 2014). Specifically the increased reactivity of the amygdala (Gaffrey et al., 2013; Hall et al., 2014; Mingtian et al., 2012; Redlich et al., 2018; Tao et al., 2012; Yang et al., 2010) in adolescent MDD and the altered activation in the subgenual anterior cingulate cortex (sgACC, Drevets, Savitz, & Trimble, 2008), a region associated with mediating affective signals from the amygdala to cortical structures (Disner et al., 2011), may be directly related to the cognitive bias in MDD

[a] Department of Child and Adolescent Psychiatry and Psychotherapy, University Hospital of Psychiatry Zurich, University of Zurich, Switzerland
[b] Neuroscience Center Zurich, University of Zurich and ETH Zurich, Switzerland
[c] Center for Interdisciplinary Brain Sciences Research, Stanford University School of Medicine, Stanford, CA, USA

(Dannlowski et al., 2007). While the amygdala and the sgACC have been independently identified to show altered activity patterns during different forms of emotion processing, also their functional coupling has been found disrupted during emotion processing (Ho et al., 2014). Crucially, brain activity in the anterior cingulate cortex is predictive of treatment response to different therapies or medications (Keedwell et al., 2010; Roiser, Elliott, & Sahakian, 2012; Siegle et al., 2012).

In search for a better mechanistic understanding, only few studies have aimed at identifying a link of behavioral measures and brain activity in these regions (Ho et al., 2016). A major challenge in the field of psychiatry is to determine appropriate treatments for patients based on their symptoms. The heterogeneity of possible symptoms summarized in a diagnosis of MDD suggest that different symptom characteristics follow different disease mechanisms, and might therefore be an important factor of why the diagnostic label only has limited predictive value for treatment outcome and complicate the selection of treatment (Cuthbert & Insel, 2013). Thus, to advance treatment efficacy, a promising approach in the field of computational psychiatry is to establish measures that enable treatment of patients based on a specific neurocomputational disease mechanism rather than a cluster of symptoms (Huys et al., 2016). Although other medical fields have benefited from the application of biomarkers to inform treatment selection (Aronson & Ferner, 2017), neurocomputational biomarkers have thus far only limited clinical utility in psychiatry. Employing computational models to describe the cognitive processes allows for disentangling subprocesses and identifying related neural correlates. Importantly, finding a computational signature reliably linked to neurophysiological changes would have the potential to facilitate the translation of research into clinical tools for diagnosis and treatment selection.

Using a face- and shape-matching task with dynamic faces and varying valence (negative, positive, neutral; Willinger et al., 2019) we investigated emotion processing in adolescent depression. In this analytic framework we applied the linear ballistic accumulator (LBA; Brown & Heathcote, 2008) model to discern effects of valence on information processing and to subsequently identify the neural correlates of emotion processing related to the drift rate parameters of the behavioral model. Given the evidence, that negative and positive valence engage the prefrontal-amygdala network differentially (Chapter 4), the contextual factor (valence) was allowed to modulate the model parameter. We hypothesized that participants with MDD show aberrant perceptual decision making, reflected by differences in the drift rate (Ho et al., 2016). Moreover, we investigated whether any such effect on the decision process is related to changes in activity in the amygdala or the sgACC.

5.3 Methods

5.3.1 Participants

Thirty MDD patients and 33 healthy individuals matched for age, IQ, gender, and handedness participated in this study. Demographic and clinical characteristics have been reported in Chapter 3. We conducted a semistructured clinical interview (K-SADS-PL, Kaufman et al., 1997 or MINI-KID, Sheehan et al., 2010) with all participants. All participants gave their written informed consent and were financially reimbursed at the end of the study. This study was approved by the ethics committee of the Kanton Zurich and was conducted in accordance with the Declaration of Helsinki.

5.3.2 Experimental task

In this study, healthy controls and participants with MDD performed the dynamic face- and shape-matching task (Willinger et al., 2019). Details of the task can be found in the Methods section of Chapter 4, however the task here differed in two aspects. First, here, subjects were presented with 4 blocks of 5 trials each for each condition (positive, negative, neutral, shapes) in randomized order, resulting in 80 trials in total. Second, to create the neutral condition, we presented faces from neutral (valence rating from Langner et al., 2010, scale: 1 (negative) – 5 (positive); M = 2.95, SD = 0.02) and contemptuous (M = 2.79, SD = 0.05) category, leading to more nuanced dynamic neutral expressions. Third, in the positive condition, we used happy faces (M = 4.60, SD = 0.15) and the most positively rated surprised faces (M = 3.04, SD = 0.06). Sad (M = 1.80, SD = 0.05) and disgusted (M = 1.83, SD = 0.11) categories were used for the negative condition.

5.3.3 Computational modeling

In this study, we used the R implementation of a hierarchical LBA model distributed with the Dynamic Models of Choice toolbox (DMC; Heathcote et al., 2019; https://osf.io/pbwx8/) to fit the response data for correct and incorrect trials.

Here, the LBA contained two evidence accumulators gathering information for the two possible responses (left or right face matches the target). The drift rate parameter v quantifies the speed of the evidence accumulation and thus the information processing efficiency. Within the present framework, the drift rates for correct and error trials are drawn from a normal distribution with the separately estimated between-trial variability sv. To make the model identifiable, typically the sv for error responses is fixed at 1 (Donkin, Brown, & Heathcote, 2009). A button press is initiated, as soon as one accumulator surpasses the response threshold, encapsulated by the decision threshold parameter B. Lastly, the parameter $t0$ captures any effects of non-decision processes (e.g. motor preparation) and the parameter A encodes the starting point of the accumulation process.

We assumed that either the drift rate (i.e. evidence accumulation) or the decision threshold (i.e. the evidence required for a decision) are influenced by the valence context. Thus, we created two models that allowed (1) the drift rate (2) the threshold parameter to vary as a function of condition. In addition, we created a null model, where both parameters

were constant across conditions. In total, three models were fitted to the response data of healthy adolescent controls and participants with MDD separately with Differential evolution Markov Chain Monte Carlo simulations (DE-MCMC, Turner, Sederberg, Brown, & Steyvers, 2013). Thirty-six chains were used for sampling the posterior distribution of the parameters thinned by keeping only every 10^{th} sample. Initial values for the hierarchical sampling were determined using fixed-effects fits. In the burn-in period, we used a 5% probability of migration for individual and group levels, after burn-in only the crossover steps of the DE-MCMC algorithm were performed during subsequent sampling. We set a convergence threshold of Gelman-Rubin (1992) values < 1.1.

To determine the best model, we performed model selection using the default metric in the DMC software, the Bayesian predictive information criterion (BPIC; Ando, 2011). Finally, we assessed group differences on the group-level distribution of the hierarchical model by computing the odds ratio (OR) of the observed difference of the posterior distributions. To this end, we divided the number of samples in the difference distribution above 0 by the number of samples below 0. An OR > 100 was regarded as decisive evidence, an OR > 10 as strong evidence (Jeffreys, 1998).

5.3.4 Data acquisition and preprocessing

All MRI data were acquired with a 3T Philips Achieva MRI (Philips Medical Systems, Best, The Netherlands). For details of the functional data acquisition please refer to Chapter 2. We recorded 295 volumes. The data was preprocessed performing slice-timing correction and realignment. Realigned functional data was coregistered to the individual T1-weighted anatomical scan and transformed into MNI-152 space using deformation field normalization. Finally, we spatially smoothed the data with a $6mm$ FWHM kernel.

5.3.5 Functional MRI data analysis

In the first-level analysis, the GLM contained an individual regressor for each face condition (negative, positive, neutral) and one for the shapes using the onsets of each trial convolved with the hemodynamic response function. Each condition comprised 20 individual trials, in 4 blocks of 5 trials. Moreover, six realignment parameters derived from preprocessing and a regressor flagging bad scans were used as nuisance regressors. Two-sample t-tests were used to test group differences on the effects of the task conditions. Linear regression analyses were performed to study the relationship between the behavioral model parameters and the whole-brain activity.

For all analyses we used a cluster-extent threshold to perform family-wise error correction using an uncorrected voxel-wise threshold of $p_{CDT} = 0.001$. To derive the cluster size we used Monte Carlo Simulation (Slotnick, 2017; Slotnick, Moo, Segal, & Hart Jr, 2003), running 10,000 iterations resulting in a minimum cluster-size of $k > 440mm^3$ corresponding to 55 contingent voxels, $p_{FWEc} < 0.05$.

5.4 Results

5.4.1 Summary statistics of behavioral measures

An analysis with a linear mixed-effects model of logRTs showed a significant effect of valence, $F(3, 4666.9) = 510.07$, $p < 10^{-15}$. Participants were fastest when presented with positive faces, then negative faces, and slowest in the neutral condition. Neither the main effect of group, $F(1, 61) = 0.26$, $p = 0.615$, nor the interaction term, $F(3, 4666.9) = 0.91$, $p = 0.424$, reached significance. The number of incorrect responses showed no main effect of group, $F(1, 61) = 0.46$, $p = 0.499$, and no significant group-by-condition interaction, $F(3, 183) = 0.04$, $p = 0.989$. Across conditions, however, there was a significant main effect, $F(3, 183) = 129.94$, $p < 10^{-15}$. Participants were more accurate when presented with negative and positive faces, and shapes, compared to neutral faces. We present the summary of the data in Table 5.1.

Table 5.1. Descriptive summary statistics of response data for all conditions in both groups.

		HC	MDD	Test statistic	*p*-value
Response time (s)	Negative	2.42 (0.33)	2.45 (0.32)	$t(61) = -0.403$	.688
	Positive	2.21 (0.36)	2.24 (0.30)	$t(61) = -0.307$	.760
	Neutral	2.93 (0.22)	3.01 (0.22)	$t(61) = -1.387$	.171
	Shapes	2.90 (0.29)	2.89 (0.31)	$t(61) = 0.111$	.912
Accuracy (%)	Negative	97.6 (3.8)	96.7 (4.0)	$t(61) = 0.927$	.358
	Positive	98.8 (2.8)	98.2 (4.3)	$t(61) = 0.691$	.493
	Neutral	72.6 (18.0)	71.5 (14.4)	$t(61) = 0.260$	.796
	Shapes	77.7 (10.2)	76.0 (12.8)	$t(61) = 0.595$	.554

Data is shown in mean (SD). Tests are performed on the means of each subject.
Abbreviations: HC, healthy controls; MDD, major depressive disorder.

5.4.2 Computational modeling

We used the LBA model to account for the response time and accuracy data of the participants. First, we were interested in how the parameters were modulated by valence. Model comparison based on the BPIC showed that the model v where the drift rate was allowed to vary by valence accounted for the data best in both groups (Table 5.2). Subsequently, we tested whether the parameters in the winning model v revealed any group differences. The density of the posterior distribution of the group-level for the parameters of interest are shown in Figure. We found decisive evidence that during neutrally valenced face matching MDD patients showed slower information accumulation compared to healthy controls (OR = 141.11:1). Furthermore, there was evidence for slower information accumulation also during positive (OR = 24.84:1) and negative (OR = 21.59:1) face matching conditions. In contrast, we found only little evidence that drift rate for shapes (OR = 2.53:1), overall decision threshold B (OR = 0.24:1), bias A (OR = 4.08:1) or non-decision time t_0 (OR = 2.07:1) differed between patients and controls.

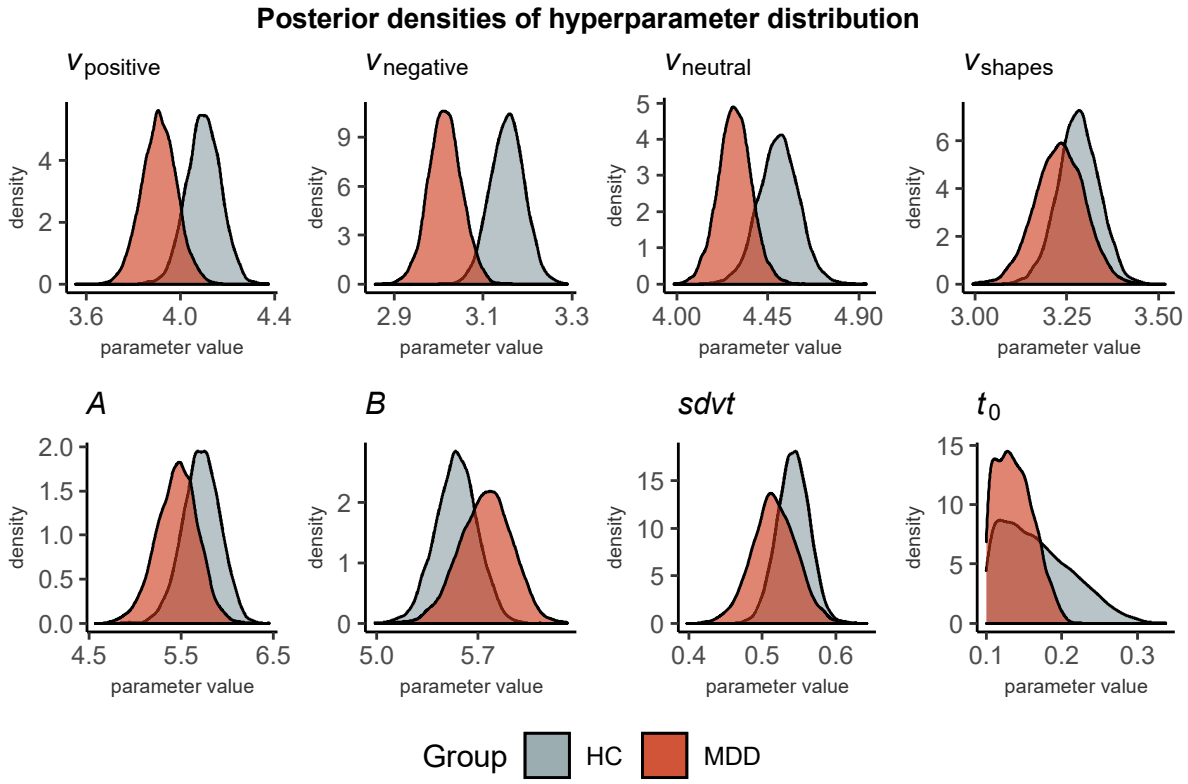

Figure 5.1. Behavioral parameters. Patients exhibited slower information processing efficiency represented by the drift rate during the neutral condition. Abbreviations: A, width of starting point distribution; B, decision threshold; sdvt, standard deviation of the drift rate; t_0, non-decision time; v, mean of the drift rate.

Table 5.2. Results of the model comparison based on Bayesian predictive information criterion. In both groups, the model with drift rate *v* varying as a function of valence described the response data best.

	Group	Model 0	Model *v*	Model *B*
BPIC	HC	7784.665	5615.285	5671.337
	MDD	6983.194	5120.881	5155.672

Abbreviations: BPIC, Bayesian predictive information criterion; HC, healthy controls; MDD, major depressive disorder. 0=null model, *v*=drift rate, *B*=decision threshold

5.4.3 fMRI results

When assessing group differences of each of the four different task conditions, we did not find any group differences surviving the preset cluster-extent threshold. As the behavioral results suggested variability in information processing between patients and controls, we investigated the neural correlates of this phenomenon. Importantly, our analysis revealed a significant cluster in the sgACC whose activity during neutrally (i.e. ambiguously) valenced face processing varied as a function of the information processing efficiency (Figure 5.2, Table 5.3).

During processing positive face processing, we found significant clusters in the bilateral inferior temporal gyrus, right middle and superior temporal gyri, right supramarginal

gyrus and left postcentral gyrus, positively associated with drift rate (Table 5.4). During negative faces, no cluster was significantly associated with drift rate. Lastly, in none of the contrasts, an association with the activity in the amygdala and the drift rate was found. The task elicited activation in brain areas commonly reported in face-matching or emotion processing paradigms (Chapter 4; Fusar-Poli et al., 2009, Figure 5.3, Table 5.5).

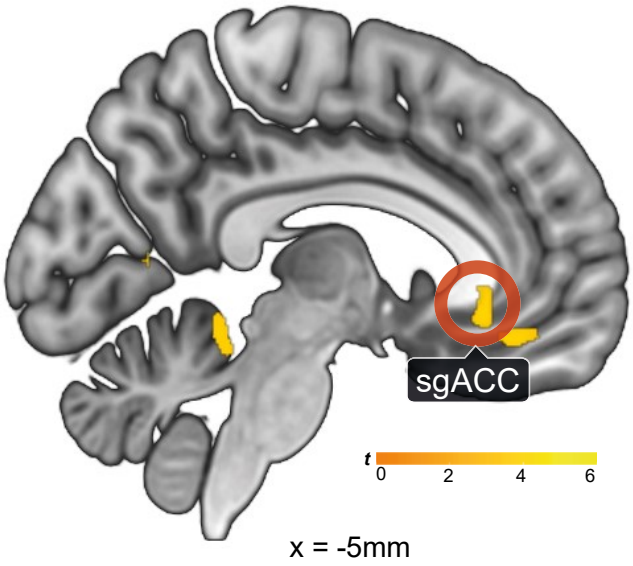

Figure 5.2. Brain activity positively associated with the drift rate (i.e. information processing efficiency) during the neutrally/ambiguous valenced dynamic face matching. We found that slower a slower drift rate is related to a cluster in the subgenual anterior cingulate cortex (sgACC). p_{FWEc}<.05, p_{CDT} <.001, N = 63.

Table 5.3. Positive association between the drift rate *v* for faces with neutral/ambiguous valence and brain activity

	MNI coordinates [mm]				
Brain region	x	y	z	k	Peak Z
R Precentral gyrus	53	-2	22	305	4.54
L Lingual gyrus	-11	-42	-10	136	4.43
L Subgenual anterior cingulate	1	24	-4	81	4.38
L Medial orbital frontal gyrus	-9	38	-14	103	4.33
L Cerebellum 6	-19	-60	-20	70	4.00
R Cuneus	19	-72	26	89	4.00
L Calcarine	-11	-68	6	96	3.94
R Calcarine	19	-62	6	95	3.78

Significance level at whole-brain cluster-level p_{FWEc} < 0.05.
Abbreviations: k, cluster size; R, right; L, left; MNI, Montreal Neurological Institute.

Table 5.4. Positive association between the drift rate v for faces with positive valence and brain activity

	MNI coordinates [mm]				
Brain region	x	y	z	k	Peak Z
L Inferior orbital gyrus	-53	80	-6	85	5.26
L Postcentral gyrus	-51	-20	42	218	4.29
R Superior temporal gyrus	67	-36	20	148	4.23
R Inferior temporal gyrus	55	-72	-8	85	4.21
R Middle temporal gyrus	55	-32	-4	130	4.16
R Precentral gyrus	63	10	28	230	4.12

Significance level at whole-brain cluster-level $p_{FWEc} < 0.05$.
Abbreviations: k, cluster size; R, right; L, left; MNI, Montreal Neurological Institute.

Table 5.5. Task effect (faces > shapes) across both groups

	MNI coordinates [mm]				
Brain region	x	y	z	k	Peak Z
R Inferior occipital gyrus	27	-96	-4	866	> 8
L Inferior occipital gyrus	-23	-98	-6	637	> 8
R Middle cingulate	5	-52	34	2222	> 8
R Angular gyrus	59	-66	26	4774	> 8
L Angular gyrus	-51	-64	50	4447	> 8
R Amygdala	19	-8	-18	526	> 8
L Amygdala	-19	-8	-16	535	> 8
R Fusiform gyrus	41	-52	22	139	> 8
L Fusiform gyrus	-41	-52	-24	84	7.30
L Inferior frontal gyrus, triangular	-53	26	0	424	7.27
L Superior frontal gyrus	-15	44	44	145	6.97
NA	-11	-26	24	175	6.84
L Medial orbital frontal gyrus	1	52	14	125	6.44
R Inferior frontal gyrus, triangular	53	32	0	89	6.22
L Rolandic operculum	-39	-18	20	96	6.06
L Superior frontal gyrus, dorsolateral	-13	58	18	362	6.04
NA	13	-32	22	56	6.03
R Superior frontal gyrus, dorsolateral	15	36	50	69	5.93
R Paracentral lobule	3	-30	60	63	5.76

Significance level at whole-brain correction $p_{FWEc} < 0.05$, minimum cluster size k > 55.
Abbreviations: k, cluster size; R, right; L, left; NA, not available; MNI, Montreal Neurological Institute.

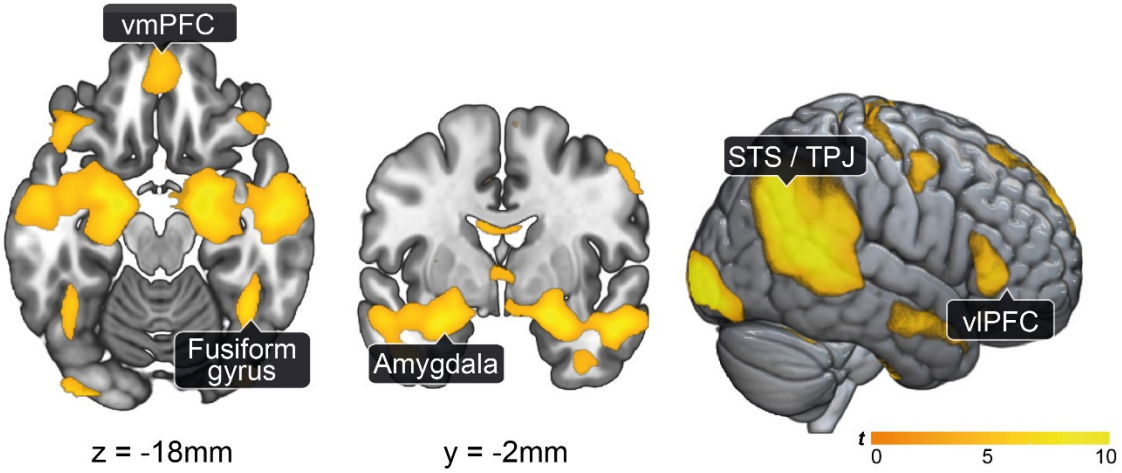

Figure 5.3. Brain activity for the task (faces-shapes) across both groups. We found activity in the amygdala, the fusiform gyrus, the ventromedial prefrontal cortex (vmPFC), ventrolateral prefrontal cortex (vlPFC), and in a cluster comprising the superior temporal sulcus (STS) and the temporo-parietal junction (TPJ). $p_{FWEc} < 0.05$, $p_{CDT} < 0.001$, $N = 63$.

5.5 Discussion

Reading and understanding emotions of others is essential for living in a social environment. Interpersonal theories of MDD hypothesize that deficits in social behaviors may impede rewarding or even increase negative social experiences and thereby contribute to maintenance or aggravation of depressive episodes (Bistricky, Ingram, & Atchley, 2011). Previous research has identified that facial affect processing and specifically the evaluation process of emotion is negatively biased in MDD, and it has been proposed that the ventral-affective brain network is particularly vulnerable to such biased processing in adolescence (Badcock, Davey, Whittle, Allen, & Friston, 2017).

Here, we showed that using a sequential sampling model, the behavior in MDD and HC can be reliably discerned during processing of faces with ambiguous facial emotion. By applying a computational model for the behavioral data (response time and accuracy), we were able to investigate distinct components of the decision process during face- and shape- matching in participants with and without MDD with varying valence. First, using model selection, we detected that contextual factors primarily influence information processing efficiency (and thus task difficulty), instead of response threshold (i.e. quantified caution of subjects). That is, participants did not adjust their decision threshold in ambiguous situations, but rather processing efficiency was decreased. Second, when comparing the parameter estimates between groups, we found strong evidence for a lower drift rate in MDD when processing neutral (i.e. ambiguous) valence, and thus less efficient evidence accumulation. We extend previous findings (Ho et al., 2016) by showing that the evidence accumulation is not only dependent on the contextual factor of facial expressions (i.e. valence), but it is also most sensitive for group differences for more subtle facial expressions.

Moreover, in a natural environment, it is impossible to fully infer a person's intention based merely based on a facial expression, but it is necessary to deduce latent causes from noisy sensory information (Huys et al., 2015). When combining the computational model with fMRI, we determined that the slower accumulation process was associated with aberrant activity in the sgACC during the decision process. The sgACC is in a position to tune sampling of the sensory evidence (Barrett & Simmons, 2015), hence slower drift rates in MDD might reflect an adaptive response of additional evidence sampling afforded by the more nuanced expressions. This regulatory function of the sgACC might be crucial to encode precise predictions about changes in the environment for a preparatory bodily response for anticipated situations that, supported by the amygdala, eventually generate a patterned physiological response (Barrett & Simmons, 2015; Dixon, Thiruchselvam, Todd, & Christoff, 2017). Any failure of appropriate recruitment of the sgACC might yield disrupted autonomous regulation with a decreased ability to create an appropriate physiological response and inefficient inference on the correct latent causes from the environment especially during ambiguous interpersonal situations.

Many studies have implicated the sgACC in the etiology of depression, with reports on functional abnormalities (Drevets et al., 2008; Greicius et al., 2007; Skaf et al., 2002) that are predictive of treatment outcome (Keedwell et al., 2010; Roiser et al., 2012; Siegle et al., 2012), structural abnormalities (Lehmbeck, Brassen, Braus, & Weber-Fahr, 2008), and its use as successful stimulation target in treatment-resistant MDD (Mayberg et al., 2005). Importantly, here we show a mechanistic link of behavior with activity in a brain region that is predictive for treatment response. Note, the conventional group comparisons relying on group means of response time or brain activity did not reveal any significant differences in behavior and brain activity between adolescents with MDD and healthy controls. That shows not only that the separation of decision making into its subprocesses by computational modeling allowed for a more sensitive analysis of the data, it also emphasizes the need to investigate psychiatric disorders by means of disease mechanisms. Establishing such relationships is of particular importance, as the reliability of biomarkers for depression based on task using emotion processing paradigms has been recently questioned (Nord, Gray, Charpentier, Robinson, & Roiser, 2017).

One point to consider in future studies is the optimal task design to improve modeling of perceptual decision making. In this study, there were only few error trials in the positive and negative valence condition. Although in DMC framework this can be at least partially accounted for by including conditions with a sufficient error rate (Heathcote et al., 2019), future studies should ensure a more advantageous ratio of correct and incorrect trials which should even further improve parameter estimation and inference on cognitive processes. It would also allow contextual factors to modulate different model parameters simultaneously, which might improve the fit of the data, however, we did not explore this possibility due to the limited amount of trials in the MRI (Gesiarz, Cahill, & Sharot, 2019). In addition, future studies in search of biomarkers might benefit from higher field

strengths and higher resolution for imaging the ventral brain to increase reliability (Geissberger et al., 2020).

On balance, in this chapter, we present an analytic approach to study the behavioral and neural correlates of cognitive bias in adolescent depression. Disrupted evaluation of interpersonal situations is a hallmark feature of MDD and, thus, aberrant processing of nuanced social cues in patients with MDD might represent a critical cognitive vulnerability factor (Bistricky et al., 2011). The current work is a next step towards determining mechanisms of face processing and significantly advances our understanding how altered emotion processing is related to neurophysiology in adolescent MDD.

6 General Discussion

This thesis assessed the application of computational models of reinforcement learning and perceptual decision making in combination with fMRI to unravel mechanisms of neurodevelopment and psychiatric disorders in youth. It provides novel mechanistic insights into behavioral and neural correlates of reinforcement learning in normative development (Chapter 2) and adolescent MDD (Chapter 3). In addition, it advances the understanding of valence-dependent emotion processing in the prefrontal-amygdala network in healthy adults (Chapter 4) and provides preliminary evidence of emotion processing deficiencies in MDD (Chapter 5). The current chapter will discuss the main findings in a more general, broader neuroscientific context and addresses important open questions in the field.

6.1 Studying typical and altered developmental trajectories of the brain with neuroimaging

Aberrant information processing, decision making, and learning in psychiatric patients has been explored extensively with traditional brain mapping methods. Early neuroimaging studies provided important insights into the function of brain networks implicated in mental disorders. In the last years, more researchers have begun to integrate computational models into their analysis. Compared to more conventional neuropsychological analyses, which limited behavioral analysis to response times or measures of accuracy, model-based analyses have specific hypotheses about the mechanisms that generate observable behavior. Model-based fMRI provides a way to integrate parameters of interest derived from computational models and functional neuroimaging allowing to understand which computations are carried out by the brain to accomplish a specific task (O'Doherty et al., 2007). Thus, model-based fMRI can give answers to the question how specific cognitive processes are implemented in particular brain areas. As a growing body of work shows, this might be particularly relevant for understanding psychopathologies, where characterizing a mental disorder with a computational model may help inferring latent disease mechanisms from observable behavior or neurophysiology (Stephan et al., 2015).

6.1.1 Unravelling the emergence of psychiatric disorders during adolescence

The burgeoning field of computational psychiatry (Huys et al., 2016; Montague et al., 2012) is concerned with the development of viable, predictive models of behavioral and biological data to improve clinical utility and advances conventional methods to overcome some of their limitations. Generative computational models allow for determining the computations performed by an agent that are utilized to generate overt behavior with explicit mechanistic meaning depending on their specific role in the model. Thus, research in this field is now focusing on revealing mechanisms that give rise to normative and psychopathological cognitive function. Eventually, this approach holds promise to

enable the revelation of disrupted mechanisms on behavioral and neural levels in cognitive processes of interest (Hauser, Will, Dubois, & Dolan, 2019).

The main focus of this thesis was the behavioral modeling of perceptual decision making in a face- and shape-matching paradigm and the use of a reinforcement learning model to determine response vigor in an incentivized task, both in combination with neuroimaging. Utilizing the models allowed us to characterize information processing and instrumental learning in adolescents with MDD. Since psychiatric disorders are associated with impairments in a variety of vital cognitive functions, computational modeling could aid the research to improve and quantify any functional deficiencies in patients. However, in MDD and even more so in adolescent MDD, the evidence for impaired learning or decision making is still weak, because only few studies have investigated the disorder with methods of computational psychiatry (Ho et al., 2014; Ho et al., 2016).

Recent computational accounts on MDD mainly based on an extensive literature on *reward* processing proposed that maladaptive beliefs of one's state and the environment impairs goal-directed decision making and learning in patients (Huys et al., 2015; Kube et al., 2020). The role of aversive (e.g. loss-avoidance or punishment) processing in MDD is much less studied, however evidence suggests that altered neural loss processing might represent a predictor for developing MDD during adolescence (Ho & Auerbach, 2017). The present work in Chapter 3 critically extends prior findings by showing altered processing in loss-avoidance context in youth with MDD, an effect that is in line with prior behavioral reports of disrupted processing of negative feedback in adult MDD (Tavares et al., 2008). We identified a potential mechanism that might be able to explain the persistence of maladaptive behavior in patients with MDD. Excessively avoiding aversive outcomes might result in less adaptation of behavior and deficits in motivation (Nord, Lawson, Huys, Pilling, & Roiser, 2018). While holding a negative belief when facing an uncertain environment might have an adaptive function, in clinical MDD, this might become maladaptive leading to persistent negative anticipation and evaluation of outcomes (Badcock et al., 2017).

Such negative attentional bias is central to cognitive theories of depression (Disner et al., 2011). A negative bias results in greater attention away from positive towards (more expected) negative stimuli and lets individuals evaluate neutral information in a more negative manner (Platt et al., 2017). One way to quantify this bias is to model response data, which was explored in this thesis. The results of our preliminary work in Chapter 5 suggest, that in adolescent MDD negatively valenced stimuli (i.e. sad or disgusted faces) are not processed differently than in typically developing adolescents. However, interestingly we found that brain activity in the subgenual anterior cingulate cortex, a brain region central for emotional processing, was associated with decreased processing efficiency of ambiguous (neutral) face stimuli. This preliminary finding suggests that a negativity bias affects social functioning in ambiguous interpersonal contexts in adolescent MDD. Automatic negative evaluation of ambiguous situations might contribute to the

cognitive vulnerability and negative self-evaluation (Platt et al., 2017). Mood, as a complex aspect of MDD, has been recently conceptualized in terms of an aberrant prior over the belief of the experience or perception of emotion that is mediated by the prefrontal-amygdala network (Clark et al., 2018). Thus, the next steps of the work provided in this thesis is to characterize the prefrontal-amygdala network dynamics in adolescent MDD during emotion processing. Results from Chapter 4 suggest that a negative bias might be reflected in impaired valence-sensitive coupling, however, more work is needed to shed light on how functional pathways within the prefrontal-amygdala network are affected over the course of MDD.

This work underscores the potential to explore the domain of instrumental learning and emotion processing to reveal reliable computational markers in adolescent MDD. This could be not only informative about the prediction of depression onset (Jin et al., 2017), but also about different depression subtypes and their respective optimal treatment strategies. Moreover, this thesis has provided a general framework of using and integrating generative models on the behavioral and neural level that can serve as an example on how to combine data across analytic levels (Marr, 1982). This provides an exciting and important next step for a better mechanistic explanation of the behavioral manifestation of psychopathology. The development and tailoring of novel tasks that are able to isolate the computations gone awry in depression is critical to improve the understanding of MDD and psychiatric disorders in general. Eventually, this can further increase our understanding and could be introduced as diagnostic tool for differential diagnosis or treatment selection.

6.1.2 Computational models to describe behavioral and neural developmental patterns

The herein presented studies illustrated the use of computational models to identify disrupted mechanisms in adolescent MDD, however, it is a general approach to study mechanisms underlying specific behavior. A second application of computational models is neurodevelopmental research. Only a few studies exist, that have employed computational methods to investigate normative neurodevelopment (e.g. Hauser et al., 2015) but recently this approach has attracted more interest in the literature (Davidow et al., 2018). It is well known that behavioral changes over time along with structural and functional shifts of the brain during adolescence (Sturman & Moghaddam, 2011). An establishment of computational developmental neuroscience might help to understand phenomena related to value guidance and the emergence of cognitive control across behavior. The findings presented in this thesis (Chapter 2) corroborate the idea that the fine-tuning of functional circuits within the incentive network gives rise to motivated behavior across adolescence (C. Insel et al., 2017). Thus, the work here provides a next step forward towards understanding the emergence of value-driven, motivated behavior across development. Given the restructuring of priorities in the transition from childhood to adolescence into valuing and maintaining social relationship, future work should focus on constructing models for testing mechanisms for social judgement, inclusion and exclusion,

as they might allow to reveal cognitive vulnerabilities in interpersonal situations, an important risk-factor for developing depression (E. E. Nelson, Leibenluft, McClure, & Pine, 2005).

6.1.3 Limitations

The studies covered within the framework of this thesis have advanced developmental and psychiatric research; nonetheless, some limitations need to be considered.

First, research in *pediatric samples* faces specific challenges for the conduct of the experiment that confine data quality and quantity. Task performance has to be considered for within-group and between-group variability. Thus, tasks are usually designed in a manner that performance is only minimally affected by the developmental stage. Moreover, experiments and tasks have to be limited in time, reducing the amount of data that can be acquired for an individual. Having fewer within-subject data reduces the statistical power of the experiment, and, together with modest effect sizes emphasizes the need for future large-scale studies. A larger scale can help to improve replicability, especially when performed across multiple sites (McShane, Tackett, Böckenholt, & Gelman, 2019; Nuijten, 2019) and represents one step forward to improve the replication crisis in the fields of psychology and cognitive neuroscience (Button et al., 2013).

A second limitation of this work is the potential confound of *treatment* in youth with major depressive disorder. Most treated participants received selective serotonin reuptake inhibitors that are known to affect the brain response in neural circuits supporting processing of incentives (Herzallah et al., 2013; McCabe et al., 2010) and emotion (Sladky, Spies, et al., 2015). Furthermore, due to the scarcity of replication studies that tested pre- and post-treatment effects, it is unclear whether drug effects normalize aberrant brain function in an affected neural circuit or aid compensation in another. Therefore, for a better understanding of the true disease state, future analyses should attempt to recruit unmedicated patients to uncover disorder related effects to reduce confounds by any acute or long-term medication.

Thirdly, in Chapter 2, we examined the development from adolescence into adulthood *treating age as continuous* variable. This deviates from previous work, which compared different stages of development based on age groups. The latter faces the challenge that it is not clear which age thresholds to set for distinguishing children and adolescents, or when adulthood begins (Casey, 2019). With our approach, we were able to test the measures of interest with several developmental models (e.g. linear, quadratic or inverse), without relying on arbitrary age bounds for different developmental phases. In this work, participants from the age of 11 until 35 were enrolled. While this broad age range is much larger than previously studies have used (Van Den Bos et al., 2012; Van Den Bos et al., 2015), and allow for an investigation of brain maturation up to adulthood, the lower bound of 11 precludes any inference on the development in earlier phases. Thus, future work examining earlier development up to young adulthood is needed to shed

light on the transient maturation from childhood into adolescence and eventually adulthood.

Fourth, in the studies involving adolescents, *age represented the sole proxy for the developmental stage*. While the age range was well-matched between the MDD group and healthy controls, not taking into account the gonadal hormones and pubertal development might obscure important driving factors for behavior and brain physiology (Varlinskaya, Vetter-O'Hagen, & Spear, 2013). Evidence highly suggests that gonadal hormones drive maturation of brain structure and function across puberty (Piekarski et al., 2017). Moreover, the timing of pubertal development seems to have a great variability within the population (Sisk & Zehr, 2005). Interactions of gonadal hormones and the environment might also explain sex differences of neurobiological mechanisms in major depression (Jenkins et al., 2018). Given this limitation, future studies should include fine-grained measures of pubertal development to assess the effect on brain maturation and behavior.

Lastly, all studies presented in this thesis employed a *cross-sectional study design* to investigate neurodevelopment from adolescence into adulthood, as well as group differences between adolescents with and without MDD. However, this design does not allow to establish causal relationships, nor is it able to reveal true developmental trajectories. In the context of brain maturation, a longitudinal study design allows for investigating genuine brain development by separating developmental from cohort effects (Telzer et al., 2018). Furthermore, it enables to characterize behavioral and physiological transitions or the lack thereof and allows the assessment of causal relationships between events across time. In the psychiatric context, cross-sectional studies only allow to establish potential biomarkers for a disease state. Here, it is important to note, that any association cannot imply causation. To facilitate clinical utility and prediction of clinical trajectories and treatment outcome, the viability of any biomarker needs to be tested in longitudinal designs (Abi-Dargham & Horga, 2016).

Taken together, the above-mentioned factors constrain the interpretation of the overall findings of this thesis. Despite those limitations, the present thesis is a significant step forward in building the foundation of a methodological framework for advancing the current understanding of neurodevelopment and neuropsychiatric processes. In the next section, possible directions for future studies are discussed to tackle open research questions in psychiatric research.

6.2 Future directions and challenges in studying pediatric major depressive disorder

Since the advent of neuroimaging many studies have attempted to find a common, underlying neurobiological biomarker that can inform about psychiatric psychopathologies. Research has focused on establishing differences between healthy controls and patients. These studies were very important to reveal candidate processes and neural cir-

cuits related to psychopathology, however, hitherto this approach was not very successful to develop useful tools for clinical practice. An inherent challenge to achieve this goal and hence for the identification of a reliable biomarker for any psychiatric disorder is the heterogeneity of symptoms and their possible combination under the same diagnostic label. Neurobiological disease mechanisms underlying psychiatric disorders do not necessarily adhere to the current clinical nosology, and there might be several distinct mechanisms leading to the same symptom. Pursuing a search for a common biological correlate of a diagnostic label, ignores the fact the same symptoms might arise through different biological pathologies. Thus, this approach has only limited potential to extend existing nosology. In contrast to symptom-based approaches, the Research Domain Criteria (RDoC, T. R. Insel, 2014) represents a multidimensional, transdiagnostic approach to understand mental disorders, where functional constructs (e.g. approach motivation) are studied and integrated across multiple levels of analysis (e.g. genetics, circuits, behavior). Within this research framework, computational psychiatry has the potential to provide tools that give insights to expand the knowledge of disease mechanism to bridge neuroscience and psychiatry, as recently emphasized by the National Institute of Mental Health (Ferrante et al., 2019). This thesis set a particular focus to extend the understanding of the neurobiological mechanisms associated with pediatric MDD. To enable the use for clinical practice, the field of neuroscience will have to identify reliable neurobiological markers predictive of treatment outcome and risk determination to improve prevention. To this end, the following sections will discuss open questions and suggestions for extensions for future studies and shed light on the importance of integrating modalities to advance the understanding of the susceptibility window to psychopathologies across development.

6.2.1 Beyond task-based neuroimaging: resting-state & structural data

The main modality employed throughout this work was task-based BOLD-weighted MRI. Nevertheless, other MRI modalities might give complementary insights into disrupted brain circuits in psychopathology. For instance, resting-state fMRI constitutes a powerful tool to examine brain function in mental disorders, which measures spontaneous BOLD fluctuations during rest (Biswal, 2012). As it is independent of task performance or compliance, it is a useful tool to study brain function in populations for whom task-based fMRI is cumbersome or not feasible such as elderly, or individuals with neuropsychiatric disorders. Resting-state data can be acquired within consortia across multiple sites to collect large samples. In adolescent MDD, first studies have shown atypical functional connectivity within the prefrontal-amygdala network (Connolly et al., 2013; Cullen et al., 2014; Davey, Harrison, Yücel, & Allen, 2012; Gabbay et al., 2013; Luking et al., 2011). Interestingly, previous reports showed opposing effects in resting-state functional connectivity patterns between adolescent and adult depression, which suggests that developmental processes or illness-related compensatory mechanisms affect the FC between striatum and the prefrontal cortex across lifespan (Furman, Hamilton, & Gotlib, 2011). Furthermore, based on resting-state functional connectivity one study was able to

determine biotypes of adult MDD differentiating the patients in responders and non-responses to stimulation treatment (Drysdale et al., 2017).

In addition, converging evidence from several diffusion tensor imaging studies report *structural* abnormalities of white matter integrity within the affective network in MDD. Lower fractional anisotropy in thalamic, temporal and prefrontal projection fibers have been consistently found in adults (G. Chen et al., 2016; Sexton, Mackay, & Ebmeier, 2009) and adolescents with a history of (Cullen et al., 2010) or at risk for MDD (Huang, Fan, Williamson, & Rao, 2011). The latter supports the hypothesis that white matter deficits are not a result of chronic depression or its treatment, but emerge in early puberty. A combined longitudinal cohort study investigating resting-state functional and structural connectivity showed differential sensitivity to predict depressive symptoms with both approaches across development, demonstrating that assessing functional and structural data could complement each other for prediction (Jalbrzikowski et al., 2017). Future studies will have to characterize white matter tracts in earlier developmental stages and how they are related to functional abnormalities further. A particular application for this is to integrate structural data with models of effective connectivity, which has been shown to improve reliability of results (Sokolov et al., 2019; Stephan, Tittgemeyer, Knösche, Moran, & Friston, 2009), which is of particular importance to aid replicability.

The studies discussed above, provide fair evidence that besides functional disruptions, the combination of multiple modalities could improve the distinction of healthy from psychopathologies based on neurobiological data (Gao, Calhoun, & Sui, 2018). Combining multiple modalities usually leads to high-dimensional data sets to characterize an individual. Thus, a recent challenge in research is to develop novel approaches to reduce the dimensions of features to advance clinical utility (e.g. subtyping of patients or predict clinical trajectories) and foster the understanding of disease mechanisms.

6.2.2 Subtyping and prediction with multimodal approaches and machine learning

Methodological advancements have ushered in the advent of data-driven approaches (e.g. machine learning analyses) to aid clinically relevant challenges, such as computer aided diagnosis or prediction of treatment outcome to guide selection of medication and the longitudinal course of disorders (Huys et al., 2016). For instance, a particular challenge for the clinician is usually not to decide if a patient should receive treatment, but rather which treatment is the most appropriate. Thus, first studies have employed machine learning to identify biological patterns with predictive value for treatment selection. These approaches have the potential to extend existing nosology based on clinical useful typologies. To classify depression patients from functional neuroimaging data, pattern classification approaches have been applied to task-based (Fu et al., 2008) and resting-state fMRI (Craddock, Holtzheimer III, Hu, & Mayberg, 2009). In a prospective study with adolescents, prediction of first onset of MDD has been performed successfully

based on structural brain images (Foland-Ross et al., 2015). Moreover, other studies successfully predicted treatment response to escitalopram (Zhdanov et al., 2020) and transcranial magnetic stimulation (Drysdale et al., 2017). That shows that data-driven models can be useful to answer critical clinical questions by harnessing complex, high-dimensional data sets. The advantages notwithstanding, due to the complexity and non-linearity of the models, data-driven approaches often do not lend themselves to a straight-forward interpretation (Haufe et al., 2014). However, mechanistic understanding is critical for trusting the model predictions for clinical usage or to identify biases in the model (Walter et al., 2019).

Therefore, in contrast to the data-driven approaches, others rely heavily on theoretical groundwork and often provide predictions about mechanisms of behavior. This can help to reduce the huge dimensionality of, for instance, neuroimaging data, to few, but informative and more easily interpretable features. Models built under this premise have meaningful parameters facilitating its comprehensibility and accessibility for non-experts. For instance, on an algorithmic level (Marr, 1982), such models might have their foundation in reinforcement learning theory and are able to make strong predictions about computational mechanisms. Computational models of behavior have been shown to be potentially clinically useful. Recent work has shown that such models are able to predict the outcome of cognitive behavioral therapy based on neural activity related to prediction errors derived from a reinforcement learning model (Queirazza, Fouragnan, Steele, Cavanagh, & Philiastides, 2019). In a different study, the authors were able to predict relapse after antidepressant discontinuation by applying computational modeling to a physical effort task (Berwian et al., 2020). On the implementation level, there are models that try to relate dynamics in neural networks to an observed BOLD signal using a biophysical plausible model (DCM; Friston et al., 2003). Developing and refining models on the different levels of analysis is an active field of research. Here, it is important to note, that the levels of description are not always sharply separated. As shown in this thesis, DCMs that model multivariate timeseries of brain regions innervated by dopaminergic projections might incorporate prediction error signals derived from a reinforcement-learning model to inform the other.

In psychiatric research, theoretical approaches had great success in advancing our understanding of how maladaptive behavior is generated and manifested in the brain. More recently, researchers increased their efforts to combine both theory- and data-driven approaches. When theoretical knowledge about a brain mechanism is available, recent work indicates that data-driven machine learning approaches benefit from a priori selection of informative features as input. Hence, after reducing the dimensionality of a data set to a few mechanistically interpretable features, machine learning can be used on a reduced set to generate clinical useful predictions (Huys et al., 2016). A promising avenue that aims at fusing theoretical and data-driven approaches is *generative embedding* (Brodersen et al., 2011). Here, generative models, most prominently DCMs, are used to

make clinically useful predictions with mechanistically interpretable features (e.g. effective connectivity). It has been demonstrated that this procedure has external validity in schizophrenia (Brodersen et al., 2014). Moreover, generative embedding was successfully applied to predict clinical trajectories of patients with MDD from brain connectivity during emotional face processing (Frässle et al., 2020). Importantly, because the DCM parameters have a particular meaning within the dynamic brain system it models, the authors could reveal that the modulation of the coupling of the amygdala and face perception areas was most predictive for the course of the disorder, thus allowing for a straight-forward, mechanistic interpretation.

Altogether, this shows that first studies have emerged that employed large-sample classification for (1) prediction of onset of MDD, (2) subtyping and diagnosis, and (3) prediction of treatment outcome. The work presented in the thesis represents a significant step forward in the identification of plausible mechanistic models of brain function that will be vital for future integration of different neuroimaging modalities and appropriate dimension reduction to meaningful parameters. In tandem with longitudinal designs, this holds promise to develop more reliable biomarkers with improved clinical utility.

6.3 Conclusion

The focus of the thesis was the development and application of integrated methods to study behavior in adolescents with typical and aberrant development in MDD within the scope of incentive and emotion processing.

The findings presented here provide novel insights to a better understanding of the functional neurodevelopment, and extend previous work of the emergence of cognitive control across adolescence significantly. Moreover, we determined that aberrant neural processing of monetary loss and inefficient processing of ambiguous social cues might form an early marker in adolescent MDD. In addition, we detected a valence-dependent functional pathway within the prefrontal-amygdala network in healthy adults that can serve as a target in future to investigate negative bias in MDD. Taken together, these results demonstrate behavioral and neural differences in critical cognitive functions in adolescent MDD that pave the way for the development of novel diagnostic and therapeutic tools. Besides the mechanistic insights into adolescent MDD, the present thesis provides a methodological framework for computational psychiatry to improve the understanding of neuropsychiatric disorders in the future.

Finally, to achieve clinical utility of these findings, future longitudinal studies should investigate how neurobiological factors during childhood and adolescence contribute to the emergence of depressive symptoms. A comprehensive understanding of the underlying neurobiological mechanisms of how and why MDD emerges in youth might result in novel opportunities to intervene early on the trajectory of one of the most debilitating psychiatric disorders.

Bibliography

Abi-Dargham, A., & Horga, G. (2016). The search for imaging biomarkers in psychiatric disorders. *Nature medicine, 22*(11), 1248.

Adolphs, R. (2002). Recognizing emotion from facial expressions: psychological and neurological mechanisms. *Behavioral and cognitive neuroscience reviews, 1*(1), 21-62.

Almeida, J., Kronhaus, D., Sibille, E., Langenecker, S., Versace, A., LaBarbara, E., & Phillips, M. L. (2011). Abnormal left-sided orbitomedial prefrontal cortical–amygdala connectivity during happy and fear face processing: a potential neural mechanism of female MDD. *Frontiers in psychiatry, 2*, 69.

Almeida, J., Mechelli, A., Hassel, S., Versace, A., Kupfer, D. J., & Phillips, M. L. (2009). Abnormally increased effective connectivity between parahippocampal gyrus and ventromedial prefrontal regions during emotion labeling in bipolar disorder. *Psychiatry Research: Neuroimaging, 174*(3), 195-201.

Almeida, J., Versace, A., Mechelli, A., Hassel, S., Quevedo, K., Kupfer, D. J., & Phillips, M. L. (2009). Abnormal amygdala-prefrontal effective connectivity to happy faces differentiates bipolar from major depression. *Biol Psychiatry, 66*(5), 451-459.

American Psychiatric Association. (2013). *Diagnostic and statistical manual of mental disorders (DSM-5®)*: American Psychiatric Pub.

Andersen, S. L., & Teicher, M. H. (2008). Stress, sensitive periods and maturational events in adolescent depression. *Trends in Neurosciences, 31*(4), 183-191. doi:10.1016/j.tins.2008.01.004

Anderson, A. K., Christoff, K., Panitz, D., De Rosa, E., & Gabrieli, J. D. (2003). Neural correlates of the automatic processing of threat facial signals. *J Neurosci, 23*(13), 5627-5633. doi:10.1523/jneurosci.23-13-05627.2003

Anderson, M. C., Ochsner, K. N., Kuhl, B., Cooper, J., Robertson, E., Gabrieli, S. W., Glover, G. H., & Gabrieli, J. D. (2004). Neural systems underlying the suppression of unwanted memories. *Science, 303*(5655), 232-235.

Ando, T. (2011). Predictive Bayesian model selection. *American Journal of Mathematical and Management Sciences, 31*(1-2), 13-38.

Aronson, J. K., & Ferner, R. E. (2017). Biomarkers—a general review. *Current protocols in pharmacology, 76*(1), 9.23. 21-29.23. 17.

Arsalidou, M., Morris, D., & Taylor, M. J. (2011). Converging evidence for the advantage of dynamic facial expressions. *Brain topography, 24*(2), 149-163.

Avenevoli, S., Swendsen, J., He, J.-P., Burstein, M., & Merikangas, K. R. (2015). Major depression in the National Comorbidity Survey–Adolescent Supplement: prevalence, correlates, and treatment. *Journal of the American Academy of Child & Adolescent Psychiatry, 54*(1), 37-44. e32.

Averbeck, B. B., & Costa, V. D. (2017). Motivational neural circuits underlying reinforcement learning. *Nature Neuroscience, 20*(4), 505.

Aybek, S., Nicholson, T. R., O'Daly, O., Zelaya, F., Kanaan, R. A., & David, A. S. (2015). Emotion-motion interactions in conversion disorder: an FMRI study. *PLoS One, 10*(4), e0123273.

Badcock, P. B., Davey, C. G., Whittle, S., Allen, N. B., & Friston, K. J. (2017). The depressed brain: an evolutionary systems theory. *Trends in Cognitive Sciences, 21*(3), 182-194.

Ball, T., Derix, J., Wentlandt, J., Wieckhorst, B., Speck, O., Schulze-Bonhage, A., & Mutschler, I. (2009). Anatomical specificity of functional amygdala imaging of responses to stimuli with positive and negative emotional valence. *Journal of neuroscience methods, 180*(1), 57-70.

Ballard, I. C., Murty, V. P., Carter, R. M., MacInnes, J. J., Huettel, S. A., & Adcock, R. A. (2011). Dorsolateral prefrontal cortex drives mesolimbic dopaminergic regions to initiate motivated behavior. *Journal of Neuroscience, 31*(28), 10340-10346.

Barkley-Levenson, E., & Galván, A. (2014). Neural representation of expected value in the adolescent brain. *Proceedings of the National Academy of Sciences, 111*(4), 1646-1651.

Barkley-Levenson, E., Van Leijenhorst, L., & Galván, A. (2013). Behavioral and neural correlates of loss aversion and risk avoidance in adolescents and adults. *Developmental cognitive neuroscience, 3*, 72-83.

Barrett, L. F., & Simmons, W. K. (2015). Interoceptive predictions in the brain. *Nature Reviews Neuroscience, 16*(7), 419.

Barron, H. C., Garvert, M. M., & Behrens, T. E. (2015). Reassessing VMPFC: full of confidence? *Nature neuroscience, 18*(8), 1064.

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

Beck, A. T. (1967). *Depression: Clinical, experimental, and theoretical aspects*: Hoeber Medical Division, Harper & Row.

Behrens, T. E., Woolrich, M. W., Walton, M. E., & Rushworth, M. F. (2007). Learning the value of information in an uncertain world. *Nature neuroscience, 10*(9), 1214-1221.

Beierholm, U., Guitart-Masip, M., Economides, M., Chowdhury, R., Düzel, E., Dolan, R., & Dayan, P. (2013). Dopamine modulates reward-related vigor. *Neuropsychopharmacology, 38*(8), 1495.

Benjamini, Y., & Hochberg, Y. (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J R Stat Soc Series B Stat Methodol*, 289-300.

Berwian, I. M., Wenzel, J. G., Collins, A. G., Seifritz, E., Stephan, K. E., Walter, H., & Huys, Q. J. (2020). Computational mechanisms of effort and reward decisions in patients with depression and their association with relapse after antidepressant discontinuation. *Jama Psychiatry, 77*(5), 513-522.

Bestmann, S., Ruge, D., Rothwell, J., & Galea, J. M. (2014). The role of dopamine in motor flexibility. *Journal of cognitive neuroscience, 27*(2), 365-376.

Bishop, S., Duncan, J., Brett, M., & Lawrence, A. D. (2004). Prefrontal cortical function and anxiety: controlling attention to threat-related stimuli. *Nature neuroscience, 7*(2), 184.

Bistricky, S. L., Ingram, R. E., & Atchley, R. A. (2011). Facial affect processing and depression susceptibility: Cognitive biases and cognitive neuroscience. *Psychological Bulletin, 137*(6), 998.

Biswal, B. B. (2012). Resting state fMRI: a personal history. *Neuroimage, 62*(2), 938-944.

Bjork, J. M., Smith, A. R., Chen, G., & Hommer, D. W. (2010). Adolescents, adults and rewards: comparing motivational neurocircuitry recruitment using fMRI. *PloS one, 5*(7), e11440.

Björklund, A., & Dunnett, S. B. (2007). Dopamine neuron systems in the brain: an update. *Trends in neurosciences, 30*(5), 194-202.

Botvinick, M., & Braver, T. (2015). Motivation and cognitive control: from behavior to neural mechanism. *Annual review of psychology, 66*, 83-113.

Braams, B. R., Peters, S., Peper, J. S., Güroğlu, B., & Crone, E. A. (2014). Gambling for self, friends, and antagonists: differential contributions of affective and social brain regions on adolescent reward processing. *Neuroimage, 100*, 281-289.

Bright, T., Felix, L., Kuper, H., & Polack, S. (2018). Systematic review of strategies to increase access to health services among children over five in low-and middle-income countries. *Tropical Medicine & International Health, 23*(5), 476-507.

Brodersen, K. H., Deserno, L., Schlagenhauf, F., Lin, Z., Penny, W. D., Buhmann, J. M., & Stephan, K. E. (2014). Dissecting psychiatric spectrum disorders by generative embedding. *NeuroImage: Clinical, 4*, 98-111.

Brodersen, K. H., Schofield, T. M., Leff, A. P., Ong, C. S., Lomakina, E. I., Buhmann, J. M., & Stephan, K. E. (2011). Generative embedding for model-based classification of fMRI data. *PLoS Comput Biol, 7*(6), e1002079.

Brown, S. D., & Heathcote, A. (2008). The simplest complete model of choice response time: Linear ballistic accumulation. *Cognitive psychology, 57*(3), 153-178.

Browning, M., Behrens, T. E., Jocham, G., O'reilly, J. X., & Bishop, S. J. (2015). Anxious individuals have difficulty learning the causal statistics of aversive environments. *Nature neuroscience, 18*(4), 590-596.

Bubic, A., Von Cramon, D. Y., & Schubotz, R. I. (2010). Prediction, cognition and the brain. *Frontiers in human neuroscience, 4*, 25.

Buckholtz, J. W., & Meyer-Lindenberg, A. (2012). Psychopathology and the human connectome: toward a transdiagnostic model of risk for mental illness. *Neuron, 74*(6), 990-1004.

Bunzeck, N., & Düzel, E. (2006). Absolute coding of stimulus novelty in the human substantia nigra/VTA. *Neuron, 51*(3), 369-379.

Button, K. S., Ioannidis, J. P., Mokrysz, C., Nosek, B. A., Flint, J., Robinson, E. S., & Munafò, M. R. (2013). Power failure: why small sample size undermines the reliability of neuroscience. *Nature reviews neuroscience, 14*(5), 365-376.

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

Cao, Z., Bennett, M., Orr, C., Icke, I., Banaschewski, T., Barker, G. J., Bokde, A. L., Bromberg, U., Büchel, C., & Quinlan, E. B. (2019). Mapping adolescent reward anticipation, receipt, and prediction error during the monetary incentive delay task. *Human brain mapping, 40*(1), 262-283.

Casey, B. J. (2019). Healthy development as a human right: lessons from developmental science. *Neuron, 102*(4), 724-727.

Casey, B. J., Giedd, J. N., & Thomas, K. M. (2000). Structural and functional brain development and its relation to cognitive development. *Biol Psychol, 54*(1-3), 241-257.

Casey, B. J., Jones, R. M., & Hare, T. A. (2008). The adolescent brain. *Ann N Y Acad Sci, 1124*, 111-126. doi:10.1196/annals.1440.010

Casey, B. J., Jones, R. M., & Somerville, L. H. (2011). Braking and accelerating of the adolescent brain. *Journal of Research on Adolescence, 21*(1), 21-33.

Cavanagh, J. F., & Frank, M. J. (2014). Frontal theta as a mechanism for cognitive control. *Trends in cognitive sciences, 18*(8), 414-421.

Cha, C. B., Franz, P. J., M. Guzmán, E., Glenn, C. R., Kleiman, E. M., & Nock, M. K. (2018). Annual Research Review: Suicide among youth–epidemiology,(potential) etiology, and treatment. *Journal of Child Psychology and psychiatry, 59*(4), 460-482.

Chai, X. J., Hirshfeld-Becker, D., Biederman, J., Uchida, M., Doehrmann, O., Leonard, J. A., Salvatore, J., Kenworthy, T., Brown, A., & Kagan, E. (2015). Functional and structural brain correlates of risk for major depression in children with familial depression. *NeuroImage: Clinical, 8*, 398-407.

Chakraborty, S., Kolling, N., Walton, M. E., & Mitchell, A. S. (2016). Critical role for the mediodorsal thalamus in permitting rapid reward-guided updating in stochastic reward environments. *Elife, 5*, e13588.

Chen, C., Takahashi, T., Nakagawa, S., Inoue, T., & Kusumi, I. (2015). Reinforcement learning in depression: a review of computational research. *Neuroscience & Biobehavioral Reviews, 55*, 247-267.

Chen, G., Hu, X., Li, L., Huang, X., Lui, S., Kuang, W., Ai, H., Bi, F., Gu, Z., & Gong, Q. (2016). Disorganization of white matter architecture in major depressive disorder: a meta-analysis of diffusion tensor imaging with tract-based spatial statistics. *Scientific reports, 6*(1), 1-11.

Chiew, K. S., & Braver, T. S. (2016). Reward favors the prepared: Incentive and task-informative cues interact to enhance attentional control. *Journal of experimental psychology: human perception and performance, 42*(1), 52.

Cho, Y. T., Fromm, S., Guyer, A. E., Detloff, A., Pine, D. S., Fudge, J. L., & Ernst, M. (2013). Nucleus accumbens, thalamus and insula connectivity during incentive anticipation in typical adults and adolescents. *Neuroimage, 66*, 508-521.

Clark, J. E., Watson, S., & Friston, K. J. (2018). What is mood? A computational perspective. *Psychological Medicine, 48*(14), 2277-2284.

Cohen, J. R., Asarnow, R. F., Sabb, F. W., Bilder, R. M., Bookheimer, S. Y., Knowlton, B. J., & Poldrack, R. A. (2010). A unique adolescent response to reward prediction errors. *Nature neuroscience, 13*(6), 669.

Comte, M., Schon, D., Coull, J. T., Reynaud, E., Khalfa, S., Belzeaux, R., Ibrahim, E. C., Guedj, E., Blin, O., Weinberger, D. R., & Fakra, E. (2016). Dissociating Bottom-Up and Top-Down Mechanisms in the Cortico-Limbic System during Emotion Processing. *Cerebral Cortex, 26*(1), 144-155. doi:10.1093/cercor/bhu185

Connolly, C. G., Wu, J., Ho, T. C., Hoeft, F., Wolkowitz, O., Eisendrath, S., Frank, G., Hendren, R., Max, J. E., & Paulus, M. P. (2013). Resting-state functional connectivity of subgenual anterior cingulate cortex in depressed adolescents. *Biological psychiatry, 74*(12), 898-907.

Craddock, R. C., Holtzheimer III, P. E., Hu, X. P., & Mayberg, H. S. (2009). Disease state prediction from resting state functional connectivity. *Magnetic Resonance in Medicine: An Official Journal of the International Society for Magnetic Resonance in Medicine, 62*(6), 1619-1628.

Crone, E. A., & Dahl, R. E. (2012). Understanding adolescence as a period of social–affective engagement and goal flexibility. *Nature Reviews Neuroscience, 13*(9), 636.

Cullen, K. R., Klimes-Dougan, B., Muetzel, R., Mueller, B. A., Camchong, J., Houri, A., Kurma, S., & Lim, K. O. (2010). Altered white matter microstructure in adolescents with major depression: a preliminary study. *Journal of the American Academy of Child & Adolescent Psychiatry, 49*(2), 173-183. e171.

Cullen, K. R., Westlund, M. K., Klimes-Dougan, B., Mueller, B. A., Houri, A., Eberly, L. E., & Lim, K. O. (2014). Abnormal amygdala resting-state functional connectivity in adolescent depression. *JAMA psychiatry, 71*(10), 1138-1147.

Cuthbert, B. N., & Insel, T. R. (2013). Toward the future of psychiatric diagnosis: the seven pillars of RDoC. *BMC medicine, 11*(1), 126.

Dannlowski, U., Ohrmann, P., Bauer, J., Kugel, H., Arolt, V., Heindel, W., Kersting, A., Baune, B. T., & Suslow, T. (2007). Amygdala reactivity to masked negative faces is associated with automatic judgmental bias in major depression: a 3 T fMRI study. *Journal of psychiatry & neuroscience: JPN, 32*(6), 423.

Davey, C. G., Harrison, B. J., Yücel, M., & Allen, N. B. (2012). Regionally specific alterations in functional connectivity of the anterior cingulate cortex in major depressive disorder.

Davidow, J. Y., Insel, C., & Somerville, L. H. (2018). Adolescent development of value-guided goal pursuit. *Trends in cognitive sciences, 22*(8), 725-736.

Dayan, P. (2012). Instrumental vigour in punishment and reward. *European Journal of Neuroscience, 35*(7), 1152-1168.

Delevich, K., Tucciarone, J., Huang, Z. J., & Li, B. (2015). The mediodorsal thalamus drives feedforward inhibition in the anterior cingulate cortex via parvalbumin interneurons. *Journal of Neuroscience, 35*(14), 5743-5753.

Delgado, M. R., Nearing, K. I., LeDoux, J. E., & Phelps, E. A. (2008). Neural circuitry underlying the regulation of conditioned fear and its relation to extinction. *Neuron, 59*(5), 829-838.

Demenescu, L., Kortekaas, R., Cremers, H., Renken, R., van Tol, M., van der Wee, N., Veltman, D., den Boer, J., Roelofs, K., & Aleman, A. (2013). Amygdala activation and its functional connectivity during perception of emotional faces in social phobia and panic disorder. *Journal of psychiatric research, 47*(8), 1024-1031.

den Ouden, H. E., Daunizeau, J., Roiser, J., Friston, K. J., & Stephan, K. E. (2010). Striatal prediction error modulates cortical coupling. *Journal of Neuroscience, 30*(9), 3210-3219.

Der-Avakian, A., & Markou, A. (2012). The neurobiology of anhedonia and other reward-related deficits. *Trends in neurosciences, 35*(1), 68-77.

Dichter, G. S., Felder, J. N., & Smoski, M. J. (2009). Affective context interferes with cognitive control in unipolar depression: an fMRI investigation. *J Affect Disord, 114*(1-3), 131-142. doi:10.1016/j.jad.2008.06.027

Dima, D., Stephan, K. E., Roiser, J. P., Friston, K. J., & Frangou, S. (2011). Effective connectivity during processing of facial affect: evidence for multiple parallel pathways. *Journal of Neuroscience, 31*(40), 14378-14385. doi:10.1523/JNEUROSCI.2400-11.2011

Disner, S. G., Beevers, C. G., Haigh, E. A., & Beck, A. T. (2011). Neural mechanisms of the cognitive model of depression. *Nature Reviews Neuroscience, 12*(8), 467-477.

Dixon, M. L., Thiruchselvam, R., Todd, R., & Christoff, K. (2017). Emotion and the prefrontal cortex: an integrative review. *Psychological bulletin, 143*(10), 1033.

Dombrovski, A. Y., Clark, L., Siegle, G. J., Butters, M. A., Ichikawa, N., Sahakian, B. J., & Szanto, K. (2010). Reward/punishment reversal learning in older suicide attempters. *American Journal of Psychiatry, 167*(6), 699-707.

Dombrovski, A. Y., Szanto, K., Clark, L., Reynolds, C. F., & Siegle, G. J. (2013). Reward signals, attempted suicide, and impulsivity in late-life depression. *JAMA psychiatry, 70*(10), 1020-1030.

Drabant, E. M., McRae, K., Manuck, S. B., Hariri, A. R., & Gross, J. J. (2009). Individual differences in typical reappraisal use predict amygdala and prefrontal responses. *Biological psychiatry, 65*(5), 367-373.

Drevets, W. C., Savitz, J., & Trimble, M. (2008). The subgenual anterior cingulate cortex in mood disorders. *CNS spectrums, 13*(8), 663.

Drysdale, A. T., Grosenick, L., Downar, J., Dunlop, K., Mansouri, F., Meng, Y., Fetcho, R. N., Zebley, B., Oathes, D. J., & Etkin, A. (2017). Resting-state connectivity biomarkers define neurophysiological subtypes of depression. *Nature medicine, 23*(1), 28-38.

Dudman, J. T., & Krakauer, J. W. (2016). The basal ganglia: from motor commands to the control of vigor. *Current opinion in neurobiology, 37*, 158-166.

Eickhoff, S. B., Paus, T., Caspers, S., Grosbras, M.-H., Evans, A. C., Zilles, K., & Amunts, K. (2007). Assignment of functional activations to probabilistic cytoarchitectonic areas revisited. *Neuroimage, 36*(3), 511-521.

Emslie, G. J., Mayes, T. L., & Ruberu, M. (2005). Continuation and maintenance therapy of early-onset major depressive disorder. *Paediatr Drugs, 7*(4), 203-217.

Ernst, M., Pine, D. S., & Hardin, M. (2006). Triadic model of the neurobiology of motivated behavior in adolescence. *Psychological medicine, 36*(3), 299-312.

Etkin, A., Buchel, C., & Gross, J. J. (2015). The neural bases of emotion regulation. *Nat Rev Neurosci, 16*(11), 693-700. doi:10.1038/nrn4044

Etkin, A., Prater, K. E., Hoeft, F., Menon, V., & Schatzberg, A. F. (2010). Failure of anterior cingulate activation and connectivity with the amygdala during implicit regulation of emotional processing in generalized anxiety disorder. *American Journal of Psychiatry, 167*(5), 545-554.

Fairhall, S. L., & Ishai, A. (2006). Effective connectivity within the distributed cortical network for face perception. *Cerebral cortex, 17*(10), 2400-2406.

Fergusson, D. M., & Woodward, L. J. (2002). Mental health, educational, and social role outcomes of adolescents with depression. *Archives of general psychiatry, 59*(3), 225-231.

Ferrante, M., Redish, A. D., Oquendo, M. A., Averbeck, B. B., Kinnane, M. E., & Gordon, J. A. (2019). Computational psychiatry: a report from the 2017 NIMH workshop on opportunities and challenges. *Molecular psychiatry, 24*(4), 479.

Foland-Ross, L. C., & Gotlib, I. H. (2012). Cognitive and neural aspects of information processing in major depressive disorder: an integrative perspective. *Frontiers in psychology, 3*, 489.

Foland-Ross, L. C., Sacchet, M. D., Prasad, G., Gilbert, B., Thompson, P. M., & Gotlib, I. H. (2015). Cortical thickness predicts the first onset of major depression in adolescence. *International Journal of Developmental Neuroscience, 46*, 125-131.

Forbes, E. E., Hariri, A. R., Martin, S. L., Silk, J. S., Moyles, D. L., Fisher, P. M., Brown, S. M., Ryan, N. D., Birmaher, B., & Axelson, D. A. (2009). Altered striatal activation predicting real-world positive affect in adolescent major depressive disorder. *American Journal of Psychiatry, 166*(1), 64-73.

Forbes, E. E., Ryan, N. D., Phillips, M. L., Manuck, S. B., Worthman, C. M., Moyles, D. L., Tarr, J. A., Sciarrillo, S. R., & Dahl, R. E. (2010). Healthy adolescents' neural response to reward: associations with puberty, positive affect, and depressive symptoms. *Journal of the American Academy of Child & Adolescent Psychiatry, 49*(2), 162-172. e165.

Forstmann, B. U., Anwander, A., Schäfer, A., Neumann, J., Brown, S., Wagenmakers, E.-J., Bogacz, R., & Turner, R. (2010). Cortico-striatal connections predict control over speed and accuracy in perceptual decision making. *Proceedings of the National Academy of Sciences, 107*(36), 15916-15920.

Frässle, S., Marquand, A. F., Schmaal, L., Dinga, R., Veltman, D. J., Van der Wee, N. J., van Tol, M.-J., Schöbi, D., Penninx, B. W., & Stephan, K. E. (2020). Predicting

individual clinical trajectories of depression with generative embedding. *NeuroImage: Clinical, 26*, 102213.

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

Friston, K. J., Holmes, A. P., Price, C., Büchel, C., & Worsley, K. (1999). Multisubject fMRI studies and conjunction analyses. *Neuroimage, 10*(4), 385-396.

Friston, K. J., Kahan, J., Biswal, B., & Razi, A. (2014). A DCM for resting state fMRI. *Neuroimage, 94*, 396-407.

Friston, K. J., Litvak, V., Oswal, A., Razi, A., Stephan, K. E., Van Wijk, B. C., Ziegler, G., & Zeidman, P. (2016). Bayesian model reduction and empirical Bayes for group (DCM) studies. *Neuroimage, 128*, 413-431.

Fu, C. H., Mourao-Miranda, J., Costafreda, S. G., Khanna, A., Marquand, A. F., Williams, S. C., & Brammer, M. J. (2008). Pattern classification of sad facial processing: toward the development of neurobiological markers in depression. *Biological psychiatry, 63*(7), 656-662.

Furl, N., Henson, R. N., Friston, K. J., & Calder, A. J. (2014). Network interactions explain sensitivity to dynamic faces in the superior temporal sulcus. *Cereb Cortex, 25*(9), 2876-2882.

Furman, D. J., Hamilton, J. P., & Gotlib, I. H. (2011). Frontostriatal functional connectivity in major depressive disorder. *Biology of mood & anxiety disorders, 1*(1), 11.

Fusar-Poli, P., Placentino, A., Carletti, F., Landi, P., Allen, P., Surguladze, S., Benedetti, F., Abbamonte, M., Gasparotti, R., & Barale, F. (2009). Functional atlas of emotional faces processing: a voxel-based meta-analysis of 105 functional magnetic resonance imaging studies. *Journal of psychiatry & neuroscience*.

Gabbay, V., Ely, B. A., Li, Q., Bangaru, S. D., Panzer, A. M., Alonso, C. M., Castellanos, F. X., & Milham, M. P. (2013). Striatum-based circuitry of adolescent depression and anhedonia. *Journal of the American Academy of Child & Adolescent Psychiatry, 52*(6), 628-641. e613.

Gaffrey, M. S., Barch, D. M., Singer, J., Shenoy, R., & Luby, J. L. (2013). Disrupted amygdala reactivity in depressed 4-to 6-year-old children. *Journal of the American Academy of Child & Adolescent Psychiatry, 52*(7), 737-746.

Galván, A., & McGlennen, K. M. (2013). Enhanced striatal sensitivity to aversive reinforcement in adolescents versus adults. *Journal of cognitive neuroscience, 25*(2), 284-296.

Gao, S., Calhoun, V. D., & Sui, J. (2018). Machine learning in major depression: From classification to treatment outcome prediction. *CNS neuroscience & therapeutics, 24*(11), 1037-1052.

Geier, C. F., & Luna, B. (2009). The maturation of incentive processing and cognitive control. *Pharmacology Biochemistry and Behavior, 93*(3), 212-221.

Geier, C. F., Terwilliger, R., Teslovich, T., Velanova, K., & Luna, B. (2009). Immaturities in reward processing and its influence on inhibitory control in adolescence. *Cerebral cortex, 20*(7), 1613-1629.

Geissberger, N., Tik, M., Sladky, R., Woletz, M., Schuler, A.-L., Willinger, D., & Windischberger, C. (2020). Reproducibility of amygdala activation in facial emotion processing at 7T. *Neuroimage, 211*, 116585.

Gelman, A., & Rubin, D. B. (1992). Inference from iterative simulation using multiple sequences. *Statistical science, 7*(4), 457-472.

Gesiarz, F., Cahill, D., & Sharot, T. (2019). Evidence accumulation is biased by motivation: A computational account. *PLoS computational biology, 15*(6), e1007089.

Giedd, J. N. (2004). Structural magnetic resonance imaging of the adolescent brain. *Ann N Y Acad Sci, 1021*, 77-85.

Giedd, J. N., Blumenthal, J., Jeffries, N. O., Castellanos, F. X., Liu, H., Zijdenbos, A., Paus, T., Evans, A. C., & Rapoport, J. L. (1999). Brain development during childhood and adolescence: a longitudinal MRI study. *Nat Neurosci, 2*(10), 861-863.

Gläscher, J. P., & O'Doherty, J. P. (2010). Model-based approaches to neuroimaging: combining reinforcement learning theory with fMRI data. *Wiley Interdisciplinary Reviews: Cognitive Science, 1*(4), 501-510.

Gläscher, J. P., Tüscher, O., Weiller, C., & Büchel, C. (2004). Elevated responses to constant facial emotions in different faces in the human amygdala: an fMRI study of facial identity and expression. *BMC neuroscience, 5*(1), 45.

Gogolla, N. (2017). The insular cortex. *Current Biology, 27*(12), R580-R586.

Gogtay, N., Giedd, J. N., Lusk, L., Hayashi, K. M., Greenstein, D., Vaituzis, A. C., Nugent, T. F., Herman, D. H., Clasen, L. S., & Toga, A. W. (2004). Dynamic mapping of human cortical development during childhood through early adulthood. *Proceedings of the National Academy of Sciences, 101*(21), 8174-8179.

Gollan, J. K., Hoxha, D., Hunnicutt-Ferguson, K., Norris, C. J., Rosebrock, L., Sankin, L., & Cacioppo, J. (2016). The negativity bias predicts response rate to Behavioral Activation for depression. *Journal of behavior therapy and experimental psychiatry, 52*, 171-178.

Gore, F. M., Bloem, P. J., Patton, G. C., Ferguson, J., Joseph, V., Coffey, C., Sawyer, S. M., & Mathers, C. D. (2011). Global burden of disease in young people aged 10-24 years: a systematic analysis. *Lancet, 377*(9783), 2093-2102. doi:10.1016/S0140-6736(11)60512-6

Gould, M. S., King, R., Greenwald, S., Fisher, P., Schwab-Stone, M., Kramer, R., Flisher, A. J., Goodman, S., Canino, G., & Shaffer, D. (1998). Psychopathology associated with suicidal ideation and attempts among children and adolescents. *J Am Acad Child Adolesc Psychiatry, 37*(9), 915-923. doi:10.1097/00004583-199809000-00011

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

Greicius, M. D., Flores, B. H., Menon, V., Glover, G. H., Solvason, H. B., Kenna, H., Reiss, A. L., & Schatzberg, A. F. (2007). Resting-state functional connectivity in major depression: abnormally increased contributions from subgenual cingulate cortex and thalamus. *Biological psychiatry, 62*(5), 429-437.

Groenewold, N. A., Opmeer, E. M., de Jonge, P., Aleman, A., & Costafreda, S. G. (2013). Emotional valence modulates brain functional abnormalities in depression: evidence from a meta-analysis of fMRI studies. *Neuroscience & Biobehavioral Reviews, 37*(2), 152-163.

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

Häberling, I., Baumgartner, N., Emery, S., Keller, P., Strumberger, M., Nalani, K., Schmeck, K., Erb, S., Bachmann, S., & Wöckel, L. (2019). Anxious depression as a clinically relevant subtype of pediatric major depressive disorder. *Journal of Neural Transmission, 126*(9), 1217-1230.

Hall, L. M., Klimes-Dougan, B., Hunt, R. H., Thomas, K. M., Houri, A., Noack, E., Mueller, B. A., Lim, K. O., & Cullen, K. R. (2014). An fMRI study of emotional face processing in adolescent major depression. *Journal of affective disorders, 168*, 44-50.

Hämmerer, D., Li, S.-C., Müller, V., & Lindenberger, U. (2011). Life span differences in electrophysiological correlates of monitoring gains and losses during probabilistic reinforcement learning. *Journal of Cognitive Neuroscience, 23*(3), 579-592.

Hariri, A. R., Mattay, V. S., Tessitore, A., Fera, F., & Weinberger, D. R. (2003). Neocortical modulation of the amygdala response to fearful stimuli. *Biological psychiatry, 53*(6), 494-501.

Hariri, A. R., Tessitore, A., Mattay, V. S., Fera, F., & Weinberger, D. R. (2002). The amygdala response to emotional stimuli: a comparison of faces and scenes. *Neuroimage, 17*(1), 317-323.

Haufe, S., Meinecke, F., Görgen, K., Dähne, S., Haynes, J.-D., Blankertz, B., & Bießmann, F. (2014). On the interpretation of weight vectors of linear models in multivariate neuroimaging. *Neuroimage, 87*, 96-110.

Hauser, T. U., Iannaccone, R., Stämpfli, P., Drechsler, R., Brandeis, D., Walitza, S., & Brem, S. (2014). The feedback-related negativity (FRN) revisited: new insights into the localization, meaning and network organization. *Neuroimage, 84*, 159-168.

Hauser, T. U., Iannaccone, R., Walitza, S., Brandeis, D., & 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.

Hauser, T. U., Will, G. J., Dubois, M., & Dolan, R. J. (2019). Annual research review: developmental computational psychiatry. *Journal of Child psychology and Psychiatry, 60*(4), 412-426.

Heathcote, A., Lin, Y.-S., Reynolds, A., Strickland, L., Gretton, M., & Matzke, D. (2019). Dynamic models of choice. *Behavior research methods, 51*(2), 961-985.

Heinzle, J., & Stephan, K. E. (2018). Dynamic causal modeling and its application to psychiatric disorders *Computational Psychiatry* (pp. 117-144): Elsevier.

Herzallah, M. M., Moustafa, A. A., Natsheh, J. Y., Abdellatif, S. M., Taha, M. B., Tayem, Y. I., Sehwail, M. A., Amleh, I., Petrides, G., & Myers, C. E. (2013). Learning from

negative feedback in patients with major depressive disorder is attenuated by SSRI antidepressants. *Frontiers in integrative neuroscience, 7*, 67.

Ho, T. C., & Auerbach, R. P. (2017). Toward an Improved Understanding of Corticobasal Ganglia Reward Circuitry in Adolescent Depression. *Biological Psychiatry: Cognitive Neuroscience and Neuroimaging, 2*(7), 554-555.

Ho, T. C., Yang, G., Wu, J., Cassey, P., Brown, S. D., Hoang, N., Chan, M., Connolly, C. G., Henje-Blom, E., & Duncan, L. G. (2014). Functional connectivity of negative emotional processing in adolescent depression. *Journal of affective disorders, 155*, 65-74.

Ho, T. C., Zhang, S., Sacchet, M. D., Weng, H., Connolly, C. G., Henje Blom, E., Han, L. K., Mobayed, N. O., & Yang, T. T. (2016). Fusiform gyrus dysfunction is associated with perceptual processing efficiency to emotional faces in adolescent depression: a model-based approach. *Frontiers in psychology, 7*, 40.

Huang, H., Fan, X., Williamson, D. E., & Rao, U. (2011). White matter changes in healthy adolescents at familial risk for unipolar depression: a diffusion tensor imaging study. *Neuropsychopharmacology, 36*(3), 684-691.

Huettel, S. A. (2012). Event-related fMRI in cognition. *Neuroimage, 62*(2), 1152-1156.

Huys, Q. J., Daw, N. D., & Dayan, P. (2015). Depression: a decision-theoretic analysis. *Annual review of neuroscience, 38*, 1-23.

Huys, Q. J., Maia, T. V., & Frank, M. J. (2016). Computational psychiatry as a bridge from neuroscience to clinical applications. *Nature neuroscience, 19*(3), 404.

Ingram, R. E., Miranda, J., & Segal, Z. V. (1998). *Cognitive vulnerability to depression*: Guilford Press.

Insel, C., Kastman, E. K., Glenn, C. R., & Somerville, L. H. (2017). Development of corticostriatal connectivity constrains goal-directed behavior during adolescence. *Nature communications, 8*(1), 1605.

Insel, T. R. (2014). The NIMH research domain criteria (RDoC) project: precision medicine for psychiatry. *American Journal of Psychiatry, 171*(4), 395-397.

Jacobs, R. H., Reinecke, M. A., Gollan, J. K., & Kane, P. (2008). Empirical evidence of cognitive vulnerability for depression among children and adolescents: A cognitive science and developmental perspective. *Clinical psychology review, 28*(5), 759-782.

Jalbrzikowski, M., Larsen, B., Hallquist, M. N., Foran, W., Calabro, F., & Luna, B. (2017). Development of white matter microstructure and intrinsic functional connectivity between the amygdala and ventromedial prefrontal cortex: associations with anxiety and depression. *Biological Psychiatry, 82*(7), 511-521.

James, S. L., Abate, D., Abate, K. H., Abay, S. M., Abbafati, C., Abbasi, N., Abbastabar, H., Abd-Allah, F., Abdela, J., & Abdelalim, A. (2018). Global, regional, and national incidence, prevalence, and years lived with disability for 354 diseases and injuries for 195 countries and territories, 1990–2017: a systematic analysis for the Global Burden of Disease Study 2017. *The Lancet, 392*(10159), 1789-1858.

Jeffreys, H. (1998). *The theory of probability*: OUP Oxford.

Jenkins, L., Kendall, A., Kassel, M., Patrón, V., Gowins, J., Dion, C., Shankman, S. A., Weisenbach, S., Maki, P., & Langenecker, S. (2018). Considering sex differences clarifies the effects of depression on facial emotion processing during fMRI. *Journal of affective disorders, 225*, 129-136.

Jensen, J., McIntosh, A. R., Crawley, A. P., Mikulis, D. J., Remington, G., & Kapur, S. (2003). Direct activation of the ventral striatum in anticipation of aversive stimuli. *Neuron, 40*(6), 1251-1257.

Jin, J., Narayanan, A., Perlman, G., Luking, K., DeLorenzo, C., Hajcak, G., Klein, D. N., Kotov, R., & Mohanty, A. (2017). Orbitofrontal cortex activity and connectivity predict future depression symptoms in adolescence. *Biological Psychiatry: Cognitive Neuroscience and Neuroimaging, 2*(7), 610-618.

Jin, J., Zelano, C., Gottfried, J. A., & Mohanty, A. (2015). Human amygdala represents the complete spectrum of subjective valence. *Journal of Neuroscience, 35*(45), 15145-15156.

Johnston, V., Liberato, S., & Thomas, D. (2012). Incentives for preventing smoking in children and adolescents. *Cochrane Database of Systematic Reviews*(10).

Johnstone, T., van Reekum, C. M., Urry, H. L., Kalin, N. H., & Davidson, R. J. (2007). Failure to regulate: counterproductive recruitment of top-down prefrontal-subcortical circuitry in major depression. *Journal of Neuroscience, 27*(33), 8877-8884.

Joseph, J. E., Zhu, X., Lynam, D., & Kelly, T. H. (2016). Modulation of meso-limbic reward processing by motivational tendencies in young adolescents and adults. *NeuroImage, 129*, 40-54.

Kahnt, T., & Tobler, P. N. (2017). Dopamine modulates the functional organization of the orbitofrontal cortex. *Journal of Neuroscience, 37*(6), 1493-1504.

Kann, L., McManus, T., Harris, W. A., Shanklin, S. L., Flint, K. H., Queen, B., Lowry, R., Chyen, D., Whittle, L., & Thornton, J. (2018). Youth risk behavior surveillance-United States, 2017. *MMWR Surveillance Summaries, 67*(8), 1.

Kanwisher, N., McDermott, J., & Chun, M. M. (1997). The fusiform face area: a module in human extrastriate cortex specialized for face perception. *Journal of neuroscience, 17*(11), 4302-4311.

Kaufman, J., Birmaher, B., Brent, D., Rao, U., Flynn, C., Moreci, P., Williamson, D., & Ryan, N. (1997). Schedule for affective disorders and schizophrenia for school-age children-present and lifetime version (K-SADS-PL): initial reliability and validity data. *Journal of the American Academy of Child & Adolescent Psychiatry, 36*(7), 980-988.

Kawai, T., Yamada, H., Sato, N., Takada, M., & Matsumoto, M. (2018). Preferential representation of past outcome information and future choice behavior by putative inhibitory interneurons rather than putative pyramidal neurons in the primate dorsal anterior cingulate cortex. *Cerebral Cortex, 29*(6), 2339-2352.

Kawasaki, H., Adolphs, R., Kaufman, O., Damasio, H., Damasio, A. R., Granner, M., Bakken, H., Hori, T., & Howard III, M. A. (2001). Single-neuron responses to emotional visual stimuli recorded in human ventral prefrontal cortex. *Nature neuroscience, 4*(1), 15.

Keedwell, P. A., Drapier, D., Surguladze, S., Giampietro, V., Brammer, M., & Phillips, M. (2010). Subgenual cingulate and visual cortex responses to sad faces predict clinical outcome during antidepressant treatment for depression. *Journal of affective disorders, 120*(1-3), 120-125.

Keenan-Miller, D., Hammen, C. L., & Brennan, P. A. (2007). Health outcomes related to early adolescent depression. *Journal of Adolescent Health, 41*(3), 256-262.

Kerestes, R., Davey, C. G., Stephanou, K., Whittle, S., & Harrison, B. J. (2014). Functional brain imaging studies of youth depression: a systematic review. *NeuroImage: Clinical, 4*, 209-231.

Keshavan, M. S., Giedd, J., Lau, J. Y., Lewis, D. A., & Paus, T. (2014). Changes in the adolescent brain and the pathophysiology of psychotic disorders. *The Lancet Psychiatry, 1*(7), 549-558.

Kessler, H., Doyen-Waldecker, C., Hofer, C., Hoffmann, H., Traue, H. C., & Abler, B. (2011). Neural correlates of the perception of dynamic versus static facial expressions of emotion. *GMS Psycho-Social-Medicine, 8*.

Kessler, R. C. (2012). The costs of depression. *Psychiatric Clinics, 35*(1), 1-14.

Kessler, R. C., Berglund, P., Demler, O., Jin, R., Merikangas, K., & Walters, E. (2005). Lifetime prevalence and age-of-onset distributions of DSM-IV disorders in the National Comorbidity Survey Replication. *Arch Gen Psychiatry, 62*(6), 593-602. doi:10.1001/archpsyc.62.6.593

Kieling, C., Adewuya, A., Fisher, H. L., Karmacharya, R., Kohrt, B. A., Swartz, J. R., & Mondelli, V. (2019). Identifying depression early in adolescence. *The Lancet Child & Adolescent Health, 3*(4), 211-213.

Kim, H., Somerville, L. H., Johnstone, T., Polis, S., Alexander, A. L., Shin, L. M., & Whalen, P. J. (2004). Contextual modulation of amygdala responsivity to surprised faces. *Journal of cognitive neuroscience, 16*(10), 1730-1745.

Kim, M. J., Loucks, R. A., Palmer, A. L., Brown, A. C., Solomon, K. M., Marchante, A. N., & Whalen, P. J. (2011). The structural and functional connectivity of the amygdala: from normal emotion to pathological anxiety. *Behavioural brain research, 223*(2), 403-410.

Kinnison, J., Padmala, S., Choi, J.-M., & Pessoa, L. (2012). Network analysis reveals increased integration during emotional and motivational processing. *Journal of Neuroscience, 32*(24), 8361-8372.

Knutson, B., Fong, G. W., Bennett, S. M., Adams, C. M., & Hommer, D. (2003). A region of mesial prefrontal cortex tracks monetarily rewarding outcomes: characterization with rapid event-related fMRI. *Neuroimage, 18*(2), 263-272.

Knutson, B., Katovich, K., & Suri, G. (2014). Inferring affect from fMRI data. *Trends in cognitive sciences, 18*(8), 422-428.

Knutson, B., Westdorp, A., Kaiser, E., & Hommer, D. (2000). FMRI visualization of brain activity during a monetary incentive delay task. *Neuroimage, 12*(1), 20-27.

Kool, W., Gershman, S. J., & Cushman, F. A. (2017). Cost-benefit arbitration between multiple reinforcement-learning systems. *Psychological science, 28*(9), 1321-1333.

Kouneiher, F., Charron, S., & Koechlin, E. (2009). Motivation and cognitive control in the human prefrontal cortex. *Nature neuroscience, 12*(7), 939.

Kriegeskorte, N., & Douglas, P. K. (2018). Cognitive computational neuroscience. *Nature neuroscience, 21*(9), 1148-1160.

Kube, T., Schwarting, R., Rozenkrantz, L., Glombiewski, J. A., & Rief, W. (2020). Distorted cognitive processes in major depression: A predictive processing perspective. *Biological psychiatry, 87*(5), 388-398.

Kühn, S., Mascharek, A., Banaschewski, T., Bodke, A., Bromberg, U., Büchel, C., Quinlan, E. B., Desrivieres, S., Flor, H., & Grigis, A. (2019). Predicting development of adolescent drinking behaviour from whole brain structure at 14 years of age. *Elife, 8*, e44056.

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

Kumar, P., Waiter, G., Ahearn, T., Milders, M., Reid, I., & Steele, J. (2008). Abnormal temporal difference reward-learning signals in major depression. *Brain, 131*(8), 2084-2093.

Ladouceur, C. D., Kerestes, R., Schlund, M. W., Shirtcliff, E. A., Lee, Y., & Dahl, R. E. (2019). Neural systems underlying reward cue processing in early adolescence: The role of puberty and pubertal hormones. *Psychoneuroendocrinology, 102*, 281-291.

Lamm, C., Benson, B., Guyer, A., Perez-Edgar, K., Fox, N., Pine, D., & Ernst, M. (2014). Longitudinal study of striatal activation to reward and loss anticipation from mid-adolescence into late adolescence/early adulthood. *Brain and cognition, 89*, 51-60.

Langner, O., Dotsch, R., Bijlstra, G., Wigboldus, D. H., Hawk, S. T., & Van Knippenberg, A. (2010). Presentation and validation of the Radboud Faces Database. *Cognition and emotion, 24*(8), 1377-1388.

Lawson, R. P., Nord, C. L., Seymour, B., Thomas, D., Dayan, P., Pilling, S., & Roiser, J. P. (2017). Disrupted habenula function in major depression. *Molecular psychiatry, 22*(2), 202-208.

Lawson, R. P., Seymour, B., Loh, E., Lutti, A., Dolan, R. J., Dayan, P., Weiskopf, N., & Roiser, J. P. (2014). The habenula encodes negative motivational value associated with primary punishment in humans. *Proceedings of the National Academy of Sciences, 111*(32), 11858-11863.

Lebreton, M., Abitbol, R., Daunizeau, J., & Pessiglione, M. (2015). Automatic integration of confidence in the brain valuation signal. *Nature neuroscience, 18*(8), 1159.

Lehmbeck, J. T., Brassen, S., Braus, D. F., & Weber-Fahr, W. (2008). Subgenual anterior cingulate cortex alterations in late-onset depression are related to "pessimistic thoughts". *The American Journal of Geriatric Psychiatry, 16*(3), 248-249.

Leung, B. K., & Balleine, B. W. (2015). Ventral pallidal projections to mediodorsal thalamus and ventral tegmental area play distinct roles in outcome-specific Pavlovian-instrumental transfer. *Journal of Neuroscience, 35*(12), 4953-4964.

Levitt, S. D., List, J. A., Neckermann, S., & Sadoff, S. (2016). The behavioralist goes to school: Leveraging behavioral economics to improve educational performance. *American Economic Journal: Economic Policy, 8*(4), 183-219.

Li, Z., Yan, C., Xie, W.-Z., Li, K., Zeng, Y.-W., Jin, Z., Cheung, E. F., & Chan, R. C. (2015). Anticipatory pleasure predicts effective connectivity in the mesolimbic system. *Frontiers in behavioral neuroscience, 9*, 217.

Liao, W., Qiu, C., Gentili, C., Walter, M., Pan, Z., Ding, J., Zhang, W., Gong, Q., & Chen, H. (2010). Altered effective connectivity network of the amygdala in social anxiety disorder: a resting-state FMRI study. *PloS one, 5*(12), e15238.

Lindquist, K. A., Satpute, A. B., Wager, T. D., Weber, J., & Barrett, L. F. (2015). The brain basis of positive and negative affect: evidence from a meta-analysis of the human neuroimaging literature. *Cerebral Cortex, 26*(5), 1910-1922.

Locke, H. S., & Braver, T. S. (2008). Motivational influences on cognitive control: behavior, brain activation, and individual differences. *Cognitive, Affective, & Behavioral Neuroscience, 8*(1), 99-112.

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

Lopez, A. D., Mathers, C. D., Ezzati, M., Jamison, D. T., & Murray, C. J. (2006). *Global burden of disease and risk factors*: The World Bank.

Luking, K. R., Pagliaccio, D., Luby, J. L., & Barch, D. M. (2016a). Depression risk predicts blunted neural responses to gains and enhanced responses to losses in healthy children. *Journal of the American Academy of Child & Adolescent Psychiatry, 55*(4), 328-337.

Luking, K. R., Pagliaccio, D., Luby, J. L., & Barch, D. M. (2016b). Reward processing and risk for depression across development. *Trends in cognitive sciences, 20*(6), 456-468.

Luking, K. R., Repovs, G., Belden, A. C., Gaffrey, M. S., Botteron, K. N., Luby, J. L., & Barch, D. M. (2011). Functional connectivity of the amygdala in early-childhood-onset depression. *Journal of the American Academy of Child & Adolescent Psychiatry, 50*(10), 1027-1041. e1023.

Manohar, S. G., Chong, T. T.-J., Apps, M. A., Batla, A., Stamelou, M., Jarman, P. R., Bhatia, K. P., & Husain, M. (2015). Reward pays the cost of noise reduction in motor and cognitive control. *Current Biology, 25*(13), 1707-1716.

Markham, J. A., Morris, J. R., & Juraska, J. M. (2007). Neuron number decreases in the rat ventral, but not dorsal, medial prefrontal cortex between adolescence and adulthood. *Neuroscience, 144*(3), 961-968. doi:10.1016/j.neuroscience.2006.10.015

Marr, D. (1982). 2010. Vision. A Computational Investigation into the Human Representation and Processing of Visual Information: Cambridge, MA: MIT Press.

Marsman, A., Mandl, R. C., van den Heuvel, M. P., Boer, V. O., Wijnen, J. P., Klomp, D. W., & Luijten, P. R. (2013). Glutamate changes in healthy young adulthood. *European Neuropsychopharmacology, 23*(11), 1484-1490.

Mayberg, H. S. (1997). Limbic-cortical dysregulation: a proposed model of depression. *The Journal of neuropsychiatry and clinical neurosciences*.

Mayberg, H. S., Lozano, A. M., Voon, V., McNeely, H. E., Seminowicz, D., Hamani, C., Schwalb, J. M., & Kennedy, S. H. (2005). Deep brain stimulation for treatment-resistant depression. *Neuron, 45*(5), 651-660.

McCabe, C., Mishor, Z., Cowen, P. J., & Harmer, C. J. (2010). Diminished neural processing of aversive and rewarding stimuli during selective serotonin reuptake inhibitor treatment. *Biological psychiatry, 67*(5), 439-445.

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

McRae, K., Misra, S., Prasad, A. K., Pereira, S. C., & Gross, J. J. (2011). Bottom-up and top-down emotion generation: implications for emotion regulation. *Social cognitive and affective neuroscience, 7*(3), 253-262.

McShane, B. B., Tackett, J. L., Böckenholt, U., & Gelman, A. (2019). Large-scale replication projects in contemporary psychological research. *The American Statistician, 73*(sup1), 99-105.

Miller, C. H., Hamilton, J. P., Sacchet, M. D., & Gotlib, I. H. (2015). Meta-analysis of functional neuroimaging of major depressive disorder in youth. *JAMA psychiatry, 72*(10), 1045-1053.

Mingtian, Z., Shuqiao, Y., Xiongzhao, Z., Jinyao, Y., Xueling, Z., Xiang, W., Yingzi, L., Jian, L., & Wei, W. (2012). Elevated amygdala activity to negative faces in young adults with early onset major depressive disorder. *Psychiatry Research: Neuroimaging, 201*(2), 107-112.

Minkova, L., Sladky, R., Kranz, G. S., Woletz, M., Geissberger, N., Kraus, C., Lanzenberger, R., & Windischberger, C. (2017). Task-dependent modulation of amygdala connectivity in social anxiety disorder. *Psychiatry Res Neuroimaging, 262*, 39-46. doi:10.1016/j.pscychresns.2016.12.016

Mir, P., Trender-Gerhard, I., Edwards, M. J., Schneider, S. A., Bhatia, K. P., & Jahanshahi, M. (2011). Motivation and movement: the effect of monetary incentive on performance speed. *Experimental brain research, 209*(4), 551-559.

Monk, C. S., Klein, R. G., Telzer, E. H., Schroth, E. A., Mannuzza, S., Moulton III, P. D., John L, Guardino, M., Masten, C. L., McClure-Tone, E. B., & Fromm, S. (2008). Amygdala and nucleus accumbens activation to emotional facial expressions in children and adolescents at risk for major depression. *american Journal of Psychiatry, 165*(1), 90-98.

Monk, C. S., Telzer, E. H., Mogg, K., Bradley, B. P., Mai, X., Louro, H. M., Chen, G., McClure-Tone, E. B., Ernst, M., & Pine, D. S. (2008). Amygdala and ventrolateral prefrontal cortex activation to masked angry faces in children and adolescents with generalized anxiety disorder. *Archives of general psychiatry, 65*(5), 568-576.

Monosov, I. E. (2017). Anterior cingulate is a source of valence-specific information about value and uncertainty. *Nature communications, 8*(1), 134.

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

Morris, R., Sparks, A., Mitchell, P., Weickert, C., & Green, M. (2012). Lack of cortico-limbic coupling in bipolar disorder and schizophrenia during emotion regulation. *Translational psychiatry, 2*(3), e90.

Musgrove, D. R., Eberly, L. E., Klimes-Dougan, B., Basgoze, Z., Thomas, K. M., Mueller, B. A., Houri, A., Lim, K. O., & Cullen, K. R. (2015). Impaired bottom-up effective connectivity between amygdala and subgenual anterior cingulate cortex in unmedicated adolescents with major depression: results from a dynamic causal modeling analysis. *Brain connectivity, 5*(10), 608-619.

Nakajima, M., Schmitt, L. I., & Halassa, M. M. (2019). Prefrontal Cortex Regulates Sensory Filtering through a Basal Ganglia-to-Thalamus Pathway. *Neuron*.

Nelson, B. D., Perlman, G., Klein, D. N., Kotov, R., & Hajcak, G. (2016). Blunted neural response to rewards as a prospective predictor of the development of depression in adolescent girls. *American Journal of Psychiatry, 173*(12), 1223-1230.

Nelson, E. E., Leibenluft, E., McClure, E. B., & Pine, D. S. (2005). The social re-orientation of adolescence: a neuroscience perspective on the process and its relation to psychopathology. *Psychological medicine, 35*(2), 163.

Neta, M., Davis, F. C., & Whalen, P. J. (2011). Valence resolution of ambiguous facial expressions using an emotional oddball task. *Emotion, 11*(6), 1425.

Niv, Y., Daw, N. D., Joel, D., & Dayan, P. (2007). Tonic dopamine: opportunity costs and the control of response vigor. *Psychopharmacology, 191*(3), 507-520.

Nord, C. L., Gray, A., Charpentier, C. J., Robinson, O. J., & Roiser, J. P. (2017). Unreliability of putative fMRI biomarkers during emotional face processing. *NeuroImage, 156*, 119-127.

Nord, C. L., Lawson, R. P., Huys, Q. J., Pilling, S., & Roiser, J. P. (2018). Depression is associated with enhanced aversive Pavlovian control over instrumental behaviour. *Scientific reports, 8*(1), 1-10.

Nuijten, M. B. (2019). Practical tools and strategies for researchers to increase replicability. *Developmental Medicine & Child Neurology, 61*(5), 535-539.

Nussenbaum, K., & Hartley, C. A. (2019). Reinforcement learning across development: What insights can we draw from a decade of research? *Developmental cognitive neuroscience*, 100733.

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

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

O'Doherty, J. P., Hampton, A., & 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.

Ochsner, K. N., & Gross, J. J. (2005). The cognitive control of emotion. *Trends Cogn Sci, 9*(5), 242-249. doi:10.1016/j.tics.2005.03.010

Ochsner, K. N., Ray, R. D., Cooper, J. C., Robertson, E. R., Chopra, S., Gabrieli, J. D., & Gross, J. J. (2004). For better or for worse: neural systems supporting the cognitive down-and up-regulation of negative emotion. *Neuroimage, 23*(2), 483-499.

Ochsner, K. N., Ray, R. R., Hughes, B., McRae, K., Cooper, J. C., Weber, J., Gabrieli, J. D., & Gross, J. J. (2009). Bottom-up and top-down processes in emotion generation: common and distinct neural mechanisms. *Psychol Sci, 20*(11), 1322-1331. doi:10.1111/j.1467-9280.2009.02459.x

Ochsner, K. N., Silvers, J. A., & Buhle, J. T. (2012). Functional imaging studies of emotion regulation: a synthetic review and evolving model of the cognitive control of emotion. *Annals of the New York Academy of Sciences, 1251*(1), E1-E24.

Ogawa, S., Lee, T.-M., Kay, A. R., & 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.

Oldham, S., Murawski, C., Fornito, A., Youssef, G., Yücel, M., & Lorenzetti, V. (2018). The anticipation and outcome phases of reward and loss processing: A neuroimaging meta-analysis of the monetary incentive delay task. *Human brain mapping, 39*(8), 3398-3418.

Overbye, K., Walhovd, K. B., Paus, T., Fjell, A. M., Huster, R. J., & Tamnes, C. K. (2019). Error processing in the adolescent brain: Age-related differences in electrophysiology, behavioral adaptation, and brain morphology. *Developmental cognitive neuroscience, 38*, 100665.

Palminteri, S., Justo, D., Jauffret, C., Pavlicek, B., Dauta, A., Delmaire, C., Czernecki, V., Karachi, C., Capelle, L., & Durr, A. (2012). Critical roles for anterior insula and dorsal striatum in punishment-based avoidance learning. *Neuron, 76*(5), 998-1009.

Parnaudeau, S., O'Neill, P.-K., Bolkan, S. S., Ward, R. D., Abbas, A. I., Roth, B. L., Balsam, P. D., Gordon, J. A., & Kellendonk, C. (2013). Inhibition of mediodorsal thalamus disrupts thalamofrontal connectivity and cognition. *Neuron, 77*(6), 1151-1162.

Parro, C., Dixon, M. L., & Christoff, K. (2018). The neural basis of motivational influences on cognitive control. *Human brain mapping, 39*(12), 5097-5111.

Paulsen, D. J., Hallquist, M. N., Geier, C. F., & Luna, B. (2015). Effects of incentives, age, and behavior on brain activation during inhibitory control: a longitudinal fMRI study. *Developmental cognitive neuroscience, 11*, 105-115.

Pe, M. L., Vandekerckhove, J., & Kuppens, P. (2013). A diffusion model account of the relationship between the emotional flanker task and rumination and depression. *Emotion, 13*(4), 739.

Penny, W. D. (2012). Comparing dynamic causal models using AIC, BIC and free energy. *Neuroimage, 59*(1), 319-330.

Penny, W. D., Stephan, K. E., Daunizeau, J., Rosa, M. J., Friston, K. J., Schofield, T. M., & Leff, A. P. (2010). Comparing families of dynamic causal models. *PLoS Comput Biol, 6*(3), e1000709. doi:10.1371/journal.pcbi.1000709

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

Peper, J., Pol, H. H., Crone, E., & Van Honk, J. (2011). Sex steroids and brain structure in pubertal boys and girls: a mini-review of neuroimaging studies. *Neuroscience, 191*, 28-37.

Pergola, G., Danet, L., Pitel, A.-L., Carlesimo, G. A., Segobin, S., Pariente, J., Suchan, B., Mitchell, A. S., & Barbeau, E. J. (2018). The regulatory role of the human mediodorsal thalamus. *Trends in cognitive sciences, 22*(11), 1011-1025.

Perlman, G., Simmons, A. N., Wu, J., Hahn, K. S., Tapert, S. F., Max, J. E., Paulus, M. P., Brown, G. G., Frank, G. K., & Campbell-Sills, L. (2012). Amygdala response and functional connectivity during emotion regulation: a study of 14 depressed adolescents. *Journal of affective disorders, 139*(1), 75-84.

Pessiglione, M., Schmidt, L., Draganski, B., Kalisch, R., Lau, H., Dolan, R. J., & Frith, C. D. (2007). How the brain translates money into force: a neuroimaging study of subliminal motivation. *Science, 316*(5826), 904-906.

Pessoa, L. (2017). A Network Model of the Emotional Brain. *Trends Cogn Sci, 21*(5), 357-371. doi:10.1016/j.tics.2017.03.002

Pessoa, L., & Adolphs, R. (2010). Emotion processing and the amygdala: from a'low road'to'many roads' of evaluating biological significance. *Nature reviews neuroscience, 11*(11), 773.

Pfabigan, D. M., Seidel, E.-M., Sladky, R., Hahn, A., Paul, K., Grahl, A., Küblböck, M., Kraus, C., Hummer, A., & Kranz, G. S. (2014). P300 amplitude variation is related to ventral striatum BOLD response during gain and loss anticipation: an EEG and fMRI experiment. *NeuroImage, 96*, 12-21.

Phelps, E. A. (2006). Emotion and cognition: insights from studies of the human amygdala. *Annu Rev Psychol, 57*, 27-53. doi:10.1146/annurev.psych.56.091103.070234

Phillips, M. L., Ladouceur, C. D., & Drevets, W. C. (2008). A neural model of voluntary and automatic emotion regulation: implications for understanding the pathophysiology and neurodevelopment of bipolar disorder. *Mol Psychiatry, 13*(9), 829, 833-857. doi:10.1038/mp.2008.65

Piekarski, D. J., Johnson, C. M., Boivin, J. R., Thomas, A. W., Lin, W. C., Delevich, K., Galarce, E. M., & Wilbrecht, L. (2017). Does puberty mark a transition in sensitive periods for plasticity in the associative neocortex? *Brain research, 1654*, 123-144.

Pine, D. S., Cohen, P., Gurley, D., Brook, J., & Ma, Y. (1998). The risk for early-adulthood anxiety and depressive disorders in adolescents with anxiety and depressive disorders. *Archives of general psychiatry, 55*(1), 56-64.

Pizzagalli, D. A., Holmes, A. J., Dillon, D. G., Goetz, E. L., Birk, J. L., Bogdan, R., Dougherty, D. D., Iosifescu, D. V., Rauch, S. L., & Fava, M. (2009). Reduced caudate and nucleus accumbens response to rewards in unmedicated individuals with major depressive disorder. *American Journal of Psychiatry, 166*(6), 702-710.

Platt, B., Waters, A. M., Schulte-Koerne, G., Engelmann, L., & Salemink, E. (2017). A review of cognitive biases in youth depression: attention, interpretation and memory. *Cognition and Emotion, 31*(3), 462-483.

Pohlack, S. T., Nees, F., Ruttorf, M., Schad, L. R., & Flor, H. (2012). Activation of the ventral striatum during aversive contextual conditioning in humans. *Biological Psychology, 91*(1), 74-80.

Power, J. D., Barnes, K. A., Snyder, A. Z., Schlaggar, B. L., & Petersen, S. E. (2012). Spurious but systematic correlations in functional connectivity MRI networks arise from subject motion. *Neuroimage, 59*(3), 2142-2154.

Price, J. L. (2005). Free will versus survival: brain systems that underlie intrinsic constraints on behavior. *Journal of Comparative Neurology, 493*(1), 132-139.

Puce, A., Allison, T., Asgari, M., Gore, J. C., & McCarthy, G. (1996). Differential sensitivity of human visual cortex to faces, letterstrings, and textures: a functional magnetic resonance imaging study. *Journal of neuroscience, 16*(16), 5205-5215.

Queirazza, F., Fouragnan, E., Steele, J. D., Cavanagh, J., & Philiastides, M. G. (2019). Neural correlates of weighted reward prediction error during reinforcement learning classify response to cognitive behavioral therapy in depression. *Science advances, 5*(7), eaav4962.

Quirk, G. J., & Beer, J. S. (2006). Prefrontal involvement in the regulation of emotion: convergence of rat and human studies. *Current opinion in neurobiology, 16*(6), 723-727.

Ramnani, N., Elliott, R., Athwal, B., & Passingham, R. (2004). Prediction error for free monetary reward in the human prefrontal cortex. *Neuroimage, 23*(3), 777-786.

Ray, R. D., & Zald, D. H. (2012). Anatomical insights into the interaction of emotion and cognition in the prefrontal cortex. *Neuroscience and Biobehavioral Reviews, 36*(1), 479-501. doi:10.1016/j.neubiorev.2011.08.005

Redlich, R., Opel, N., Bürger, C., Dohm, K., Grotegerd, D., Förster, K., Zaremba, D., Meinert, S., Repple, J., & Enneking, V. (2018). The limbic system in youth depression: brain structural and functional alterations in adolescent in-patients with severe depression. *Neuropsychopharmacology, 43*(3), 546-554.

Remington, N. A., Fabrigar, L. R., & Visser, P. S. (2000). Reexamining the circumplex model of affect. *Journal of personality and social psychology, 79*(2), 286.

Rescorla, R. A., & Wagner, A. R. (1972). A theory of Pavlovian conditioning: Variations in the effectiveness of reinforcement and nonreinforcement. *Classical conditioning II: Current research and theory, 2*, 64-99.

Rigoli, F., Chew, B., Dayan, P., & Dolan, R. J. (2016). Multiple value signals in dopaminergic midbrain and their role in avoidance contexts. *Neuroimage, 135*, 197-203.

Rikhye, R. V., Gilra, A., & Halassa, M. M. (2018). Thalamic regulation of switching between cortical representations enables cognitive flexibility. *Nature neuroscience, 21*(12), 1753.

Robbins, T. W., & Arnsten, A. F. (2009). The neuropsychopharmacology of fronto-executive function: monoaminergic modulation. *Annual review of neuroscience, 32*, 267-287.

Roiser, J. P., Elliott, R., & Sahakian, B. J. (2012). Cognitive mechanisms of treatment in depression. *Neuropsychopharmacology, 37*(1), 117-136.

Rolls, E. T., Huang, C.-C., Lin, C.-P., Feng, J., & Joliot, M. (2020). Automated anatomical labelling atlas 3. *NeuroImage, 206*, 116189.

Rothkirch, M., Tonn, J., Köhler, S., & Sterzer, P. (2017). Neural mechanisms of reinforcement learning in unmedicated patients with major depressive disorder. *Brain, 140*(4), 1147-1157.

Roy, M., Shohamy, D., & Wager, T. D. (2012). Ventromedial prefrontal-subcortical systems and the generation of affective meaning. *Trends Cogn Sci, 16*(3), 147-156. doi:10.1016/j.tics.2012.01.005

Rutledge, R. B., Moutoussis, M., Smittenaar, P., Zeidman, P., Taylor, T., Hrynkiewicz, L., Lam, J., Skandali, N., Siegel, J. Z., & Ousdal, O. T. (2017). Association of neural and emotional impacts of reward prediction errors with major depression. *JAMA psychiatry, 74*(8), 790-797.

Sajad, A., Godlove, D. C., & Schall, J. D. (2019). Cortical microcircuitry of performance monitoring. *Nature neuroscience, 22*(2), 265.

Sander, D., Grafman, J., & Zalla, T. (2003). The human amygdala: an evolved system for relevance detection. *Reviews in the Neurosciences, 14*(4), 303-316.

Sander, D., & Scherer, K. (2014). Surprise. *Oxford companion to emotion and the affective sciences*: Oxford University Press.

Scholl, J., & Klein-Flügge, M. (2018). Understanding psychiatric disorder by capturing ecologically relevant features of learning and decision-making. *Behavioural brain research, 355*, 56-75.

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

Sexton, C. E., Mackay, C. E., & Ebmeier, K. P. (2009). A systematic review of diffusion tensor imaging studies in affective disorders. *Biological psychiatry, 66*(9), 814-823.

Sharp, C., Kim, S., Herman, L., Pane, H., Reuter, T., & Strathearn, L. (2014). Major depression in mothers predicts reduced ventral striatum activation in adolescent female offspring with and without depression. *Journal of abnormal psychology, 123*(2), 298.

Sheehan, D. V., Sheehan, K. H., Shytle, R. D., Janavs, J., Bannon, Y., Rogers, J. E., Milo, K. M., Stock, S. L., & Wilkinson, B. (2010). Reliability and validity of the mini international neuropsychiatric interview for children and adolescents (MINI-KID). *The Journal of clinical psychiatry*.

Siegle, G. J., Thompson, W. K., Collier, A., Berman, S. R., Feldmiller, J., Thase, M. E., & Friedman, E. S. (2012). Toward clinically useful neuroimaging in depression treatment: prognostic utility of subgenual cingulate activity for determining depression outcome in cognitive therapy across studies, scanners, and patient characteristics. *Archives of General Psychiatry, 69*(9), 913-924.

Silvers, J. A., Shu, J., Hubbard, A. D., Weber, J., & Ochsner, K. N. (2015). Concurrent and lasting effects of emotion regulation on amygdala response in adolescence and young adulthood. *Dev Sci, 18*(5), 771-784. doi:10.1111/desc.12260

Sisk, C. L., & Zehr, J. L. (2005). Pubertal hormones organize the adolescent brain and behavior. *Frontiers in neuroendocrinology, 26*(3-4), 163-174.

Skaf, C. R., Yamada, A., Garrido, G. E., Buchpiguel, C. A., Akamine, S., Castro, C. C., & Busatto, G. F. (2002). Psychotic symptoms in major depressive disorder are associated with reduced regional cerebral blood flow in the subgenual anterior cingulate cortex: a voxel-based single photon emission computed tomography (SPECT) study. *Journal of affective disorders, 68*(2-3), 295-305.

Sladky, R., Hoflich, A., Kublbock, M., Kraus, C., Baldinger, P., Moser, E., Lanzenberger, R., & Windischberger, C. (2015). Disrupted effective connectivity between the amygdala and orbitofrontal cortex in social anxiety disorder during emotion discrimination revealed by dynamic causal modeling for FMRI. *Cereb Cortex, 25*(4), 895-903. doi:10.1093/cercor/bht279

Sladky, R., Spies, M., Hoffmann, A., Kranz, G., Hummer, A., Gryglewski, G., Lanzenberger, R., Windischberger, C., & Kasper, S. (2015). (S)-citalopram influences amygdala modulation in healthy subjects: a randomized placebo-controlled double-blind fMRI study using dynamic causal modeling. *Neuroimage, 108*, 243-250.

Slotnick, S. D. (2017). Cluster success: fMRI inferences for spatial extent have acceptable false-positive rates. *Cognitive neuroscience, 8*(3), 150-155.

Slotnick, S. D., Moo, L. R., Segal, J. B., & Hart Jr, J. (2003). Distinct prefrontal cortex activity associated with item memory and source memory for visual shapes. *Cognitive Brain Research, 17*(1), 75-82.

Smith, D. V., Hayden, B. Y., Truong, T.-K., Song, A. W., Platt, M. L., & Huettel, S. A. (2010). Distinct value signals in anterior and posterior ventromedial prefrontal cortex. *Journal of Neuroscience, 30*(7), 2490-2495.

Snyder, H. R. (2013). Major depressive disorder is associated with broad impairments on neuropsychological measures of executive function: a meta-analysis and review. *Psychological bulletin, 139*(1), 81.

Sokolov, A. A., Zeidman, P., Erb, M., Ryvlin, P., Pavlova, M. A., & Friston, K. J. (2019). Linking structural and effective brain connectivity: structurally informed parametric empirical bayes (si-peb). *Brain Structure and Function, 224*(1), 205-217.

Somerville, L. H. (2016). Searching for signatures of brain maturity: what are we searching for? *Neuron, 92*(6), 1164-1167.

Somerville, L. H., Hare, T., & Casey, B. J. (2011). Frontostriatal maturation predicts cognitive control failure to appetitive cues in adolescents. *Journal of cognitive neuroscience, 23*(9), 2123-2134.

Soriano Salinas, C. M., Fontaine, J. R., & Scherer, K. R. (2015). Surprise in the GRID. *Review of Cognitive Linguistics, 13*(2), 436-460.

Steele, J., Kumar, P., & Ebmeier, K. P. (2007). Blunted response to feedback information in depressive illness. *Brain, 130*(9), 2367-2374.

Steinberg, L. (2010). A dual systems model of adolescent risk-taking. *Developmental Psychobiology: The Journal of the International Society for Developmental Psychobiology, 52*(3), 216-224.

Stephan, K. E., Iglesias, S., Heinzle, J., & Diaconescu, A. O. (2015). Translational perspectives for computational neuroimaging. *Neuron, 87*(4), 716-732.

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

Stephan, K. E., Penny, W. D., Daunizeau, J., Moran, R. J., & Friston, K. J. (2009). Bayesian model selection for group studies. *Neuroimage, 46*(4), 1004-1017.

Stephan, K. E., Penny, W. D., Moran, R. J., den Ouden, H. E., Daunizeau, J., & Friston, K. J. (2010). Ten simple rules for dynamic causal modeling. *Neuroimage, 49*(4), 3099-3109.

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

Stiensmeier-Pelster, J., Braune-Krickau, M., Schürmann, M., & Duda, K. (2014). DIKJ - Depressionsinventar für Kinder und Jugendliche.

Störmer, V., Eppinger, B., & Li, S.-C. (2014). Reward speeds up and increases consistency of visual selective attention: A lifespan comparison. *Cognitive, Affective, & Behavioral Neuroscience, 14*(2), 659-671.

Strang, N. M., & Pollak, S. D. (2014). Developmental continuity in reward-related enhancement of cognitive control. *Developmental cognitive neuroscience, 10*, 34-43.

Stringaris, A., Vidal-Ribas Belil, P., Artiges, E., Lemaitre, H., Gollier-Briant, F., Wolke, S., Vulser, H., Miranda, R., Penttilä, J., & Struve, M. (2015). The brain's response to reward anticipation and depression in adolescence: dimensionality, specificity, and longitudinal predictions in a community-based sample. *American Journal of Psychiatry, 172*(12), 1215-1223.

Sturman, D. A., & Moghaddam, B. (2011). The neurobiology of adolescence: changes in brain architecture, functional dynamics, and behavioral tendencies. *Neuroscience & Biobehavioral Reviews, 35*(8), 1704-1712.

Sul, J. H., Kim, H., Huh, N., Lee, D., & Jung, M. W. (2010). Distinct roles of rodent orbitofrontal and medial prefrontal cortex in decision making. *Neuron, 66*(3), 449-460.

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

Tao, R., Calley, C. S., Hart, J., Mayes, T. L., Nakonezny, P. A., Lu, H., Kennard, B. D., Tamminga, C. A., & Emslie, G. J. (2012). Brain activity in adolescent major depressive disorder before and after fluoxetine treatment. *American Journal of Psychiatry, 169*(4), 381-388.

Tavares, J. V. T., Clark, L., Furey, M. L., Williams, G. B., Sahakian, B. J., & Drevets, W. C. (2008). Neural basis of abnormal response to negative feedback in unmedicated mood disorders. *Neuroimage, 42*(3), 1118-1126.

Taylor, J. B., Visser, T. A., Fueggle, S. N., Bellgrove, M. A., & Fox, A. M. (2018). The error-related negativity (ERN) is an electrophysiological marker of motor impulsiveness on the Barratt Impulsiveness Scale (BIS-11) during adolescence. *Developmental cognitive neuroscience, 30*, 77-86.

Taylor, S. F., Martis, B., Fitzgerald, K. D., Welsh, R. C., Abelson, J. L., Liberzon, I., Himle, J. A., & Gehring, W. J. (2006). Medial frontal cortex activity and loss-related responses to errors. *Journal of Neuroscience, 26*(15), 4063-4070.

Telzer, E. H., McCormick, E. M., Peters, S., Cosme, D., Pfeifer, J. H., & van Duijvenvoorde, A. C. (2018). Methodological considerations for developmental longitudinal fMRI research. *Developmental cognitive neuroscience, 33*, 149-160.

Thapar, A., Collishaw, S., Pine, D. S., & Thapar, A. K. (2012). Depression in adolescence. *The Lancet, 379*(9820), 1056-1067.

Trautmann, S. A., Fehr, T., & Herrmann, M. (2009). Emotions in motion: dynamic compared to static facial expressions of disgust and happiness reveal more widespread emotion-specific activations. *Brain research, 1284*, 100-115.

Tsvetanov, K. A., Henson, R. N., Tyler, L. K., Razi, A., Geerligs, L., Ham, T. E., & Rowe, J. B. (2016). Extrinsic and intrinsic brain network connectivity maintains cognition across the lifespan despite accelerated decay of regional brain activation. *Journal of Neuroscience, 36*(11), 3115-3126.

Turner, B. M., Sederberg, P. B., Brown, S. D., & Steyvers, M. (2013). A method for efficiently sampling from distributions with correlated dimensions. *Psychological methods, 18*(3), 368.

Uddin, L. Q. (2015). Salience processing and insular cortical function and dysfunction. *Nature Reviews Neuroscience, 16*(1), 55-61.

Urry, H. L., Van Reekum, C. M., Johnstone, T., Kalin, N. H., Thurow, M. E., Schaefer, H. S., Jackson, C. A., Frye, C. J., Greischar, L. L., & Alexander, A. L. (2006). Amygdala and ventromedial prefrontal cortex are inversely coupled during regulation of negative affect and predict the diurnal pattern of cortisol secretion among older adults. *Journal of Neuroscience, 26*(16), 4415-4425.

Van Den Bos, W., Cohen, M. X., Kahnt, T., & Crone, E. A. (2012). Striatum–medial prefrontal cortex connectivity predicts developmental changes in reinforcement learning. *Cerebral cortex, 22*(6), 1247-1255.

Van Den Bos, W., Rodriguez, C. A., Schweitzer, J. B., & McClure, S. M. (2015). Adolescent impatience decreases with increased frontostriatal connectivity. *Proceedings of the National Academy of Sciences, 112*(29), E3765-E3774.

Van Duijvenvoorde, A. C., Zanolie, K., Rombouts, S. A., Raijmakers, M. E., & Crone, E. A. (2008). Evaluating the negative or valuing the positive? Neural mechanisms supporting feedback-based learning across development. *Journal of Neuroscience, 28*(38), 9495-9503.

Varlinskaya, E. I., Vetter-O'Hagen, C. S., & Spear, L. P. (2013). Puberty and gonadal hormones: role in adolescent-typical behavioral alterations. *Hormones and behavior, 64*(2), 343-349.

Vrieze, E., Pizzagalli, D. A., Demyttenaere, K., Hompes, T., Sienaert, P., de Boer, P., Schmidt, M., & Claes, S. (2013). Reduced reward learning predicts outcome in major depressive disorder. *Biological psychiatry, 73*(7), 639-645.

Vuilleumier, P., Richardson, M. P., Armony, J. L., Driver, J., & Dolan, R. J. (2004). Distant influences of amygdala lesion on visual cortical activation during emotional face processing. *Nature neuroscience, 7*(11), 1271.

Wagner, D. D., & Heatherton, T. F. (2012). Self-regulatory depletion increases emotional reactivity in the amygdala. *Social cognitive and affective neuroscience, 8*(4), 410-417.

Walter, M., Alizadeh, S., Jamalabadi, H., Lueken, U., Dannlowski, U., Walter, H., Olbrich, S., Colic, L., Kambeitz, J., & Koutsouleris, N. (2019). Translational machine learning for psychiatric neuroimaging. *Progress in Neuro-Psychopharmacology and Biological Psychiatry, 91*, 113-121.

Wei, Z., Yang, N., Liu, Y., Yang, L., Wang, Y., Han, L., Zha, R., Huang, R., Zhang, P., & Zhou, Y. (2016). Resting-state functional connectivity between the dorsal anterior cingulate cortex and thalamus is associated with risky decision-making in nicotine addicts. *Scientific reports, 6*, 21778.

Wessel, J. R. (2018). An adaptive orienting theory of error processing. *Psychophysiology, 55*(3), e13041.

Whalen, P. J. (1998). Fear, vigilance, and ambiguity: Initial neuroimaging studies of the human amygdala. *Current directions in psychological science, 7*(6), 177-188.

Whalen, P. J., Raila, H., Bennett, R., Mattek, A., Brown, A., Taylor, J., van Tieghem, M., Tanner, A., Miner, M., & Palmer, A. (2013). Neuroscience and facial expressions of emotion: The role of amygdala–prefrontal interactions. *Emotion Review, 5*(1), 78-83.

Whiteford, H. A., Degenhardt, L., Rehm, J., Baxter, A. J., Ferrari, A. J., Erskine, H. E., Charlson, F. J., Norman, R. E., Flaxman, A. D., & Johns, N. (2013). Global burden of disease attributable to mental and substance use disorders: findings from the Global Burden of Disease Study 2010. *The lancet, 382*(9904), 1575-1586.

Whitten, L. A. (2013). Translational neuroscience and potential contributions of functional magnetic resonance imaging (fMRI) to the prevention of substance misuse and antisocial behavior. *Prevention Science, 14*(3), 238-246.

Wilkinson, P., Kelvin, R., Roberts, C., Dubicka, B., & Goodyer, I. (2011). Clinical and psychosocial predictors of suicide attempts and nonsuicidal self-injury in the Adolescent Depression Antidepressants and Psychotherapy Trial (ADAPT). *American journal of psychiatry, 168*(5), 495-501.

Willinger, D., Karipidis, I. I., Beltrani, S., Di Pietro, S. V., Sladky, R., Walitza, S., Stämpfli, P., & Brem, S. (2019). Valence-dependent coupling of prefrontal-amygdala effective connectivity during facial affect processing. *Eneuro, 6*(4).

Wilson, R. C., Takahashi, Y. K., Schoenbaum, G., & Niv, Y. (2014). Orbitofrontal cortex as a cognitive map of task space. *Neuron, 81*(2), 267-279.

Winecoff, A., Clithero, J. A., Carter, R. M., Bergman, S. R., Wang, L., & Huettel, S. A. (2013). Ventromedial prefrontal cortex encodes emotional value. *Journal of Neuroscience, 33*(27), 11032-11039.

World Health Organization. (1993). *The ICD-10 classification of mental and behavioural disorders: diagnostic criteria for research* (Vol. 2): World Health Organization.

Wrase, J., Kahnt, T., Schlagenhauf, F., Beck, A., Cohen, M. X., Knutson, B., & Heinz, A. (2007). Different neural systems adjust motor behavior in response to reward and punishment. *Neuroimage, 36*(4), 1253-1262.

Wu, C. C., Samanez-Larkin, G. R., Katovich, K., & Knutson, B. (2014). Affective traits link to reliable neural markers of incentive anticipation. *Neuroimage, 84*, 279-289.

Yang, T. T., Simmons, A. N., Matthews, S. C., Tapert, S. F., Frank, G. K., Max, J. E., Bischoff-Grethe, A., Lansing, A. E., Brown, G., & Strigo, I. A. (2010). Adolescents with major depression demonstrate increased amygdala activation. *Journal of the American Academy of Child & Adolescent Psychiatry, 49*(1), 42-51.

Zeidman, P., Jafarian, A., Corbin, N., Seghier, M. L., Razi, A., Price, C. J., & Friston, K. J. (2019). A guide to group effective connectivity analysis, part 1: First level analysis with DCM for fMRI. *NeuroImage*.

Zhdanov, A., Atluri, S., Wong, W., Vaghei, Y., Daskalakis, Z. J., Blumberger, D. M., Frey, B. N., Giacobbe, P., Lam, R. W., & Milev, R. (2020). Use of machine learning for predicting escitalopram treatment outcome from electroencephalography recordings in adult patients with depression. *JAMA network open, 3*(1), e1918377-e1918377.

Ziegler, G., Hauser, T. U., Moutoussis, M., Bullmore, E. T., Goodyer, I. M., Fonagy, P., Jones, P. B., Lindenberger, U., & Dolan, R. J. (2019). Compulsivity and impulsivity traits linked to attenuated developmental frontostriatal myelination trajectories. *Nature neuroscience, 22*(6), 992.

Zinchenko, O., Yaple, Z. A., & Arsalidou, M. (2018). Brain responses to dynamic facial expressions: a normative meta-analysis. *Frontiers in Human Neuroscience, 12*.

Acknowledgements

First and foremost, I want to thank Prof. Dr. Silvia Brem for giving me the opportunity and entrusting me with this PhD project. I always appreciated her excellent guidance and support in performing this research and beyond, which has been always much more than what is expected from a supervisor. I am grateful that she shared her exceptional expertise in developmental neuroimaging and encouraged my scientific endeavors. Her outstanding mentoring shaped my academic development over the past years and her devotion to this research projects was very inspiring.

I also want to express my gratitude to Prof. Dr. Nicole Wenderoth for being the head of the dissertation committee and for the interesting discussions and helpful advice to improve the work continuously.

My sincere gratitude also goes to Prof. Dr. med. Dipl.-Psych. Susanne Walitza who supported this project with her expertise and comments in clinical questions and her admirable commitment and enthusiasm for clinical research.

Moreover, I want to thank Dr. Iliana Karipidis and Dr. Georgette Pleisch who shared their experiences, guided me through the project in the beginning and taught me many technical aspects of conducting research. After all these years, I can only hope that some day I will reach their level of diligence and care as well as their incredible eye for detail.

This project would not have been possible without the dedicated support of the project members. Particularly, I want to thank Selina Beltrani, Plamina Dimanova, Selina Neuer, Alina Strehl, and Nada Frei. Their support with conducting the measurements and communicating our research was extremely valuable. Furthermore, I would like to thank Dr. med. Gregor Berger, Dr. Isabelle Häberling, Dr. Noemi Baumgartner, Sophie Emery, Mona Albermann, Kristin Nalani, and Carolina Rauch, who provided clinical support and assistance with great determination.

Moreover, I want to express my earnest gratitude for the many interesting discussions during our lab meetings, at lunch or the coffee break, and the fruitful collaborations to the Developmental Neuroimaging Group, Sarah Di Pietro, Dr. Gorka Fraga González, Patrick Haller, Seline Coraj, Christina Lutz, Dr. Alexander Roth, Dr. Julia Werhahn, Marco Hartmann, and Susanna Mohl. I also want to thank Dr. Philipp Stämpfli, who always strived to provide the best possible research environment at the MR-center.

My gratitude also goes to my family, who always supported me during this period. Special thanks go to my Katka for her loving support, her patience and the endless travels between Zurich, Vienna and Barcelona.

Finally, I want to thank all study participants and their families. It was a pleasure to see with how much motivation, interest and curiosity the adolescents took part in the experiments and assessments.

Abbreviations

AMY	Amygdala
BIC	Bayesian information criterion
BMA	Bayesian model averaging
BOLD	Blood-oxygenation-level dependent
BPIC	Bayesian predictive information criterion
CDI	Child Depression Inventory
CD-RISC	Connor-Davidson Resilience Scale
CDT	Cluster-defining threshold
dACC	Dorsal anterior cingulate cortex
DCM	Dynamic causal modeling
DE-MCMC	Differential evolution Markov Chain Monte Carlo
DMC	Dynamic models of choice
DSM	Diagnostic and Statistical Manual of Mental Disorders
EPI	Echo-planar imaging
FD	Framewise displacement
FDR	False-discovery rate
FFA	Fusiform face area
fMRI	Functional magnetic resonance imaging
FWE	Family-wise error
GLM	General linear model
HC	Healthy controls
ICD	International Statistical Classification of Diseases and Related Health Problems
INS	Insula
HRF	Hemodynamic response
K-SADS-PL	Kiddie Schedule for Affective Disorders and Schizophrenia for School-Age Children Present and Lifetime Version
LBA	Linear ballistic accumulator
logRT	Log-transformed response time
LOOCV	Leave-one-out cross-validation
LPFC	Lateral prefrontal cortex
MDD	Major depressive disorder
MID	Monetary Incentive Delay Task
MINI-KID	Mini-International Neuropsychiatric Interview for Children and Adolescents
MNI	Montreal Neurological Institute
MPFC	Medial prefrontal cortex
OFC	Orbitofrontal cortex
OR	Odds ratio
PE	Prediction error

PEB	Parametric Empirical Bayes
PP	Posterior probability
PSS	Perceive Stress Scale
RDoC	Research domain criteria
RIAS	Reynolds Intellectual Assessment Scales
SDQ-K	Strength and Difficulty Questionnaire for Children
sgACC	Subgenual anterior cingulate cortex
SSRI	Selective serotonin reuptake inhibitor
STS	Superior temporal sulcus
TAPAS	TNU Algorithms for Psychiatry-Advancing Science
TD	Temporal difference
THL	Thalamus
TPJ	Temporoparietal junction
vlPFC	Ventrolateral prefrontal cortex
vmPFC	Ventromedial prefrontal cortex
VOI	Volume of interest
VS	Ventral striatum
WISC	Wechsler Intelligence Scale for Children
XP	Exceedance probability

List of figures

List of tables