Uitnodiging

Voor het bijwonen van de openbare verdediging van het proefschrift
Disentangling the functional brain architecture in mental disorders

op woensdag 30 augustus 2017 om 12.30 uur precies
in de Aula van de Radboud Universiteit, Comeniuslaan 2, te Nijmegen

U bent van harte welkom bij deze plechtigheid en de aansluitende receptie ter plaatse

R.H.R. Pruim
r.h.r.pruim@gmail.com

Paranimfen
Patrick Sijbel
Reinhard Prophitius

Disentangling the functional brain architecture in mental disorders

R.H.R. Pruim
Acknowledgement:
The work described in this thesis has been supported by NWO Large Investment grant 1750102007010 (Dr. Buitelaar), ZonMW Addiction: Risk Behaviour and Dependency Grant 60-60600-97-193 (Dr. Buitelaar), NWO Brain & Cognition: an Integrative Approach grant 433-09-242 (Dr. Buitelaar), NWO National Initiative Brain & Cognition 056-13-015 (Dr. Buitelaar), the EU FP7 grants TACTICS (278948), IMAGEMEND (602450), MATRICS (603016) and AGGRESSOTYPE (602805), EU IMI grant EU-AIMS (115300), National Institutes of Health grant R01MH62873 (Dr. Faraone), and grants from Radboudumc, University Medical Center Groningen, Accare, and VU University Amsterdam. Dr. Mennes has received funding from the European Research Council under the European Union’s Seventh Framework Programme (FP7/2007-2013) grant agreement 327340 (BRAIN FINGERPRINT). Dr. Beckmann is supported by the Netherlands Organisation for Scientific Research (NWO-Vidi 864-12-003) and we gratefully acknowledge the funding from the Wellcome Trust UK Strategic Award [098369/Z/12/Z]. Dr. Faraone is supported by the K.G. Jebsen Centre for Research on Neuropsychiatric Disorders, University of Bergen, Bergen, Norway, the European Community’s Seventh Framework Programme (FP7/2007-2013) under grant agreement n°602805 and NIMH grant R01MH094469. Dr. Franke’s research is supported by grants from the Netherlands Organization for Scientific Research (NWO), i.e. the NWO Brain & Cognition Excellence Program (grant 433-09-229) and a Vici grant from NWO (grant 016-130-669). Her research also received funding from the European Community’s Seventh Framework Programme (FP7/2007-2013) under grant agreements n° 602805 (Aggressotype) and n° 602450 (IMAGEMEND), and from the European Community’s Horizon 2020 Programme (H2020/2014-2020) under grant agreement n° 643051 (MiND). In addition, her work is supported by a grant for the ENIGMA Consortium (grant number U54 EB020403) from the BD2K Initiative of a cross-NIH partnership. I acknowledge the Department of Pediatrics of the VU University Medical Center for the opportunity to use the mock scanner for preparation of our participants.
Disentangling
the functional brain architecture
in mental disorders

Proefschrift

ter verkrijging van de graad van doctor
aan de Radboud Universiteit Nijmegen
op gezag van de rector magnificus prof. dr. J.H.J.M. van Krieken,
volgens besluit van het college van decanen
in het openbaar te verdedigen op woensdag 30 augustus 2017
om 12.30 uur precies

door

Raimon Pruim
geboren op 25 september 1987
te Oldebroek
Promotoren:
Prof. dr. J.K. Buitelaar
Prof. dr. C.F. Beckmann

Copromotor:
Dr. M.J.J. Mennes

Manuscriptcommissie:
Prof. dr. R. Cools
Prof. dr. S.A.R.B. Rombouts (Universiteit Leiden)
Dr. J.P. Marques
## Contents

**Introduction**  
7

**Chapter 1**  
ICA-AROMA: A robust ICA-based strategy for removing motion artifacts from fMRI data  
31

**Chapter 2**  
Evaluation of ICA-AROMA and alternative strategies for motion artifact removal in resting state fMR  
73

**Chapter 3**  
Modelling categorical and dimensional mechanisms in mental disorders  
113

**Chapter 4**  
Bio-behavioural network correlates of categorical and dimensional attention-deficit/hyperactivity disorder  
137

**Summary and general discussion**  
177

**Appendix**  
201
Introduction
Introduction

‘Wij zijn ons brein’
‘We are our brain’
Dick Swaab (book title) (Swaab, 2010)

We leven in ‘borderline tijden’
We live in ‘borderline times’
Dirk de Wachter (book title) (de Wachter, 2012)

‘Etiketkinderen’
‘Labelchildren’
Zembla (documentary title) (VARA, 2013)

‘Terroristen zijn niet gek. Maar wat dan wel?’
‘Terrorists are not crazy. But then, what are they?’
NRC (title newspaper article) (Dirk Vlasblom, 2016)

Media as well as mental health professionals intensively debate whether certain behaviors, which are considered to be ‘abnormal’, are due to a mental disorder. This debate is complicated because various definitions and conceptualizations of mental disorders exist (Meynen and Ralston, 2011). On the one hand behavioral abnormalities are seen as a consequence of biological (dys)function, while on the other hand mental disorders are viewed as social constructs. In the media this raises debates on for example the so-called ‘epidemic and hype’ of mental disorders (Dehue, 2008; VARA, 2010; van Hintum, 2013), the social and moral acceptance of certain ‘abnormal’ behavior with presumed biological underpinnings (e.g. sexual orientation (FRA, 2016)), and the question too which extent criminals and terrorists can be considered mentally ill and legally accountable (Melle, 2013; Dirk Vlasblom, 2016). This thesis will not discuss the pros and cons of various definitions of mental disorders. Rather, it takes the perspective that underlying biological functioning is central and exploits this fundamental notion as a starting point to reconceptualize research towards disentangling biological dysfunction (i.e. pathophysiology) associated with mental disorders. In this chapter I will first introduce the main concepts and topics of the thesis, and subsequently present the aims and outline of the thesis.
Introduction

Overview

Pathophysiological underpinnings of mental disorders can be present at different levels, such as the genetic, cognitive and/or neural architectural domain. Importantly, it is believed that dissociable impairments across and within these domains might result in similar behavioral symptoms, i.e. the existence of biological subgroups of patients. Interestingly, the last decade a new neuroimaging protocol has emerged, called resting-state fMRI (rfMRI), which allows to simultaneously study the functional architecture of multiple neural networks (Biswal et al., 1995; Fox and Raichle, 2007; Biswal, 2012). This technique provides new opportunities to investigate dissociable biological underpinnings of mental disorders across brain networks. By requiring minimum participant compliance it moreover offers great potential for clinical application (Fox and Raichle, 2007; Castellanos et al., 2013). To date, research exploiting rfMRI data has indeed yielded promising insights into neurofunctional correlates associated with mental disorders such as major depressive disorder (Dutta et al., 2014; Iwabuchi et al., 2015; Mulders et al., 2015), bipolar disorder (Vargas et al., 2013; Piguet et al., 2015), schizophrenia, (Narr and Leaver, 2015; Sheffield and Barch, 2016), anxiety disorders (Peterson et al., 2014), psychosis (Satterthwaite and Baker, 2015), autism spectrum disorder (Ha et al., 2015) and attention-deficit/hyperactivity disorder (ADHD) (Castellanos et al., 2009). Given its promise and opportunity to simultaneously study functionality of multiple brain networks, the current thesis will specifically focus on this technique.

Despite these promising results, the research field is confronted with high heterogeneity between studies (Castellanos et al., 2009; Oldehinkel et al., 2013; Vargas et al., 2013; Mulders et al., 2015). These drawbacks are a consequence of accumulating effects associated with 1) categorical conceptualization of mental disorders, 2) clinical, etiological and pathophysiological heterogeneity, and 3) noise in the acquired rfMRI data.

1. Diagnostics in mental disorders are based on a categorical system implying discrete differences between diagnosed and healthy individuals (American Psychiatric Association, 2013a). Accordingly, researchers typically investigate mental disorders using case-control designs. However, evidence is increasing that most mental disorders in fact represent the impairing tail on a continuum of normal behavior (Coghill and Sonuga-Barke, 2012; Haslam et al., 2012). By only implying discrete differences, case-control designs are therefore suboptimal models for investigating the pathophysiological mechanisms associated with mental disorders.
2. The research field is moving towards neural systems based conceptualizations of mental disorders, meaning that distinct impairments of different neural networks might lead to similar (abnormal) behavior (as noted above) (Makris et al., 2009; Mulders et al., 2015). Unfortunately, researchers that exploit rfMRI data to investigate the functional architecture of the brain typically employ univariate modeling procedures which do not fully acknowledge the multivariate nature of the data, i.e. simultaneous activation of multiple neural networks. Therefore, such suboptimal modeling procedures impede efforts towards disentangling impairments across distinct neural systems in mental disorders.

3. Next to suboptimal modeling procedures, the rfMRI research field is confronted with the detrimental effects of structured noise in rfMRI data induced by head movement of the participant during the scanning procedure (Power et al., 2012; Satterthwaite et al., 2012; Van Dijk et al., 2012). Importantly, this concerns all types of metrics typically obtained from rfMRI, and is of particular concern for studies investigating (impaired) neurodevelopment (Satterthwaite et al., 2012).

In the following sections these aspects will be described in more detail. The current thesis will present new analytical methods to address these drawbacks and optimize research into the association of functional brain architecture and mental disorders. Subsequently, these methods have been applied to study multiple functional brain networks in the context of ADHD; a prime example of a mental disorder in which these drawbacks emerge.

Conceptualization of mental disorders

Categorical and dimensional conceptualizations

Currently, the standard instrument to diagnose mental disorders is the 5th edition of the Diagnostic and Statistical Manual of Mental Disorders (DSM-5); a handbook of the American Psychiatric Association (American Psychiatric Association, 2013a). This instrument regards a classification system for mental disorders based on behavioral symptoms in which every disorder is characterized by a cluster of signs and symptoms. Although, severity of symptoms has been more appreciated in the DSM-5 compared to previous versions of the handbook (American Psychiatric Association, 2013b), it primarily remains a categorical system. Accordingly, researchers typically investigate mental disorders using categorical designs that compare cases to controls.
A category-based system has high clinical utility because it is practical in its usage, it optimizes and simplifies communication, and because it aids clinical decision making which often comprises categorical decisions (i.e. start an intervention or not) (Coghill and Sonuga-Barke, 2012). However, categorical behavioral diagnoses seem to fail to align with underlying pathophysiological mechanisms (Hyman, 2007; Regier et al., 2009). Accordingly, multiple studies have now investigated whether the latent structure of certain mental disorders is indeed categorical or might in fact be dimensional; i.e. whether there is a discrete difference between healthy versus diagnosed individuals (categorical mechanism) or whether the diagnosed individuals represent the tail on a continuum of normal behavior (dimensional mechanism). Although some disorders appear to be either categorical (e.g., schizotypy) or dimensional (i.e., ADHD), the results of these studies have been somewhat inconsistent and the debate on categories versus dimensions in psychopathology is still an active debate (Coghill and Sonuga-Barke, 2012; Haslam et al., 2012).

Importantly, these findings are based on behavioral data and do not necessarily imply similar mechanisms at other levels (e.g. genetic, neural or neurocognitive domain). Moreover, mental disorders can be pathophysiological heterogeneous, meaning that subgroups of patients might exist with dissociable impairments; potentially with different categorical or dimensional structures (Coghill and Sonuga-Barke, 2012). For instance in the case of ADHD, the latent structure at the behavioral level is believed to be dimensional (Haslam et al., 2006, 2012; Coghill and Sonuga-Barke, 2012) while at the neural architectural level there is evidence for both categorical and dimensional mechanism, across different functional networks (Chabernaud et al., 2011; Elton et al., 2014). It is therefore thought that the way forward in understanding the pathophysiology associated with mental disorders is not to restrict research to either a categorical or dimensional view, but to integrate categorical and dimensional conceptualizations and measurements (Chabernaud et al., 2011; Rutter, 2011; Coghill and Sonuga-Barke, 2012).

To study these mechanisms researchers often use DSM-diagnosis as a categorical measure (e.g. ADHD-diagnosis) and symptom scores as a dimensional measure (e.g. amount of hyperactive behavior). Evidently, there is close relationship between such categorical and dimensional measures as they describe similar behavior (i.e. discrete versus continuous characterization). To model categorical-dimensional mechanisms such categorical and dimensional measures are typically combined into a single statistical model, testing for the unique effects of both variables (e.g. (Chabernaud et al., 2011; Elton et al., 2014)). It is however underappreciated that, due to close relationship (i.e. statistical dependency)
between these variables, such models suffer from reduced sensitivity and therefore an increased rate of false negative results (York, 2012; Mumford et al., 2015). Yet, no modeling framework has currently been proposed which addresses this drawback, impeding the believed importance of moving towards modelling and disentangling the categorical-dimensional complexity of mental disorders.

**Neural systems conceptualizations**

Single mental disorders might involve dissociative cognitive and neural deficits (Durston et al., 2011; Coghill et al., 2014). The notion that similar symptom profiles at the behavioral level can be caused by different neurobiological underpinnings is referred to as ‘equifinality’ (Cicchetti and Rogosch, 1996). Accordingly, mental disorders are increasingly conceptualized based on neural systems, e.g. in major depressive disorder and ADHD (Castellanos and Proal, 2012; Mulders et al., 2015).

As an example, I will here discuss a neural systems perspective on ADHD.

Attention-Deficit/Hyperactivity Disorder (ADHD) is a highly prevalent childhood mental disorder (Faraone et al., 2003), which persists into adulthood in 50-80% of the cases (Kessler et al., 2005). It is a heterogeneous disorder characterized by symptoms of inattention and hyperactivity/impulsivity (American Psychiatric Association, 2013a). Inattentiveness refers to a style of behavior involving high distractibility, inability to pay attention to details, lack of planning and disorganization (Rutter et al., 2008). Hyperactivity refers to an excess of fine or gross motor movements and is closely related to impulsiveness which can be described as acting without reflecting (Rutter et al., 2008).

From the early twentieth century ADHD-like hyperactivity syndromes were viewed as being the consequence of minimal brain damage or minimal brain dysfunction (Taylor, 2011). The transition to modern conceptualizations of ADHD began around 1980 with the introduction of the DSM-III. Here for the first time the term attention-deficit was coined as the most characteristic symptom domain. This was followed by an influential paper of Barkley in 1997 which stated that the core problem of ADHD lays in dysfunctional behavioral inhibition and which led to the view of ADHD as an executive function disorder (Barkley, 1997). However, deficits in executive functions were not the only cognitive deficits found in ADHD. Patients with ADHD also have variable response speed (Sergeant et al., 2003), show delay aversion and other abnormalities in processing of reward (Sonuga-Barke, 2003) and have variability in motor timing (Rommelse et al., 2008). Due to this wide spread of
cognitive deficits, ADHD turned out to be a very heterogeneous and multifactorial disorder. Accordingly, over the years, different models have been proposed to characterize ADHD (Sergeant, 2000, 2004; Sonuga-Barke, 2003).

Given the heterogeneity of the disorder, models of ADHD are shifting towards neurobiological models in which pathologies in specific neurocircuits are thought to be the underlying causes of the cognitive deficits. As an example, Durston et al. (2011) hypothesize the involvement of three neurocircuits in ADHD, in which dorsal fronto-striatal connections are mainly linked to cognitive control, orbitofrontal-striatal circuits are linked to reward processing and fronto-cerebellar circuits are linked to timing (see Figure 1). Moreover, a recent review and meta-analysis discusses ADHD in the context of impairments across distinct neural networks (Thomas Yeo et al., 2011). Several large-scale neural networks have now been implicated in ADHD, comprising localized networks related to visual and motor cortices as well as networks distributed across association cortex (e.g. the default mode and executive control network) (Castellanos and Proal, 2012; Cortese et al., 2012). Consequently, the research field is moving towards neural systems-based conceptualizations of ADHD (Makris et al., 2009; Castellanos and Proal, 2012; Cortese and Castellanos, 2012; Cortese et al., 2012; Faraone et al., 2015).

Figure 1 - Relations between behavior, cognition and brain circuits in ADHD (figure is copied from Durston et al. (2011))
Modelling functional brain networks

Resting-state fMRI

To investigate the functional architecture of neural networks we can exploit functional MRI. This technique is based on the fact that neural activity within a brain region alters the metabolic rate and oxygen consumption of that area, inducing a change in the concentration of deoxyhemoglobin and oxyhemoglobin (Poldrack et al., 2011). This hemodynamic response is called the Blood Oxygenation Level Dependent (BOLD) response. Since deoxyhemoglobin and oxyhemoglobin have different magnetic properties (paramagnetic versus diamagnetic) the change in hemoglobin concentration causes a change in the local magnetic field which can be measured by functional MRI. Accordingly, signal fluctuations observed using fMRI are thought to indirectly relate to neural activity.

Typically, researchers exploited fMRI to study the functionality of the brain by eliciting neural activity induced by specific stimuli (e.g. visual input) and tasks to be performed by the subject. This is referred to as task-based fMRI. Spontaneous modulation of the BOLD signal during 'rest', i.e. unrelated to explicit input or output, has long been regarded as noise. However, this view changed dramatically after a study by Biswal et al. (1995) which showed that signals from distant brain areas (in their study left and right hemispheric regions of the primary motor network) were highly correlated during rest. Such synchronization in functional time-series during rest of anatomically separated brain regions is now believed to reflect functional communication between these regions and is therefore used to investigate functional connectivity as described by Friston (1994). This field of research gained popularity when fMRI data obtained during rest was successfully used for investigating functional connectivity of a hypothesized network referred to as the default mode network (Greicius et al., 2003). This type of fMRI research is referred to as resting-state fMRI (rfMRI). Compared to 'traditional' task-based fMRI the main advantages of rfMRI are the ability to study multiple networks at the same time, requiring minimum participant compliance, and being easily applicable across development; therefore providing a high potential for clinical application (Castellanos et al., 2013).

Functional brain networks

Data-driven analyses of rfMRI data using Independent Component Analysis (ICA) has led to the identification of multiple resting-state networks (RSNs) (Beckmann et al., 2005; Damoiseaux et al., 2006; De Luca et al., 2006; Smith et al., 2009). Such networks consist each of a set of brain regions with high synchronization in
spontaneous functional time-series. Figure 2 shows an overview of the RSNs as typically identified using ICA (Beckmann et al., 2005). Interestingly, these networks are biologically meaningful as they consist of regions that overlap in both function and neuroanatomy. Importantly, these networks, obtained when the brain is at ‘rest’, correspond to networks that activate during under functional tasks (Smith et al., 2009). The RSNs reflect sensory systems as well as networks involved in higher-level cognitive processes (Beckmann et al., 2005; Smith et al., 2009). Examples of sensory network are networks related to visual and auditory processing, and motor functioning (Beckmann et al., 2005; Smith et al., 2009). RSNs related to higher-level cognitive processes (e.g. attention, inhibition and working memory) concern the executive control network and the left/right lateralized fronto-parietal networks (Beckmann et al., 2005; Smith et al., 2009). In addition, a well-studied RSN is the ‘default mode network’ (DMN) which is associated with task-irrelevant mental processes, memory and mind wandering (Raichle et al., 2001; Raichle and Snyder, 2007; Buckner et al., 2008).

Figure 2 - Resting-state networks as estimated using spatial ICA in a study by Beckmann et al. (2005) (figure is copied from (Beckmann et al., 2005)). The images are co-registered to the MNI template; the coordinates represent distance in mm. The identified networks are respectively referred to as: A) medial visual network, B) lateral visual network, C) auditory network, D) sensory-motor network, E) default mode network, F) executive control network, G) right fronto-parietal network and H) left fronto-parietal network.
Methods for modeling brain networks

In general, statistical methods for investigating functional connectivity using rfMRI analysis data can be divided into two main groups: seed-based approaches (seed-based correlation analysis; SCA) and data-driven approaches (most typically, independent component analysis; ICA) (Li et al., 2009). Box 1 discusses the methodological principles of both SCA and ICA.

Seed-based approaches require a priori definition(s) of a single or multiple regions of interest (seeds). The location of a seed is determined a priori and can depend on the research question and hypothesis. As such, this method allows testing region-specific research questions. However, results will highly depend on the accuracy and validity of the selected seed(s) and are therefore driven by methodological choices of the researcher (Ma et al., 2007; Cole et al., 2010). Researchers, for instance, frequently define regions based on a pre-defined anatomical atlas, though it is questionable whether this is valid since the correspondence between anatomical and functional regions is unclear (Fornito et al., 2013). Moreover, SCA is a univariate method which undermines the multivariate nature of fMRI data; voxel time-series are in fact mixtures of multiple signals, containing structured noise and signal related to other functional networks (Smith et al., 2012). As a result any obtained functional connectivity map using SCA cannot validly be interpreted as a distinct neurobiological system (Cole et al., 2010).

A popular alternative to SCA is ICA. This is an approach to study functional connectivity by estimating spatial maps of RSNs in a data-driven fashion. This technique operates under the assumption that the observed (fMRI) data is a mixture of multiple independent signal sources (e.g. BOLD fluctuations, artifacts, etc.). ICA decomposes the fMRI data into a set of components which are spatially independent. This results in a set of estimated spatial maps, each associated with a specific time-course, reflecting regions with similar signal fluctuations. Accordingly, some of these maps represent spatial patterns of synchronized BOLD fluctuations, i.e. RSNs. To obtain RSNs which are representative for multiple subjects, the data of these subjects can be combined in to a single group-ICA analysis (Beckmann et al., 2009). Subject-specific representations of these RSN maps can subsequently be obtained using a multivariate linear regression procedure referred to as dual regression (Beckmann et al., 2009; Filippini et al., 2009). See Box 1 for more detailed information on ICA and dual regression. By providing a valid multivariate method for modeling distinct neurobiological systems (RSNs), ICA + dual regression offers a great potential for investigating neural correlates of mental disorders across different neural networks. This approach has been shown to provide functional connectivity metrics which higher reliability and improved detection of inter-
individual differences compared to SCA (Zuo et al., 2010; Smith et al., 2014; Zuo and Xing, 2014).

**Box 1 – Methods for modeling functional brain networks**

*Seed-based correlation analysis*

Seed-based correlation analysis (SCA) is an example of a model-based approach, and has been frequently used in the field of fMRI research. This method requires a spatial model by means of a priori selected voxels or cortical region (i.e. seed region). Subsequently, the time-series are extracted from all these voxels which are then combined into one single representative time-series for the full region (e.g. by averaging the voxel-specific time-series). This time-course can now be used as a regressor for a linear correlation or linear regression against the time-series of any voxel within the brain, representing the functional connectivity of that voxel with the seed region. Typically, a regression or correlation coefficient is calculated for every voxel in the brain, providing a whole-brain functional connectivity map. Alternatively, instead of defining a single brain region and determining the functional connectivity of every voxel with that seed, one can define multiple seed regions (e.g. 264 brain regions identified by Power et al. (2011)) and determine the temporal correlation between each pair of seeds. This provides a matrix representation of whole-brain functional connectivity, although at a much sparser level.

*Data-driven analysis*

The most common data-driven approach to study functional connectivity is spatial independent component analysis (ICA). This is a so called ‘blind-source separation’ method (Jutten and Herault, 1991). The ICA algorithm operates under the assumption that each fMRI scan is an observation of a mixture of spatially independent sources (e.g. a neural network). ICA estimates these latent sources by decomposing the fMRI data into a set of spatial maps (called independent components) and associated time-courses, such that the statistical dependency of the spatial maps is minimized. The underlying rationale for this approach is based on the central limit theorem which states that the distribution of a mixture of signals will always be more Gaussian distributed than the constituent signals, implying that the decomposition which results in the most non-Gaussian spatial maps provides the solution with the highest statistical independency between components. Algorithms implementing ICA differ in which measure of non-gaussianity is used, how this measure is minimized/maximized, and whether any preprocessing has been applied to the data (e.g. data reduction and/or prewhitening). The implementation which will be used throughout this thesis is the probabilistic ICA approach called Multivariate Exploratory Linear Decomposition into Independent Components (MELODIC) which is part of the image analysis toolbox FSL (Smith et al., 2004; Woolrich et al., 2009; Jenkinson et al., 2012).

The components estimated by ICA represent various signal sources that together constitute the fMRI signal. These components are spatial maps representing neural networks (referred to as resting-state networks; RSNs) but also motion, physiological and scanner artifacts (noise components). However, ICA will require identification of the estimated components, for instance to identify the RSN of interest for further research. Accordingly, components need to be identified by visual inspection or an automated detection algorithm. Moreover, running ICA on separate subjects produces
variable results of estimated components, e.g. a networks defined in one subject might be split into two networks in another subject. These aspects raise practical challenges for the utility of ICA for group-level analysis in which one would ideally compare the same neural network in multiple subjects and therefore requires a one-to-one mapping of the components across these subjects.

Although a template-matching procedure has been proposed for automated identification of networks, this procedure has shown to be unreliable and does not solve the issue of potentially split networks across subjects (Greicius et al., 2004; Garrity et al., 2007; Zuo et al., 2010). In contrast, a principled approach is the use of a top-down group-ICA approach (Calhoun et al., 2001). Here, the first step is to combine the fMRI data of multiple subjects into a single higher-level ICA decomposition, resulting in a set of spatial components that describe common signal sources within and between subjects. Subsequently, for each subject, subject-level representations of these components are derived using back reconstruction (Calhoun et al., 2001) or dual regression (Beckmann et al., 2009; Filippini et al., 2009). Back reconstruction uses a subspace of the original subject-specific fMRI data. However, the information contained in this subspace is not guaranteed to be consistent across subjects which therefore still complicates between-subject analysis. The dual regression approach however does use the full original dataspace of the subject. The group-level components are used as spatial regressors to the full original data to acquire time-courses which are then normalized and used as temporal regressors in a second regression procedure to derive the spatial maps associated with the components (Beckmann et al., 2009; Filippini et al., 2009). This procedure therefore provides a method to obtain individualized time-courses and spatial maps of the components determined by group-ICA, which allow for valid between-subject analysis.

### Handling noise in fMRI data

Much like any raw medical imaging data, rfMRI data needs to be preprocessed to allow valid further analyses. Box 2 outlines the typical steps involved such fMRI preprocessing. Although partly accounted for in the preprocessing, various sources of noise can contribute to the rfMRI signal. Box 3 gives an idea of these various noise sources. Among all types of artifacts present in rfMRI data, head motion artifacts are of particular interest and gained increased attention in recent literature.

### Head motion artifacts

In 2012 three studies were published, respectively by Van Dijk et al. (2012), Power et al. (2012) and Satterthwaite et al. (2012), which showed that functional connectivity metrics obtained from rfMRI data are heavily impacted by participant head motion despite conventional preprocessing methods to remove motion artifacts. More specifically, these studies showed that residual motion-related noise in the
Introduction

data introduces spurious temporal similarities between voxel time-series and bias
the estimation of functional connectivity metrics (Power et al., 2012; Satterthwaite
et al., 2012; Van Dijk et al., 2012). Since younger individuals tend to present higher
levels of head motion and motion might act as a disease trait (e.g. patients with
ADHD may be expected to present more head motion artifacts than non-ADHD
controls), this becomes a major concern for investigating functional connectivity
related to neurodevelopment and specific diseases (Satterthwaite et al., 2012; Van
Dijk et al., 2012).

To remove residual motion-related noise it has been proposed to detect time-
points associated with high motion and subsequently fully delete or regress out these
fMRI volumes from the data (Power et al., 2012; Satterthwaite et al., 2013).
Detection of such time-points can be established using a summary time-series for
frame-to-frame head displacement (i.e. head movement between subsequent scans)
and/or global signal intensity. When, at a specific time-point, a predefined threshold
is exceeded, that particular volume is regarded as motion artifact and removed from
the data. Importantly, such a procedure will inherently leave subthreshold artifacts
in the data, potentially remove signal of interest, and will lead to inter-subject
variability in degrees of freedom which violates typical assumptions made in higher-
level statistical analysis (see Chapter 1 and 2). Accordingly, the neuroimaging
research field is lacking an efficient method for removal of motion artifacts from
rfMRI data which especially complicates neurodevelopment and psychiatric brain
research.

Box 2 – Preprocessing of fMRI data

Quality control
Although often overlooked, the first step in preprocessing fMRI data is a quality check
of the raw data. This can be achieved by using a tool to watch the fMRI time-series and
visually inspect the data, e.g. using fsview in the toolbox of FSL. Especially,
detrimental MR-scanner artifacts can often be detected visually (see below for more
information on MR-scanner artifacts).

Slice time correction
A single three-dimensional fMRI volume is typically collected by collecting two-
dimensional images (slice), one slice at a time, in either a sequential (descending or
ascending) or interleaved (odd-even) order. Consequently, data between slices is
systematically collected at a slightly different time-point, where the difference between
the first and last slice will be almost equal to the volume acquisition time. Potentially,
researchers can correct for these timing differences. However, slice time correction is
subject of debate as since the correction procedures might actually propagate artifacts
from one image through the full time-series. In addition, the impact of slice timing
differences appears to be minor in case of data acquired within the repetition time used
in typical fMRI sequences (<2 seconds).
Introduction

**Motion correction**
During the fMRI data acquisition subjects typically tend to (slightly) move their head. As a consequence, the 3D images acquired over time are not aligned properly, i.e. a specific voxel represents a different brain area over time. To correct for this misalignment, all acquired 3D images have to be aligned to a single target image (e.g. the image acquired in the middle of the acquisition). Since the geometry of the brain does not change during the acquisition the correction can be accomplished by a rigid body transformation of every brain image with respect to the target image using six degrees of freedom (DOF); translation in x, y and z direction, and rotation around the x, y and z axis (respectively called roll, pitch and yaw).

**Spatial smoothing**
Regions of activation are typically larger than one voxel, inducing spatial correlations between neighboring voxels. Moreover, voxels at the edge of an activated region will only partially represent the activated area (partial volume effect) and partially reflect non-activated tissue. For these reasons, fMRI images are spatially smoothed. By reducing some of the variability between neighboring voxels spatial smoothing effectively increases the signal to noise ratio. The smoothing is done using a Gaussian kernel with a common full-width half maximum (FWHM) of around 4-12mm. In practice this means that intensities of neighboring voxels are weighted and added to the voxel itself, the FWHM determines the weighting of the neighboring voxels. Since a voxel is not a physiological entity but an arbitrary measure, smoothing is said to result in data that better resembles the underlying brain tissue better. The SNR is increased if the size of the expected activation area is larger than the amount of smoothing, as such too much smoothing will in fact reduce small activation areas.

**Global intensity scaling**
A last preprocessing step is global intensity scaling, which means that the mean intensity of a brain volume over time is normalized over all subjects such that every subject has the same mean intensity over time. This is because the intensity produced by the scanner is different for every participant. While differences between voxels are meaningful and comparable across participants, the exact height is meaningless.

**Temporal filtering**
To remove low-frequency drifts in the data, either due to scanner noise or to physiological noise like respiration and the cardiac cycle, the fMRI data is temporally high-pass filtered. A typical value of the filter is around 0.01Hz. Since the Nyquist frequency for a typical TR of two seconds is 0.25Hz (Nyquist frequency = 1/2 * sample frequency) low frequency drifts below this frequency (as respiration etc.) will be represented in the data by even lower frequencies due to aliasing. Low pass filtering is also possible and will reduce high frequency artifacts. However, although rfMRI research has typically focused on BOLD fluctuations in the low-frequency range (<0.1 Hz) (Cordes et al., 2001), we currently know that useful information is also present at higher frequencies (Niazy et al., 2011; Liao et al., 2013; Kalcher et al., 2014), which advocates against the use of low pass filtering the data (Niazy et al., 2011; Liao et al., 2013; Kalcher et al., 2014).

**Nuisance regression**
Residual artifacts after the previously described image preprocessing can subsequently be attenuated or removed using nuisance regression. With this procedure
a set of confounding time-series are defined which are subsequently regressed out of the data using a linear regression (e.g. ordinary least squares regression). These confounding time-series typically comprise a set of parameters to specifically remove motion artifacts, complemented with a small set of time-series to remove any type of structured noise. The set of motion parameters describes translational and rotational head movement during the fMRI acquisition (obtained from the realignment procedure described previously in this Box). The additional time-series typically comprises averaged MRI signals across voxels respectively located within white matter and/or cerebrospinal fluid; since neural activation is expected to be specifically located within grey matter, these time-series are thought to represent a composite measure of structured noise.

Box 3 – Artifacts in functional MRI data

**MR-scanner artifacts**

The most typical examples of MR-scanner artifacts are spike, ghosting and susceptibility artifacts. Spike artifacts are caused by electrical instability of the scanner and typically appear as stripes across the fMRI image. Ghost artifacts, also referred to as Nyquist or N/2 ghost artifacts, on the other hand are specifically related to a type of sequence often used in fMRI scanning (echo planar imaging) and can be seen by the appearance of ‘ghost’ images which are shifted by half the field of view in the so called ‘phase encoding direction’ of the sequence. In addition, artifacts might appear near air-tissue interfaces in the brain. These interfaces cause local inhomogeneities in the magnetic field that cause a drop in the detected MRI signal and spatially distort the image at that specific location (susceptibility artifact).

**Motion artifacts**

Head movement during the MRI data acquisition causes acquired volumes/scans to be misaligned, which is addressed by realigning the volumes over time (see Box 2). However, head motion also induces secondary artifacts. First, head motion during the acquisition of a single brain volume will cause brain areas to move into neighboring slices. Dependent on the direction of the movement these brain areas will not get excited or excited twice, causing so called ‘spin history effects’ (Friston et al., 1996). These effects are reflected by large intensity changes in the acquired scans. Second, as noted before, the main magnetic field of the MR-scanner contains small inhomogeneities. Therefore, head motion will cause voxels to endure small differences in local magnetic fields strength, again impacting the MRI signal. Third, head movement will interact with the accuracy of the volume realignment procedure. First of all, the amount of head motion will determine the extent to which the data has to be interpolated when aligning the target image to a reference image. An additional effect on how head movement interacts with the realignment procedure is related to magnetic inhomogeneities. As noted before, these inhomogeneities are induced by air-tissue interfaces and might cause spatial distortions in the acquired scan. Head movement will change the location of these interfaces and therefore cause variable magnetic field inhomogeneity and spatial distortions over time. Since this realignment algorithm assumes the spatial properties of the brain (e.g. shape) to be constant over time, such variable spatial distortion will impact the accuracy of the algorithm.
Introduction

Physiological artifacts
Artifacts can also be induced by physiological noise related to heartbeat and respiration (Murphy et al., 2013). Indeed, it has been shown that BOLD fluctuations significantly correlate with cardiac and respiratory cycling rates, indicating the artifactual impact of such physiological noise on fMRI data (Birn et al., 2006; Shmueli et al., 2007). The heartbeat will cause pulsatile motion of blood vessels, which generates small tissue movements with a frequency related to the cardiac cycle (Dagli et al., 1999). Effects due to respiration on the other hand are caused by movement of the chest which causes changes in the inhomogeneity of the magnetic field located at the head, with a frequency related to the respiration rate (Raj et al., 2001; Brosch et al., 2002). In addition, both cardiac pulsation and respiration will induce movement of the brain stem causing deformation of brain tissue and movement of cerebrospinal fluid which induce changes in the main magnetic field (Dagli et al., 1999). Importantly, the cardiac and respiratory cycles have a relative high frequency compared to the low-frequency BOLD fluctuations. However, physiological noise related to cardiac and respiratory effects will be aliased into the low-frequency range due to the low sample rate at which fMRI data is typically acquired (a sampling frequency of ~0.3-0.5 Hz, i.e. 2-3 seconds per scan) (Lowe et al., 1998; Bhattacharyya and Lowe, 2004). Next to these movement-related effects, is also has been shown that arterial CO2 concentration and blood pressure changes related to the cardiac and respiratory cycle impact the fMRI signal (Murphy et al., 2013).

Thesis outline

The previous sections discussed the large potential of rfMRI for investigating mental disorders and potential translation to clinical application. However, three main aspects are currently suboptimally addressed in this type of research. First, the lack of an efficient method for removal of motion artifacts from rfMRI data. Second, the lack of a modeling framework to account for, and disentangle, the categorical-dimensional complexity of mental disorders. Third, insufficient incorporation of neural systems conceptualizations of mental disorders by relying on univariate modeling procedures to investigate distinct neurobiological systems, rather than exploiting multivariate approaches which allows a more valid and integrated interpretation.

The aim of the current thesis is to address these drawbacks and develop new methodology to optimize research into the association of functional brain architecture and mental disorders. Developed methodology will be applied to study the functional neural architecture of ADHD. This disorder is the ultimate test case to apply the proposed methodology since it is a prime example of a mental disorder in which the discussed drawbacks emerge: 1) it presents as a dimensional mechanism at the behavioral level while there is evidence for both categorical and
Introduction

dimensional neurofunctional underpinnings (Haslam et al., 2006; Chabernaud et al., 2011; Elton et al., 2014), 2) dissociable networks have shown to be associated with the disorder (Castellanos and Proal, 2012; Cortese et al., 2012), and 3) since the disorder is associated with hyperactive symptoms these subjects can be expected to present increased levels of head motion artifacts in rfMRI data.

Accordingly, Chapter 1 presents the development and validation of a data-driven ICA-based method for automated detection and removal from functional MRI data (ICA-AROMA: ‘ICA-based Automatic Removal of Motion Artifacts’). Subsequently, Chapter 2 presents an extensive comparison of ICA-AROMA against a broad range of alternative strategies with respect to their ability to remove motion artifacts while preserving signal of interest. Next, Chapter 3 presents a statistical framework to identify and characterize categorical and dimensional mechanisms in mental disorders, while preserving sensitivity to the effects of interest. Finally, Chapter 4 integrates all proposed methodology by using ICA-AROMA, dual regression and the proposed categorical-dimensional modeling framework to study functional networks in the context of ADHD. Importantly, this study validates the obtained results by mapping the identified functional neural correlates to cognitive and genetics measures known to be associated with ADHD.

References


Melle, I., 2013. The Breivik case and what psychiatrists can learn from it. World Psychiatry 12, 16–21.
Introduction


Introduction
Chapter 1

ICA-AROMA: A robust ICA-based strategy for removing motion artifacts from fMRI data

Published as:
Abstract

Head motion during functional MRI (fMRI) scanning can induce spurious findings and/or harm detection of true effects. Solutions have been proposed, including deleting ('scrubbing') or regressing out ('spike regression') motion volumes from fMRI time-series. These strategies remove motion-induced signal variations at the cost of destroying the autocorrelation structure of the fMRI time-series and reducing temporal degrees of freedom. ICA-based fMRI denoising strategies overcome these drawbacks but typically require re-training of a classifier, needing manual labeling of derived components (e.g. ICA-FIX; Salimi-Khorshidi et al., (2014)). Here, we propose an ICA-based strategy for Automatic Removal of Motion Artifacts (ICA-AROMA) that uses a small (n = 4), but robust set of theoretically motivated temporal and spatial features. Our strategy does not require classifier re-training, retains the data's autocorrelation structure and largely preserves temporal degrees of freedom. We describe ICA-AROMA, its implementation, and initial validation. ICA-AROMA identified motion components with high accuracy and robustness as illustrated by leave-N-out cross-validation. We additionally validated ICA-AROMA in resting-state (100 participants) and task-based fMRI data (118 participants). Our approach removed (motion-related) spurious noise from both rfMRI and task-based fMRI data to larger extent than regression using 24 motion parameters or spike regression. Furthermore, ICA-AROMA increased sensitivity to group-level activation. Our results show that ICA-AROMA effectively reduces motion-induced signal variations in fMRI data, is applicable across datasets without requiring classifier re-training, and preserves the temporal characteristics of the fMRI data.
Introduction

Head motion during functional Magnetic Resonance Imaging (fMRI) scanning results in misalignment of one volume to the next. This introduces measurement inaccuracies as imaging voxels do not represent identical brain regions over time. Primary effects of participant head motion in fMRI data are corrected by realigning volumes using linear alignment algorithms. However, head motion does not only result in misaligned volumes, but also induces secondary effects related to partial voluming, interpolation effects, magnetic field inhomogeneities, intra-volume motion, and spin-history effects (Friston et al., 1996; Beall and Lowe, 2014), which cannot be corrected for by using volume-realignment. The most common strategy to correct for these secondary effects is to model participant head motion and remove the modeled responses from the fMRI data using additional linear regressors within the framework of the General Linear Model (GLM; (Friston et al., 1996)). Recent findings have spurred renewed debate on the impact of participant head motion on resting state fMRI (rfMRI) experiments particularly. Most prominently, participant head motion during an rfMRI measurement could induce spurious temporal correlation between brain regions, even in light of generally adopted strategies for motion-induced artifact correction (Power et al., 2012; Satterthwaite et al., 2012; Van Dijk et al., 2012). Functional connectivity measures derived from rfMRI could be particularly affected by spurious temporal correlations as they investigate temporal correlations in the absence of a task-related model. Accordingly, in light of increased head motion in younger participants or the possibility of motion as a disease trait (e.g. patients with attention-deficit/hyperactive disorder), discriminating signal from noise when investigating developmental or disease-related neural signatures will be complicated by interactions with head motion (Van Dijk et al., 2012; Satterthwaite et al., 2013b).

In contrast to model-free functional connectivity analyses, typical task-based fMRI analyses investigate BOLD activity related to an experimental model and realignment-based measurements of head motion can be included as covariates. However, in addition to ‘spontaneous movement’, task-based fMRI analyses can be affected by stimulus-related motion, which cannot be separated from stimulus-related signal variations of interest in the regression model. Accordingly, regression-based motion artifact removal strategies potentially remove signal of interest, thereby decreasing sensitivity to functional activation.

The most common strategy for dealing with secondary effects of participant head motion in fMRI data is nuisance regression. Typical nuisance regression models include 6 to 24 motion-related covariates derived from volume-realignment.
parameters (Friston et al., 1996; Satterthwaite et al., 2013a; Yan et al., 2013a). Six or 12 motion-related covariates were initially considered sufficient, but currently 24 covariates are recommended (Satterthwaite et al., 2013a; Yan et al., 2013a). In addition, and specifically for rfMRI data, recent strategies have proposed excluding volumes associated with high motion from the fMRI time-series (Power et al., 2012, 2014; Satterthwaite et al., 2013a). Two (analogous) strategies accomplish this goal by respectively regressing out (‘spike regression’; Lemieux et al., 2007; Satterthwaite et al., 2013a) or deleting individual high-motion volumes from the fMRI time-series (‘scrubbing’; Power et al., 2012).

The strategies above aim to rigorously implement correction for motion-related artifacts, yet several drawbacks can be identified. First, each strategy relies on the realignment parameters (RPs) obtained from realigning volumes during primary correction for motion in fMRI data. Naturally, these parameters can only be as accurate as the algorithm used for realignment. Second, the regression-based strategies typically only model linear motion-induced signal variation while the underlying dynamics are non-linear (Fair et al., 2012), i.e. secondary effects of head motion are not necessarily captured by the obtained realignment parameters. Third, the use of a large set of nuisance regressors may lead to overfitting of the data and therefore to removal of signal of interest (Satterthwaite et al., 2013a; Zuo et al., 2013). Fourth, by removing specific high-motion volumes, spike regression and scrubbing destroy the autocorrelation structure of the data. This will impact frequency filtering typically employed within fMRI preprocessing (Carp, 2013) and prevent any analysis that is aimed at investigating frequency characteristics (e.g. amplitude low-frequency fluctuations; ALFF) or non-stationarity in functional connectivity (Yan et al., 2013a). Finally, such volume removal or regression results in a high and variable loss of temporal degrees of freedom (tDoF; Yan et al., 2013). As an example, in two reported high-motion cohorts, the mean amount of deleted volumes was respectively 26% and 58% (Power et al., 2012). The number of available volumes is typically regarded as the available number of tDoF, making it clear that the associated loss in statistical power induced by spike regression or scrubbing can be substantial and can differ substantially between subsets of subjects after volume removal/regression. Importantly, tDoF determine the estimation accuracy of subject-level statistics. Reduced tDoF for instance affect the error variance within a typical single-subject, first-level regression. Accordingly, although spike regression and scrubbing can reduce the association between motion artifact and measures of interest across a population (Yan et al., 2013b), they can introduce an association between the amount of motion and the accuracy of single-subject statistics. This, in turn, introduces heteroscedasticity at the between-subject level.
To avoid such drawbacks, Yan et al., (2013a) suggested employing of extensive nuisance regression, including 24 RPs, at the single-subject level and including motion covariates in group-level analyses. However, group-level motion covariates can share variance with variables of interest and therefore reduce sensitivity to an effect of interest. As an example, motion might act as a trait (Van Dijk et al., 2012; Couvy-Duchesne et al., 2014; Kong et al., 2014) and have a neural basis (Pujol et al., 2014; Zeng et al., 2014). Moreover, group-level covariates do not correct for inferior estimation of single-subject effect sizes in the presence of high levels of noise at the subject level.

Finally, although controversial, global signal regression and band-pass filtering are frequently considered for confound removal. Global signal regression reduces the impact of motion on functional connectivity metrics (Satterthwaite et al., 2013a; Yan et al., 2013a) but it inevitably removes signal of interest as the global signal is a superposition of both signal and noise components. Additionally, global signal regression introduces anti-correlations and alters connectivity structure (Murphy et al., 2009; Weissenbacher et al., 2009; Yan et al., 2013b). Similarly, temporal band-pass filtering removes signal of interest present at higher frequencies (Niazy et al., 2011; Liao et al., 2013; Kalcher et al., 2014).

The drawbacks of current strategies to remove motion-induced signal variations from fMRI data are at least partly addressed by alternative strategies that aim to identify and remove motion-related artifacts using Independent Component Analysis (ICA). Applied to fMRI data, ICA decomposes the data into a set of spatial independent component maps (ICs), and associated time-courses (McKeown et al., 1998; Beckmann and Smith, 2004; Beckmann et al., 2005). The resulting components represent brain activity, and/or structured noise (e.g. motion-related, physiological or scanner-induced noise). Components representing noise can be regressed out from the data, implying that ICA can be used to remove noise from fMRI data in a data-driven fashion (Thomas et al., 2002; Kelly et al., 2010; Kundu et al., 2012). However, labeling ICA components as noise or signal of interest is a subjective and time-consuming process. Multiple methods have been developed to automatically identify noise components based on temporal and/or spatial features (Thomas et al., 2002; Kochiyama et al., 2005; De Martino et al., 2007; Perlberg et al., 2007; Tohka et al., 2008; Kundu et al., 2012; Bhaganagarapu et al., 2013; Rummel et al., 2013; Storti et al., 2013; Salimi-Khorshidi et al., 2014). These methods have not been widely adopted due to lack of accuracy or extensive validation over multiple datasets. Some strategies have specific disadvantages such as being only applicable to task-based fMRI data (Kochiyama et al., 2005) or multi-echo fMRI data (Kundu et al., 2012), being limited to physiological noise (Thomas et al., 2002; Perlberg et al., 2007) or requiring to re-train the classifier for every new
dataset acquired using a different MR scanner and/or MRI protocol (Thomas et al., 2002; De Martino et al., 2007; Salimi-Khorshidi et al., 2014).

To address current issues associated with both motion parameter- and ICA-based strategies for motion artifact removal from fMRI data, we propose an alternative ICA-based strategy. In contrast to generic ICA-based denoising strategies (Salimi-Khorshidi et al., 2014), we here focus specifically on the classification and removal of components that specifically relate to head motion. In the light of an increasing number of large-scale multi-site studies, we aim to develop a robust strategy that does not require classifier re-training across datasets. To that end we construct a classifier using a limited and theoretically motivated set of features. These features are defined a priori and correspond to component characteristics which are typically evaluated during manual classification. Note that such an approach substantially differs from an approach in which a complex classifier is trained using an extensive set of features, inherently increasing the probability of a biased classifier towards the training dataset. Specifically, our strategy implements a classifier that employs two temporal and two spatial features. The ICA-based denoising is applicable to rfMRI and task-based fMRI data, largely preserves the autocorrelation structure of the fMRI time-series, and has little impact on the tDoF, thereby avoiding heteroscedasticity in group-level statistics.

In the first section of this manuscript we discuss our ICA-based strategy, its features, classifier construction, and development. The second section comprises an evaluation of the classifier and a validation study in which we applied our strategy to rfMRI and task-based fMRI data. In both datasets we investigated the removal of group-level spurious noise, sensitivity to activation, and loss in tDoF. Throughout all assessments we compared results obtained using ICA-AROMA to preprocessing with extensive nuisance regression (Satterthwaite et al., 2013a; Yan et al., 2013a) and spike regression (Satterthwaite et al., 2013a). Of note, a complete evaluation of ICA-AROMA against alternative strategies for removing motion-related artifacts was beyond the scope of this manuscript. For such an evaluation we refer the reader to a companion manuscript where we compared nine strategies by assessing the achieved quality of motion artifact removal, preservation of signal of interest, and replication across multiple rfMRI datasets (Pruim et al., 2015b).
ICA-AROMA method

Methods

ICA-AROMA

Figure 1 provides an overview of our ICA-based strategy for motion artifact removal called ‘ICA-AROMA’ or ‘ICA-based Automatic Removal of Motion Artifacts’. Within the typical fMRI participant-level preprocessing stream ICA-AROMA is applied after spatial smoothing but prior to high-pass filtering and further nuisance regression. ICA-AROMA includes three consecutive steps. The first step is a probabilistic ICA on the partly preprocessed single-subject fMRI data. Next, ICA-AROMA exploits a set of four discriminative features and a classification procedure to identify ICs representing motion artifacts. Finally, the selected components are removed from the fMRI time-series using linear regression.
Step 1—Probabilistic Independent Component Analysis

Probabilistic ICA is achieved using Multivariate Exploratory Linear Decomposition into Independent Components (MELODIC, part of the FMRIB Software Library (FSL), version v5.0, available at http://www.fmrib.ox.ac.uk/fsl; Smith et al., (2004); Woolrich et al., (2009); Jenkinson et al., (2012)). We employed MELODIC with automatic estimation of the number of independent components (Beckmann and Smith, 2004). This results in a set of spatial ICs (Z-statistical maps) and associated time-courses for each participant independently. For component classification, we additionally obtained power spectra for each IC time-course.

Step 2—component classification

Next, ICA-AROMA uses a predetermined classifier to identify components that represent motion-related artifacts by assessing each component in light of its high-frequency content, correlation with realignment parameters, edge fraction, and CSF fraction. Specifically, an IC was classified as motion-related when it exceeded at least one of three criteria: 1) exceeding a decision boundary combining the edge fraction and maximum RP correlation, 2) a CSF fraction N10%, or 3) a high-frequency content N35%. See Section 'Features for component selection' for specific details on the features. Of note, the ICA-AROMA classifier does not require retraining when applied to a new dataset.

Step 3—data denoising

Finally, ICs identified as motion artifacts are removed from the fMRI data. To this end, we conduct an ordinary least squares regression on the data \( Y = Xb + e \), using the full set of IC time-series \( X \) as a design matrix (i.e. mixing matrix as estimated by ICA), giving parameter estimates \( b = X^+ Y \). Both the design matrix and the matrix of parameter estimates can then be partitioned into a noise \( X_n \tilde{b}_n \) and signal \( X_s \tilde{b}_s \) part, according to the automatic classification. The variance specifically associated with the motion-classified ICs can subsequently be subtracted from the data; \( Y_{denoised} = Y - X_n \tilde{b}_n \). This method is implemented in the FSL command fsl_regfilt.

Features for component selection

ICA-AROMA uses a combination of four features to identify components that represent motion-related artifacts. These features were derived from the frequency spectrum, time-course, and spatial map of each IC and will be further referred to as ‘high-frequency content’ (HFC), ‘maximum correlation with realignment parameters’ (maximum RP correlation), ‘edge fraction’ and ‘CSF fraction’. Of note, these features were not selected through an extensive feature selection procedure,
but were specifically based on theoretical arguments and criteria typically used during manual denoising. Below we describe each feature in more detail.

**Temporal features**

*High-frequency content.* Time-series of BOLD-related ICs are typically dominated by low frequencies, given that the haemodynamic response function acts like a low-pass filter. In contrast, this is not specifically the case for motion-related ICs given that these are not of neuronal origin and therefore not smeared temporally by the haemodynamic effect. Especially motion ICs related to secondary motion effects like spin-history induced signal variations will show significant power at high frequencies. We defined the high-frequency content feature (HFC) as the frequency \( f_{0.12} \), expressed as a fraction of the Nyquist frequency \( f_n \), at which the higher frequencies explain 50% of the total power between 0.01 Hz and \( f_n \). In other words; \( f_{0.12} / f_n \) after solving \( f_{0.12} \) from \( \int_{0.01}^{f_{0.12}} F / \int_{f_{0.12}}^{f_n} F = 0.5 \) for a given power spectrum \( F \) and \( f_n \). It reflects the ‘tendency’ of the frequency spectrum towards high frequencies as higher values indicate increased power in the higher frequencies of the spectrum.

Note that rather than fixing a threshold value to a given value, e.g. as is typically implicitly done by means of temporal band-pass filtering into a 0.01 Hz–0.1 Hz range, we here derive a threshold as a proportion of the full frequency range (up to the Nyquist-limit). In cases where investigators are using fast TR data (e.g. data generated using multi-band sequences) the threshold frequency will also increase. This reflects the fact that an increasing body of work has now convincingly demonstrated that the low-frequency characteristics of resting-state signals is induced by the haemodynamic convolution, i.e. that the underlying neuronal signatures are better described as being broadband up to Nyquist, and that useful information therefore is also available in higher frequencies (which in the case of fast TR measurements are better resolved in the measured observations) (Niazy et al., 2011; Liao et al., 2013; Kalcher et al., 2014).

*Maximum correlation with realignment parameters.* Realignment parameter (RP) time-series are derived from the volume-realignment algorithm (e.g. FSL mcflirt, AFNI 3dvolreg, SPM realignment) and are indirect measures of (typically linear) motion effects. Consequently, they do not model the full dynamics of motion-induced signal variations, although we expect RPs to relate to at least some of the variance induced by motion. Therefore, RPs should correlate at least to some extent with the time-series of ICs that represent motion artifact. In contrast, RPs are not expected to correlate with time-series of signal of interest. We defined a 72RP model, including the standard 24RP model (6 standard RPs, their derivatives, and the quadratic terms of these 12 RPs), as well as a single time-point backward and a
forward shifted version of the 24RP model to reflect possible non-linear effects. Next, we calculated the maximum absolute correlation of each IC's time-course with each of the 72 RPs (squared RPs were correlated with squared IC time-series). We used a robust correlation, by calculating the mean correlation over 1000 random selections of 90% of the points in a time-series.

Spatial features

In addition to temporal characteristics we assessed spatial characteristics of each IC. To increase accuracy of the spatial features, IC spatial maps are first thresholded using a Gamma/Gaussian alternative testing approach (p > 0.5). This approach is automated in FSL's MELODIC. Additional standardization is established by aligning the IC spatial maps to the MNI152 template (2 mm isotropic resolution). Registration was performed by first co-registering the IC spatial maps to the participant's structural image using an affine boundary-based registration (FSL-FLIRT; Jenkinson and Smith, (2001); Jenkinson et al., (2002); Greve and Fischl, (2009)). Subsequently, IC maps were transformed to MNI152-space using the structural to MNI152-space non-linear transformation (FSL-FNIRT; Andersson et al., (2007)). We transformed the IC spatial maps to MNI152 2 mm standard space to ensure the IC classification to be independent of the initial voxel-size at acquisition and allowing the use of standardized masks to derive the spatial features described below. Note that the fMRI data itself is not transformed, i.e. the data denoising after IC classification is performed in native space.

Edge fraction. Head movements will induce strong variations in voxels that are located near intensity edges of the brain, as head motion will shift the location of the brain relative to the voxel location (i.e. voxels do not represent identical brain regions over time). Accordingly, we assessed each IC's representation near the edge of the brain. To this end we defined an edge mask by subtracting an eroded whole-brain mask (in MNI152 2 mm space, eroded using a 10 mm box kernel) from the full MNI152 2 mm mask, hence retaining the edge of the brain. Prior to the erosion we subtracted a CSF mask from the whole-brain mask such that the edges around the CSF would be included in the edge mask. Next, we calculated each IC's edge fraction as the sum of absolute Z-values of voxels overlapping the edge mask or located outside the whole-brain mask, divided by the sum of absolute Z-values of all voxels.

CSF fraction. Similar to the edge fraction this feature is based on the observation that ICs of interest are represented specifically within gray matter while motion effects are most prominently located at intensity edges within the brain (e.g. ventricle borders). To this end we defined a CSF mask by thresholding a CSF segmentation prior, supplied as part of FSL, at 95% of the robust range. We defined
ICA-AROMA method

each IC’s CSF fraction as the sum of absolute Z-values of voxels overlapping the CSF mask, divided by the sum of absolute Z-values of all voxels.

ICA-AROMA training

Next we discuss the extensive training of the ICA-AROMA classifier. We describe the dataset used for training, the preprocessing performed, and the labeling process required for classifier training and evaluation.

Training data

We included participants from the NeuroIMAGE project, a large project aimed at studying Attention-Deficit/Hyperactivity Disorder (ADHD) (von Rhein et al., 2015). To implement ICA-AROMA we used data from 30 healthy controls and included their anatomical and rfMRI scans. We will refer to this data as the training set (See Table 1 for participant characteristics).

MRI acquisition

Data were acquired at two scanning locations on similar 1.5 Tesla Siemens scanners (Siemens Sonata at VU University Medical Centre in Amsterdam; Siemens Avanto at Donders Centre for Cognitive Neuroimaging in Nijmegen) using the same Siemens 8-channel head coil and identical scanning protocols. Anatomical images were obtained using an MPRAGE sequence (TR = 2730 ms, TE = 2.95 ms, T1 = 1000 ms, flip angle = 7, matrix size = 256 × 256, FOV = 256 mm, 176 slices with 1 mm isotropic voxels). Functional images during rest were obtained using a gradient echo echo-planar imaging (GE-EPI) sequence (TR = 1960 ms, TE = 40 ms, FOV = 224 mm, 37 axial slices, flip angle = 80, matrix size = 64 × 64, in-plane resolution = 3.5 mm, slice thickness/gap = 3.0 mm/0.5 mm). Participants were instructed to relax with their eyes open during the rtMRI scan.

Table 1 - Characteristics of the participants included in the initial training and validation of ICA-AROMA.

<table>
<thead>
<tr>
<th>N</th>
<th>Age (years; mean ± SD)</th>
<th>Male (%)</th>
<th>Scan location 1 (%)</th>
<th>RMS-FD (mm; mean ± SD)</th>
<th>RMS-FD (mm; maximum)</th>
</tr>
</thead>
<tbody>
<tr>
<td>rfMRI (training)</td>
<td>30</td>
<td>16.6 ± 3.6</td>
<td>40</td>
<td>33</td>
<td>0.13 ± 0.10</td>
</tr>
<tr>
<td>rfMRI (validation)</td>
<td>100</td>
<td>16.9 ± 2.9</td>
<td>50</td>
<td>39</td>
<td>0.12 ± 0.09</td>
</tr>
<tr>
<td>Stop-signal task fMRI</td>
<td>118</td>
<td>16.9 ± 3.2</td>
<td>45</td>
<td>47</td>
<td>0.07 ± 0.04</td>
</tr>
</tbody>
</table>
fMRI data preprocessing

Preprocessing of the training set was carried out using tools from the FMRIB Software Library (FSL; http://www.fmrib.ox.ac.uk/fsl; (Smith et al., 2004; Woolrich et al., 2009; Jenkinson et al., 2012)) and involved (1) removal of the first five volumes to allow for signal equilibration, (2) head movement correction by volume-realignment to the middle volume using MCFLIRT, (3) global 4D mean intensity normalization and (4) spatial smoothing (6 mm FWHM). Importantly, no temporal filtering was applied at this stage of processing. Next, we applied ICA to the preprocessed participant-level data, using automatic dimensionality estimation as implemented in FSL MELODIC. To quantify motion in the fMRI data we used frame-wise displacement (FD) time-series as calculated by MCFLIRT (Jenkinson et al., 2002). We used the root mean squared of the FD time-series as a participant-level summary motion score (RMS-FD).

Manual labeling of independent components

The final feature set and classifier of ICA-AROMA as presented above were based on feature and classifier testing. To this end, one author (RP) manually labeled all ICs resulting from the participant-level ICA as motion, resting-state network (RSN) or ‘other’. The spatial maps, time-courses and power spectra of every component were inspected. Specifically, the spatial maps were inspected for the presented activation pattern; activation at intensity edges of the brain was regarded as motion, whereas ‘clustered’ activation with limited spurious characteristics and correspondence with well-known RSNs were regarded as RSN. A spiking pattern within the time-course or slow drifts with potentially abrupt changes that correspond with the six rigid body parameters were regarded as motion whereas time-courses dominated by low-frequency fluctuations were regarded as RSN. Since manual labeling is a subjective process only clear cases of motion or RSN ICs were classified as such. ICs where any doubt existed were classified as ‘other’.

Table 2 provides an overview of the manual labeling results averaged across participants for 1036 ICs detected across the 30 rfMRI datasets. In total, we labeled 36% of all ICs as motion and 23% as RSN. Across participants we obtained an average 34.5 ± 8.2 ICs. Participant motion (RMS-FD) was highly correlated with the total number of ICs (p < 1e-7), and the number of ICs identified as motion (p < 1e-9) or ‘other’ (p = 0.02). Participant motion did not correlate with the number of ICs identified as RSN (p = 0.76).
ICA-AROMA method

Table 2 - Results of the manual labeling of ICs resulting from participant-level ICA on 30 rfMRI datasets (training data). We show the number of ICs identified as motion, RSN or ‘other’ (mean ± SD across participants), as well as the correlation of the number of ICs with participant motion (RMS-FD). The right part of the table illustrates the same results but after removal of motion-related ICs using ICA-AROMA (see Section 'Classifier evaluation').

<table>
<thead>
<tr>
<th>IC-label</th>
<th>Initial ICA results</th>
<th>After ICA-AROMA</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean ± SD number ICs</td>
<td>Mean ± SD number ICs</td>
</tr>
<tr>
<td>Motion</td>
<td>12.4 ± 5.7</td>
<td>0.87 (&lt; 1e-9)</td>
</tr>
<tr>
<td>RSN</td>
<td>8.0 ± 2.7</td>
<td>−0.06 (0.76)</td>
</tr>
<tr>
<td>Other</td>
<td>14.2 ± 4.7</td>
<td>0.43 (0.02)</td>
</tr>
<tr>
<td>Total</td>
<td>34.5 ± 8.2</td>
<td>0.83 (&lt; 1e-7)</td>
</tr>
</tbody>
</table>

Automatic classification of independent components using ICA-AROMA. Figure 2 illustrates the manually labeled ICs mapped onto the four selected features, and the decision boundaries defined for automatic classification. The automatic ICA-AROMA classifier initially identifies motion-related ICs using a decision boundary combining the edge fraction and maximum RP correlation features. Given our focus on removing motion-related components, the decision boundary was determined by training a LDA classifier on ICs labeled as motion or RSN. As is evident from Figure 2, combining the edge fraction and maximum RP correlation already provided high discrimination between motion-related ICs and RSNs. The HFC and CSF fraction features did not provide added value to identify motion-related ICs within this multi-dimensional space. However, as hypothesized, RSNs clearly exhibited decreased high-frequency content, suggesting that components exhibiting increased high-frequency content could be classified as noise.

Figure 2 - The figure presents the feature scores of 1036 components which were manually labeled as: resting-state network (RSN), motion or ‘other’. The criteria of the classifier are presented by dashed lines in each graph. The first criterion (left panel) comprises a decision boundary combining the edge fraction and maximum correlation with RPs, determined by a linear discriminant analysis (LDA) on components manually labeled as RSN or motion. The two additional criteria concern a high-frequency content threshold of 0.35 (middle panel), and a CSF fraction threshold of 0.1 (right panel).
Similarly, RSNs exhibited very low overlap with CSF as indicated by the low CSF fraction scores. Accordingly, we defined conservative thresholds for both features (0.35 for HFC; 0.10 for CSF fraction), resulting in additional classification of suspected motion-related or general noise components while clearly preserving RSNs.

**Tuning feature implementation and testing alternative features**

We adopted an approach where we trained a classifier on the basis of an a priori set of features to maximize robustness while minimizing the bias of the classifier towards the training dataset. However, in addition to the selected features (HFC, maximum RP correlation, edge fraction, and CSF fraction), we investigated the potential added value of various additional features (see Supplementary materials). Before investigating additional features we first assessed variants of the edge fraction and maximum RP correlation features to fine-tune the implementation of our originally defined features (e.g. amount of erosion applied to edge mask). Next, we investigated the potential of alternative features that characterize different spatial or temporal aspects of the components and have been used in previous literature (e.g. spatial smoothness, kurtosis). Our results indicate that some alternative features did capture additional characteristics that could potentially be beneficial in the classification procedure. However, benefits were minimal and did not outweigh the added analytical complexity and risk of overfitting the classifier towards our training dataset.

**Results**

**Classifier evaluation**

To evaluate the ICA-AROMA classifier we assessed its classification accuracy in the training dataset and the robustness of the classifier by means of a leave-N-out cross-validation. When applying the ICA-AROMA classifier on the full set of components of the training set, including all 30 participants, ICA-AROMA removed a mean of 23.1 out of 34.5 components across participants (see Table 2). Figure 3 presents the classification results specifically for each manually defined class; RSN, motion and ‘other’. Of the components that were removed only 1.1% was manually labeled as RSN (3% of all RSNs, i.e. specificity of 97%). ICA-AROMA reduced the number of motion ICs from a mean of 12.4 per participant to 0.9 (i.e. sensitivity of 93%). In addition, the correlation between the number of motion ICs and the participant level motion summary score (RMS-FD) strongly decreased after removal of ICs classified as motion by ICA-AROMA and was no longer significant (p = 0.10). ICA-AROMA further removed a mean of 11.4 components (81%) labeled as ‘other’ (see
Supplementary Figure 2 for representative examples of ‘other’ components which were respectively removed or not removed by ICA-AROMA). Yet, participants exhibiting more motion still had more components labeled as ‘other’. In contrast, participant-level motion was no longer associated with the total number of components after denoising.

In addition, we evaluated the robustness of the ICA-AROMA classifier, using a leave-N-out cross-validation over 500 random splits of the training dataset. For each split we re-determined the edge fraction-maximum RP correlation boundary using 20 randomly selected participants and tested its accuracy in the remaining 10 participants. The high-frequency content and CSF fraction thresholds were kept constant, as they were intended to be conservative. Accuracy was defined as the sensitivity and specificity in identifying motion ICs from the total set of RSN and motion ICs. The cross-validation yielded a sensitivity and specificity of respectively 92 ± 2% and 97 ± 3%. In addition, 80 ± 3% of the ‘other’ components met the classification criteria.

As the rfMRI data of the training set were acquired at two different scan locations (Nijmegen and Amsterdam) we also validated ICA-AROMA with respect to scanner site. To this end, we determined the edge fraction/max RP correlation boundary using the participants that were scanned in Amsterdam (n = 20) and tested it on the participants scanned in Nijmegen (n = 10), and vice versa. This respectively yielded a specificity of 92% and 100% and a sensitivity of 95% and 90%. However, note that all leave-N-out cross-validation analyses included smaller
training samples than the total sample of 30 participants used for training the final ICA-AROMA classifier.

**Validation**

We validated ICA-AROMA by applying it to rfMRI and stop-signal task fMRI data (SST-fMRI). We compared results obtained with ICA-AROMA to two commonly applied motion correction strategies: extensive nuisance regression including 24 RPs, and spike regression. All three preprocessing strategies were complemented by standard nuisance regression including time-series for WM and CSF signal as well as a linear trend to remove residual (non-motion related) structured noise. We investigated two aspects in both the rfMRI and SST-fMRI data: statistical maps from group-level analyses and the loss in tDoF associated with secondary motion artifact removal. By assessing group-level spatial maps, we aimed to investigate the removal of spurious noise, and potentially increased sensitivity to activation. The loss in tDoF on the other hand, illustrates the cost of a strategy by potentially introducing heteroscedasticity in group-level analyses. Materials and methods used for this validation are detailed in the Supplementary materials.

**Resting-state fMRI**

We employed group-ICA on 100 rfMRI datasets (see Table 1 for participant characteristics), and derived participant-level time-courses and frequency spectra of the resulting spatial components (Beckmann et al., 2009; Filippini et al., 2009). The automatic dimensionality estimation in MELODIC group-ICA estimated 22 components for 24RP regression, 18 components for spike regression, and 11 components for ICA-AROMA (spatial maps are presented in Supplementary Figures 3 and 4). Amongst these maps we manually identified respectively 10, 8 and 11 of these components as RSNs, on the basis of their spatial configuration and concordance with canonical RSNs described previously in literature (Beckmann et al., 2005; Damoiseaux et al., 2006; Calhoun et al., 2008; Smith et al., 2009; Biswal et al., 2010; Rosazza and Minati, 2011).

All 11 ICs resulting from the ICA-AROMA preprocessed data represented RSNs. In contrast, group-ICA after 24RP or spike regression yielded both RSNs and ICs representing structured noise. Moreover, the RSN maps derived from ICA-AROMA preprocessed data contained decreased levels of spurious noise compared to equivalent RSN maps obtained after 24RP or spike regression. As an example, Figure 4 presents the spatial maps and frequency spectra of an IC representing the sensorimotor network. Compared to ICA-AROMA, the sensorimotor components
ICA-AROMA method

obtained after 24RP or spike regression included increased levels of spurious noise. This was reflected in the spatial as well as the frequency domain, by intensities at the brain’s edges and increased high-frequency power.

Table 3 illustrates the relationship between participant-level summary scores of RSN and motion time-courses, respectively referred to as RMS-RSN and RMS-FD (see Supplementary materials for details). After processing using ICA-AROMA, we observed no significant relationships between the RSN and motion summary scores. In contrast, after processing using 24RP or spike regression respectively 9 out of 10 and 5 out of 8 RSNs were found to be significantly correlated with the motion summary score (RMS-FD).

Figure 5 illustrates the loss in tDoF associated with each strategy. Naturally, 24RP regression resulted in a constant loss of 27 tDoF (24RP + WM + CSF + linear
trend regressors) for each participant. ICA-AROMA resulted in a similar loss in tDoF, removing on average 26 ± 7 tDoF. In contrast, spike regression resulted in a high and highly variable loss in tDoF of 55 ± 37. In the rfMRI data, spike regression resulted in removal of more than 50 tDoF in half of the included participants, with some losing close to 160 tDoF (out of 260 available tDoF). Of note, participants were excluded from our analysis when the data contained less than 125 volumes after spike regression (i.e. lost tDoF of 162).

### Task-based fMRI

We assessed fMRI data obtained during performance of a Stop Signal task (SST; 462 runs over 118 participants; see Table 1 for participant characteristics), intended to measure response inhibition (Logan et al., 1984). Figure 6 illustrates that the group-level activation maps associated with correct stop-trials derived for each preprocessing strategy were highly consistent (threshold-free cluster enhanced, FWE-corrected, p < 0.05). Compared to 24RP or spike regression, ICA-AROMA reduced activation around brain edges while resulting in increased sensitivity to (de)activation in regions associated with the default mode network. 24RP regression and spike regression yielded very similar activation maps besides a shift in mean. This mean shift results from the fact that by regressing out variance of specific volumes without fully deleting the volume from the fMRI time series, spike regression effectively replaces these volumes by the mean volume over time, therefore leading to an underestimation of the parameter estimates. The results as

<table>
<thead>
<tr>
<th>RSN</th>
<th>24RP regression</th>
<th>Spike regression</th>
<th>ICA-AROMA</th>
</tr>
</thead>
<tbody>
<tr>
<td>Visual-medial</td>
<td>0.16 (0.10)1</td>
<td>-0.29 (&lt; 0.01)</td>
<td>-0.04 (0.67)</td>
</tr>
<tr>
<td>Visual-occipital-1</td>
<td>-0.27 (&lt; 0.01)</td>
<td>-0.07 (0.48)</td>
<td></td>
</tr>
<tr>
<td>Visual-occipital-2</td>
<td>0.56 (&lt;0.01)2</td>
<td>x</td>
<td>-0.06 (0.57)</td>
</tr>
<tr>
<td>Visual-lateral</td>
<td>x</td>
<td>-0.04 (0.68)</td>
<td></td>
</tr>
<tr>
<td>DMN-posterior</td>
<td>0.29 (&lt; 0.01)</td>
<td>-0.29 (&lt; 0.01)</td>
<td>0.13 (0.20)</td>
</tr>
<tr>
<td>DMN-frontal</td>
<td>0.46 (&lt; 0.01)</td>
<td>-0.32 (&lt; 0.01)</td>
<td>0.09 (0.40)</td>
</tr>
<tr>
<td>Executive control</td>
<td>0.54 (&lt; 0.01)</td>
<td>0.12 (0.25)3</td>
<td>0.08 (0.46)</td>
</tr>
<tr>
<td>Auditory</td>
<td>0.49 (&lt; 0.01)</td>
<td></td>
<td>0.07 (0.52)</td>
</tr>
<tr>
<td>Sensorimotor-1</td>
<td>0.70 (&lt; 0.01)</td>
<td>0.05 (0.62)4</td>
<td>0.12 (0.24)4</td>
</tr>
<tr>
<td>Sensorimotor-2</td>
<td>0.65 (&lt; 0.01)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Left fronto-parietal</td>
<td>0.43 (&lt; 0.01)</td>
<td>-0.16 (0.12)</td>
<td>-0.05 (0.63)</td>
</tr>
<tr>
<td>Right fronto-parietal</td>
<td>0.55 (&lt; 0.01)</td>
<td>-0.26 (&lt; 0.01)</td>
<td>-0.17 (0.09)</td>
</tr>
</tbody>
</table>

x component was not identified
1 merged visual-medial and visual-occipital-1 components.
2 merged visual-occipital-2 and visual-lateral components.
3 merged executive control and auditory components
4 merged sensorimotor 1 and 2 components.
presented in Figure 6 replicated across all six task contrasts (Supplementary Figures 5 and 6).

In addition, as for the rfMRI data, Figure 5 illustrates the loss in tDoF in the SST-fMRI data (over 462 runs) associated with each strategy. While 24RP regression and ICA-AROMA resulted in a limited, stable loss in tDoF of respectively 27 and 23 ± 3, spike regression resulted in a loss of 32 ± 9 tDoF per run, out of a total available tDoF of 85 ± 3.6 (i.e. number of volumes).

**Discussion**

We proposed ICA-AROMA, a novel strategy for removing motion artifacts from fMRI data. ICA-AROMA exploits independent component analysis to identify participant-specific motion-related components in a data-driven fashion. These components are subsequently removed from the data. We showed that ICA-AROMA effectively removed motion artifacts from both rfMRI and task-based fMRI data, while increasing sensitivity to signal of interest. By avoiding the removal of fMRI volumes, ICA-AROMA largely preserved temporal degrees of freedom and retained the autocorrelation structure of the fMRI data. A pragmatic advantage is that ICA-AROMA can be readily and robustly applied across datasets without requiring re-training, determination of thresholds or manual labeling of components.
The effects of secondary motion artifacts on fMRI data are an active topic of debate within the neuroimaging community. Proposed strategies for removal of these artifacts typically exploit realignment parameters, derived from an initial volume-realignment procedure, for extensive nuisance regression (Friston et al., 1996; Satterthwaite et al., 2013a) and/or identification and removal of specific high-motion volumes (Power et al., 2012, 2014; Satterthwaite et al., 2013a). As an alternative, ICA-based strategies for fMRI data denoising aim for automatic classification and removal of structured noise components derived by ICA (Thomas et al., 2002; Kochiyama et al., 2005; De Martino et al., 2007; Perlbarg et al., 2007; Tohka et al., 2008; Kundu et al., 2012; Bhaganagarapu et al., 2013; Rummel et al., 2013; Storti et al., 2013; Salimi-Khorshidi et al., 2014). Generalization of such classifiers over multiple datasets, acquired using different MR scanner and/or different MRI protocols, is however not straightforward and typically requires re-training to ensure accurate classification. This will specifically impact multi-center studies, including data-sharing initiatives, where a different classifier will have to be trained for every imaging site or imaging sequence. As a result, the classifier and its associated accuracy will vary across datasets and might introduce bias when analyzing pooled data. Moreover, separate training to accommodate each dataset will result in data loss, as the participants used for training have to be excluded from further analysis.

![Figure 6 - Group-level activation maps (upper row) obtained from SST-fMRI data (118 participants), processed using three different motion artifact removal strategies: 24RP regression, spike regression and ICA-AROMA. The bottom row illustrates the significant change in sensitivity to activation between every pair of strategies. These results were obtained by first deriving participant-level difference maps between absolute parameter estimate maps obtained for two motion artifact removal strategies. Secondly, analogous to the activation maps, we tested these difference-maps at the group-level. By determining the difference between the absolute values, rather than the raw values, these results illustrate differences in effect sizes, i.e. sensitivity to activation. A significant decrease in effect size is presented in blue and a significant increase in yellow. All group-level activation and sensitivity maps show t-statistic maps thresholded at p < 0.05 (threshold-free cluster enhanced, FWE-corrected), associated with correct stop-trials.](image-url)
To mitigate concerns associated with realignment parameter- or current ICA-based strategies we designed and validated ICA-AROMA with a focus on removing motion artifacts. By focusing on one type of structured noise and designing a robust classifier we aimed for a strategy that generalizes to new datasets. It differs from currently available specific ICA-based strategies by defining limited set of features that are more standardized, span both the spatial and temporal domain, and are independent from experimental design. Hence, it is applicable to both rfMRI and task-based fMRI, and allows for discriminating between signal and motion ICs that share characteristics in the spatial or temporal domain (e.g. a signal component presenting activation near the brain-edge but temporally uncorrelated with RPs). We showed that ICA-AROMA accurately identified motion artifacts while retaining signal of interest at the participant-level. Moreover, ICA-AROMA yielded comparable results on task-based fMRI and rfMRI data without re-training the classifier.

Classification accuracy of ICA-based strategies, as also presented in this work, is generally reported by comparing automatic classification to manual labeling (De Martino et al., 2007; Bhaganagarapu et al., 2013; Rummel et al., 2013; Salimi-Khorshidi et al., 2014). However, as manual labeling has to be performed by visual inspection (Kelly et al., 2010) in the absence of ground-truth on what various components represent, it is possible that raters mislabel components. As an example, an IC that is manually labeled as general noise might in fact be related to motion in a less typical manner (e.g. due to non-linear motion effects).

We indeed found significant inter-individual correlations between the motion score and the number of components labeled as ‘other’. Suggesting that in participants that exhibited increased head motion there was an increased number of components that did not clearly represent RSN or motion artifact. Accordingly, we want to emphasize that the performance of classification strategies should not be solely judged based on their consistency with manually labeled components.

Next to nominal classification accuracy, we therefore validated ICA-AROMA by conducting group-level analyses and specifically investigated the presence of signal of interest (i.e. specificity) and motion-related noise (i.e. sensitivity). The group-ICA analysis on the data preprocessed using ICA-AROMA reflected the accuracy and robustness of ICA-AROMA by only yielding ICs representing RSNs, which were unrelated to motion in both the spatial and temporal domain. In contrast, preprocessing the data by conducting 24RP or spike regression resulted in ICs representing structured noise and multiple RSNs that were spatially and temporally related to motion. In addition, ICA-AROMA reduced motion-related
noise in the SST-fMRI data to larger extent than 24RP and spike regression, as investigated by group-level activation analysis. Moreover, ICA-AROMA increased sensitivity to activation in the SST-fMRI data. We argue that such an evaluation method is more informative and objective than solely discussing the subjective classification accuracies of ICA-based strategies.

Another reason why the classification accuracy of an ICA-based motion artifact removal strategy is not necessarily representative of its denoising capability is related to the method implemented for removal of the identified ICs from the fMRI data. ICs can be fully regressed out of the data, analogous to RP-regression (Friston et al., 1996; Satterthwaite et al., 2013a). In this case, all variance associated with these noise ICs will be removed, including shared-variance with ‘good’ ICs. Therefore, such a component regression method can be regarded as ‘aggressive denoising’. Alternatively, one can employ a linear regression on the full mixing matrix as estimated by ICA, containing both signal and noise ICs (see Section ‘Step 3 - data denoising’). This allows specifically removing the variance assigned to the identified noise ICs, yielding ‘non-aggressive denoising’. A drawback of this type of denoising is that motion-related components that were not classified as such will be specifically retained within the data. Aggressive denoising, on the other hand, might partly remove such components due to shared variance with correctly classified noise components. We chose to implement the non-aggressive approach within ICA-AROMA as this represents a more conservative approach and prevents the loss of signal of interest due to overfitting of the noise regressors, specifically when the data contains stimulus-related noise. For more radical removal of noise, Griffanti et al., (2014) proposed to incorporate 24 RPs within a non-aggressive denoising procedure in ICA-FIX (Salimi-Khorshidi et al., 2014). The authors proposed to employ ICA on the original data and classify the ICs, subsequently regress out the 24 RPs from both the IC time-series and original data, and as a last step apply non-aggressive denoising for removal of the components classified as noise. However, this strategy results in an additional loss of tDoF and potential loss in sensitivity to signal of interest.

As a result of preserving shared variance between signal and noise components, the actual lost tDoF when using ICA-AROMA will be less than presented in the current research, as we counted one tDoF for each IC that was removed. In contrast to low loss of tDoF when employing ICA-AROMA, spike regression resulted in a major and highly variable loss in tDoF across participants (see Figure 5). Such vast differences in lost tDoF at the participant level will have down-stream consequences. Decreased tDoF impacts detection power and reliability of subject-level functional connectivity estimates, and will have additional consequences for group-level analyses due to between-subject tDoF-variability (Birn
ICA-AROMA method

et al., 2013; Yan et al., 2013a, 2013b). Group-level analyses typically run under the assumption that the tested variables (e.g. participant-level functional connectivity estimates) include equal error variance. This error variance is partly determined by the available number of tDoF. By introducing high tDoF-variability between participants, spike regression and scrubbing induce group-level variability in error variance and therefore violate typical statistical inference at the group-level (e.g. parametric statistics as a two-sample T-test, ANOVA, or ordinary least squares regression). However, residual motion artifact also impacts error variance of participant level estimates. Consequently, it is clear that spike regression and scrubbing effectively substitute variability due to motion artifacts by heteroscedasticity due to the lost tDoF. Of note, deleting volumes in the scrubbing strategy (Power et al., 2012) would yield identical results.

Next to a variable loss in tDoF, volume-removal will decrease reliability of fMRI metrics (Birn et al., 2013; Zuo et al., 2013) and destroy the autocorrelation structure of the fMRI data. The latter prevents analysis assessing frequency characteristics or non-stationarity of the data, and impacts temporal filtering. To reduce the impact on temporal filtering, one can conduct volume-removal after the filtering procedure. However, as discussed by(Carp, 2013), temporal filtering of time-series that contain sharp transitions (i.e. motion-induced spikes) induce ringing artifacts. Therefore, motion artifacts are preferably removed prior to temporal filtering. To accommodate this, we designed ICA-AROMA to be applied prior to temporal filtering. Moreover, applying secondary motion artifact removal prior to temporal filtering not only prevents ringing artifacts but additionally prevents removal of low-frequency motion-related signal variance which can aid in identification of motion artifacts. Similarly, as signal from WM and CSF time-series might also contain variance related to motion artifacts, we applied ICA-AROMA prior to nuisance regression. Additionally, we applied ICA-AROMA after spatial smoothing. Spatial smoothing will blur motion-related artifacts to neighboring voxels. Yet, as spatial smoothing will increase signal to noise ratio by reducing speckled noise, it will increase the ability of ICA to detect coherent signal patterns, including better separation of structured artifacts. Note that we did not include slice-timing correction in our preprocessing pipeline. Although slice-timing correction might interact with motion-induced artifacts, it is not expected to substantially alter the characteristics of the motion components and therefore impact the ICA-AROMA classification accuracy (i.e. the components will still be prominently reflected at the edges of the brain and correlate with RPs).

ICA-AROMA classified a high fraction of components manually labeled as 'other' as motion-related. In addition, we observed no components representing structured noise in the group-ICA results. These observations indicate that the ICA-
AROMA criteria resulted in the classification of other types of structured noise as motion-related, effectively resulting in their removal (e.g. cardiac pulsation artifacts might exceed the HFC threshold due to the presence of high-frequency noise). Yet, it should be noted that ICA-AROMA is specifically designed and validated with respect to removal of motion-related artifacts. Accordingly, we advise to check your data for any remaining structured artifacts and remove residual structured noise after application of ICA-AROMA. We implemented the removal of residual noise by means of WM and CSF regression. Note that to prevent the risk of removing signal of interest due to spatial smoothing or inaccurate WM and CSF masks, a conservative approach is advisable. Here, we extracted mean WM and CSF signals using masks obtained by multiplying a participant-specific tissue prior with an MNI152-derived tissue prior, both thresholded at a very conservative threshold of 95% tissue probability (see Supplementary materials).

In the current manuscript we validated ICA-AROMA and discussed its fundamentally different approach compared to alternative strategies. However, an extensive evaluation of the performance of ICA-AROMA in the context of available denoising strategies is required. Such a comprehensive assessment is beyond the scope of the current manuscript, but is provided in a companion manuscript (Pruim et al., 2015a). In that manuscript we compared ICA-AROMA to alternative strategies including: extensive nuisance regression with 6 or 24 motion parameters (Friston et al., 1996; Satterthwaite et al., 2013a), scrubbing (Power et al., 2012), spike regression (Satterthwaite et al., 2013a), aCompCor (Behzadi et al., 2007), ICA-FIX without re-training (Salimi-Khorshidi et al., 2014) and SOCK (an alternative ICA-based strategy; Bhaganagarapu et al., 2013)). We assessed the quality of motion artifact removal of each strategy. Importantly we also addressed preservation of signal of interest, an aspect which is often overlooked when evaluating motion artifact removal strategies and is of special interest in light of recent findings that relate head motion to traits or neural activity (Van Dijk et al., 2012; Couvy-Duchesne et al., 2014; Kong et al., 2014; Pujol et al., 2014; Zeng et al., 2014). We investigated multiple functional connectivity estimates derived by dual regression, seed-based regression, and seed-based correlation matrices. The results were additionally replicated across four different resting-state datasets, including one clinical dataset. In summary, we showed that spike regression, scrubbing, ICA-FIX and ICA-AROMA minimized the impact of head motion on functional connectivity estimates. ICA-FIX and ICA-AROMA however limited the loss of tDoF and exhibited increased reproducibility of spatial maps representing RSNs, indicating that these strategies achieve more efficient noise removal. However, without re-training, ICA-FIX poorly generalized across datasets. In contrast, ICA-AROMA retained signal of interest and resulted in highly consistent results on all
investigated domains, endorsing its generalizability and robustness as put forward in the current manuscript.

**Conclusion**

We have developed a robust and automated strategy to identify and remove motion-related artifact from fMRI data. Our strategy employs independent component analysis, and a simple, four-feature classifier, that does not require re-training in new datasets. ICA-AROMA is applicable to both rfMRI and task-based fMRI data, effectively removing (motion-related) spurious noise while increasing sensitivity to activation. An initial validation showed that ICA-AROMA outperforms extensive nuisance regression and spike regression in terms of noise removal. Importantly, ICA-AROMA does not require deleting or regressing out time-points to remove motion-induced signal variations. ICA-AROMA is publicly available as an easily applicable single command (http://fsl.fmrib.ox.ac.uk/fsl/fslwiki/OtherSoftware).

**Acknowledgments**

The NeuroIMAGE project was supported by NWO Large Investment Grant 1750102007010, ZonMW Grant 60-60600-97-193, and NOW Brain and Cognition grants 433-09-242 and 056-13-015 to Dr. Buitelaar, the EU FP7 grant TACTICS (grant no. 278948), and grants from Radboud University Medical Center, University Medical Center Groningen and Accare, and VU University Amsterdam. Dr. Mennes has received funding from the European Research Council under the European Union’s Seventh Framework Programme (FP7/2007-2013)/ERC grant agreement 327340. Dr. Beckmann is supported by the Netherlands Organisation for Scientific Research (NWO-Vidi 864-12-003) and we gratefully acknowledge the funding from the Wellcome Trust UK Strategic Award [098369/Z/12/Z].

**Conflict of interest**

The authors declare that they have no conflicts of interest.
References


Fair, D.A., Nigg, J.T., Iyer, S., Bathula, D., Mills, K.L., Dosenbach, N.U.F., Schlaggar,
ICA-AROMA method


Supplementary materials

Tuning feature implementation and testing alternative features

In addition to the selected features presented in the manuscript (HFC, maximum RP correlation, edge fraction, and CSF fraction), we investigated various alternative features. We first assessed variants of the edge fraction and maximum RP correlation features to fine-tune the implementation of our originally defined features (e.g., amount of erosion applied to edge mask). Next, we investigated the potential added value of alternative features that characterize different spatial or temporal aspects of the components and have been used in previous literature.

Feature variants

To fine-tune the implementation of our initial features we evaluated several procedural variants of the edge fraction and maximum RP correlation. Assessing these variants resulted in the final procedures used to derive the features of the ICA-AROMA classifier.

We derived variants of the edge fraction feature by calculating the feature for four different types of spatial IC maps and by using four different edge masks, resulting in a total set of 16 edge fraction features. The spatial IC maps were either thresholded by mixture modeling or unthresholded, and either kept in native space or registered to standard space (MNI152 2mm). The edge masks were derived by eroding a brain mask with a box kernel of three or five times the voxel size, and including or excluding an edge around the ventricles. For the maximum RP correlation feature we derived eight variants by calculating the feature for alternative RP models: 6 standard RPs (6RP), 6 standard + 6 derivatives (12RP), 12RP + quadratic terms (24RP), 24RP + single frame forward-shifted 24RP (48fwRP), 24RP + single frame backward-shifted 24RP (48bwRP), 24RP + single frame forward- and backward-shifted 24RP (72RP) and the two models derived from the 72RPs but respectively only including the 36 parameters related to the standard or frame-wise RPs. We defined these models since different RPs relate to different types or aspects of motion artifacts. The standard 6 RPs, for instance, reflect the head position in the scanner and therefore relate to artifacts induced by magnetic field interaction and interpolation effects, frame-wise RPs relate to spin-history effects and temporally shifted RPs potentially relate to non-linear characteristics of the artifacts.

We investigated all 128 combinations between the 16 edge fraction and eight maximum RP correlation features with respect to their LDA classification accuracy;
percentage of correctly classified RSN and motion components. This yielded highly similar accuracy scores of 92±2%, ranging from 85% to 95%. In the end we selected those procedural variants that provided the best mix between ease of implementation, accuracy and theoretical motivation. The edge fraction and maximum RP correlation features as finally included in ICA-AROMA yielded an accuracy of 94%.

**Potential alternative features**

In addition to investigating variants on our initial features we assessed the potential added value of alternative features that characterize different spatial or temporal aspects of the components. The spatial features comprised: variance, kurtosis, skewness, entropy, smoothness of the spatial map and a WM+CSF fraction. We extracted the WM+CSF fraction from mixture modeled thresholded spatial maps (p>0.5) registered to standard space (MNI152 2mm isotropic), using a WM+CSF mask derived by thresholding the CSF and WM segmentation prior supplied as part of FSL at 95% of the robust range. All other spatial features were derived from the unthresholded spatial maps within native space. As an estimator for the spatial smoothness of the components we extracted the smoothness measure as implemented in the FSL ‘smoothest’ command; being the square root of the determinant of Lambda, in which Lambda is the covariance matrix of the partial derivatives of the Gaussian random field (Nichols, 2008). The investigated temporal features comprised: kurtosis, skewness, entropy, one-lag autocorrelation, two-lag autocorrelation and fractional ALFF (fALFF; Zou et al., (2008)) of the IC time-course.

We assessed the potential value of these features by first applying the original ICA-AROMA to the training set, and subsequently investigating how the remaining components (231 RSN, 27 motion and 83 ‘other’ components) map onto the different features (see Supplementary Figure 1). Indeed, the features reflect different characteristics between the motion, RSN and ‘other’ components. As an example, motion and ‘other’ components on average resulted in lower fALFF scores, lower skewness of the spatial maps, and higher smoothness scores compared to RSNs. However, as can be depicted from the figure, achieving a substantial increased identification of noise components while retaining RSN components is not straightforward and will require multiple features within a multi-dimensional classification approach. As ICA-AROMA already identified the major portion of noise components, implementation of additional features requires making a trade-off between minimally increasing classification accuracy and decreasing robustness/simplicity of the strategy. Accordingly, we regarded none of the additionally tested features to substantially add to our strategy as presented in the maintext.
Validation

To validate ICA-AROMA we applied the strategy torfMRI and stop-signal task fMRI data (SST-fMRI) and compared results to two commonly applied motion correction strategies: extensive nuisance regression including 24 realignment parameters (RPs), and spike regression. We investigated two aspects in both datasets: spatial maps from group-level analyses and the loss in tDoF associated with secondary motion artifact removal. By assessing group-level spatial maps, we aimed to investigate the removal of spurious noise, and potentially increased sensitivity to activation. The loss in tDoF on the other hand, illustrates the cost of a strategy by potentially introducing heteroscedasticity in group-level analyses.

Resting-state fMRI

We assessed rfMRI data from healthy controls from the NeuroIMAGE project. After exclusion for incomplete scanning, less than 125 non-spike labeled volumes (Power et al., 2012), 5% of participants exhibiting strongest motion based on the root mean squared of the frame-wise displacement (FD) measure proposed by Jenkinson et al., (2002)(RMS-FD), incidental neuroradiological findings and the participants used for training ICA-AROMA we included 100 healthy controls.

After initial preprocessing as described in the main manuscript we completed the preprocessing in three-fold by employing three different strategies for secondary motion artifact removal. In case of the extensive nuisance regression strategy (24RP regression) we extended standard nuisance regression (WM, CSF, linear trend) with 24 RPs: the 6 standard RPs derived from volume realignment (MCFLIRT), its derivatives and the quadratic terms of these 12 RPs. White matter and CSF time-series were derived by determining the mean time-series over voxels within a predefined WM and CSF mask. To obtain the masks, we applied FSL FAST to the
T1 structural image to derive a CSF and WM probability map. The maps were thresholded at 95% and subsequently registered to native EPI space. Likewise we registered and thresholded (95% of the robust range) the MNI152 average CSF and WM segmentation maps (priors). Multiplication of both masks resulted in the respective conservative CSF and WM masks. In case of spike regression we extended the 24RP regression by adding a single regressor for every frame defined as a spike (criterion of >0.2mm on the FD-measure proposed by Jenkinson et al. (2002)), as well as for one frame before and two frames after the spike frame (Power et al., 2012, 2014). Note that this strategy differs from the proposed preprocessing pipeline by Satterthwaite et al. (2013) which included global signal regression (and its the derivatives and squared terms), band-pass filtering, and boxcar instead of single spike regressors. Finally, in ICA-AROMA, ICA-AROMA was applied to the preprocessed data, followed by standard nuisance regression (WM, CSF, linear trend). In all strategies, data were high-pass filtered (>0.01Hz) after nuisance regression.

For every participant we then transformed the three differently preprocessed fMRI data to the participants’ structural image using a single pre-calculated affine boundary-based registration (FLIRT; Jenkinson and Smith, (2001); Jenkinson et al., (2002)). Subsequently, we registered the functional data to MNI152 standard space (4mm isotropic resolution) using non-linear registration as implemented in FSL FNIRT (Andersson et al., 2007).

To investigate residual noise versus group-level sensitivity we conducted a group-ICA on the preprocessed data of the three strategies (Beckmann et al., 2009). The group-ICA was implemented using MELODIC with automatic dimensionality estimation. We then investigated the resulting set of group-level ICs qualitatively and compared results between strategies. First, we identified which components represented resting-state networks (RSN) on the basis of their spatial configuration and concordance with canonical RSNs found previously in literature (Beckmann et al., 2005; Damoiseaux et al., 2006; Calhoun et al., 2008; Smith et al., 2009; Biswal et al., 2010; Rosazza et al., 2012). In addition, we assessed the relationship between the group-level components and subject-level motion. To this end, we obtained for each IC each participant’s IC time-course by using the set of group-level ICs for spatial regression against the preprocessed fMRI data. This resulted in IC time-course scores for every participant, for every IC and for each motion removal strategy. To obtain frequency spectra of all IC time-courses we employed a discrete Fourier transform. Next we calculated the root mean square for the IC time-courses that were manually identified as RSNs (RMS-RSN), correlated the score with the motion summary score RMS-FD across participants, and compared the correlation strengths between the different strategies.
To investigate lost temporal degrees of freedom (tDoF), we considered the number of available time-points to be the total number of tDoF, thereby ignoring temporal autocorrelation. Each strategy lost three tDoF due to the nuisance regression of signal from CSF, WM, and a linear trend. Extensive nuisance regression resulted in an additional loss of 24 tDoF. Spike regression resulted in ‘24 + number of spike regressors’ lost tDoF, and ICA-AROMA additionally lost ‘number of removed ICs’ tDoF. Lost tDoF as a result of high-pass filtering, implemented through a local Gaussian-weighted line-fitting procedure, were not accounted for since calculation of the associated lost tDoF is not straight-forward and does not differ between the three strategies.

Task-based fMRI

We assessed fMRI data obtained during performance of a Stop Signal task (multiple runs per participant), intended to measure response inhibition (Logan et al., 1984). These data were also part of the NeuroIMAGE project. Of note, for a detailed description of the task and ADHD-related findings we refer to van Rooij et al. (2015). After exclusion for insufficient accuracy, incomplete data, and 5% of the runs exhibiting strongest motion, we included 118 healthy controls (462 runs). The preprocessing pipeline was analogous to the preprocessing of the rFMRI validation dataset. The only differences being that we removed the first four instead of five volumes and applied slice timing correction (after volume realignment). After initial preprocessing we again applied the three secondary motion removal strategies as described above and concluded with high-pass temporal filtering (>0.01Hz).

After preprocessing we continued with typical first-level task-based analysis. Specifically, we constructed a General Linear Model (GLM) for each run for each participant to investigate task-related activation. To this end we modeled successful stop-trials, failed stop-trials, successful go-trials and failed go-trials. We derived six participant-specific activation maps (parameter estimate maps; PEs) by testing the following contrast: successful stop, failed stop, successful go, successful stop successful go, failed stop - successful go, and failed stop - successful stop. These maps were co-registered to the participants’ structural image using an affine boundary-based registration (FLIRT) and subsequently registered to standard space (MNI152, 4mm isotropic) using a non-linear registration algorithm (FNIRT). For every participant, the first-level PE-maps were combined across runs using a single-subject fixed effects model. To determine group-level activation we employed non-parametric permutation testing (5000 permutations, no covariates) on the participant-specific fixed-effect PE-maps. We tested a positive and negative mean group effect (activation and deactivation). The resulting maps were threshold-free cluster enhanced, family-wise error (FWE) corrected and thresholded at p<0.05. In
ICA-AROMA method

addition to these group-level activation maps we investigated sensitivity to activation more specifically. To that end we subtracted absolute participant-level PE-maps obtained by the three preprocessing strategies for each of the six contrast separately, and tested group-level effects analogous to the above procedure. We tested: |ICA-AROMA| – |24RP regression|, |ICA-AROMA| – |Spike regression| and |Spike regression| – |24RP regression|. This resulted in three group-level difference maps reflecting the respective decrease or increase in sensitivity towards task-related activation. Finally, analogous to the rfMRI analyses, we assessed the fractional loss in tDoF per run.

References
Power, J.D., Barnes, K. a., Snyder, A.Z., Schlaggar, B.L., Petersen, S.E., 2012. Spurious


Supplementary Figure 2 - Representative examples of components manually labeled as ‘other’ which either meet (left column) or did not meet (right column) the criteria of the ICA-AROMA classifier, i.e. were removed or not removed from the fMRI data during denoising. The components were derived from participant-level ICA and manually labeled as ‘other’ when the component did not uniquely represent a resting-state network or motion component.
Supplementary Figure 3 - Group-ICA components (Z-statistical spatial maps; thresholded using mixture modeling, p>0.5) manually identified as resting-state networks, obtained for each of the three different motion artifact removal strategies; 24RP regression, spike regression, and ICA-AROMA. See Supplementary Figure 4 for group components identified as structured noise. Resting-state networks included: A) Visual-medial, B & C) Visual-occipital, D) Visual-lateral, E) DMN-Posterior, F) DMN-Frontal, G) Executive control, H) Auditory, I & J) Sensorimotor, K) Left Fronto-Parietal, L) Right Fronto-Parietal.
ICA-AROMA method

Supplementary Figure 4 - Group-ICA components, manually identified as structured noise (Z-statistical spatial maps; thresholded using mixture modeling, p>0.5), obtained for the three different motion artifact removal strategies; 24RP regression, spike regression, and ICA-AROMA. Note, that no structured noise components were observed for ICA-AROMA. See Supplementary Figure 3 for the group components identified as resting-state networks.
Supplementary Figure 5 - Group-level activation maps for the SST-fMRI data, obtained for the three different motion artifact removal strategies: 24RP regression, spike regression, and ICA-AROMA. The figures illustrate t-statistical maps thresholded at p<0.05 (threshold-free cluster enhanced, FWE-corrected), for the following task-related contrasts: A) successful stop, B) failed stop, C) successful go, D) successful stop - successful go, E) failed stop - successful go, F) failed stop - successful stop.
Supplementary Figure 6 - Group-level sensitivity to task-related activity in the SST-fMRI data preprocessed using the three different motion artifact removal strategies; 24RP regression, spike regression, and ICA-AROMA. Each map depicts the difference between absolute parameter estimates obtained for two different motion artifact strategies. The figures illustrate t-statistical maps thresholded at $p<0.05$ (threshold-free cluster enhanced, FWE-corrected), for the following contrasts: A) successful stop, B) failed stop, C) successful go, D) successful stop - successful go, E) failed stop - successful go, F) failed stop - successful stop.
Chapter 2

Evaluation of ICA-AROMA and alternative strategies for motion artifact removal in resting state fMRI

Published as:
Chapter 2

Abstract

We proposed ICA-AROMA as a strategy for the removal of motion-related artifacts from fMRI data (Pruim et al., 2015). ICA-AROMA automatically identifies and subsequently removes data-driven derived components that represent motion-related artifacts. Here we present an extensive evaluation of ICA-AROMA by comparing our strategy to a range of alternative strategies for motion-related artifact removal: (i) no secondary motion correction, (ii) extensive nuisance regression utilizing 6 or (iii) 24 realignment parameters, (iv) spike regression (Satterthwaite et al., 2013a), (v) motion scrubbing (Power et al., 2012), (vi) aCompCor (Behzadi et al., 2007; Muschelli et al., 2014), (vii) SOCK (Bhaganagarapu et al., 2013), and (viii) ICA-FIX (Griffanti et al., 2014; Salimi-Khorshidi et al., 2014), without re-training the classifier. Using three different functional connectivity analysis approaches and four different multi-subject resting-state fMRI datasets, we assessed all strategies regarding their potential to remove motion artifacts, ability to preserve signal of interest, and induced loss in temporal degrees of freedom (tDoF). Results demonstrated that ICA-AROMA, spike regression, scrubbing, and ICA-FIX similarly minimized the impact of motion on functional connectivity metrics. However, both ICA-AROMA and ICA-FIX resulted in significantly improved resting-state network reproducibility and decreased loss in tDoF compared to spike regression and scrubbing. In comparison to ICA-FIX, ICA-AROMA yielded improved preservation of signal of interest across all datasets. These results demonstrate that ICA-AROMA is an effective strategy for removing motion-related artifacts from fMRI data. Our robust and generalizable strategy avoids the need for censoring fMRI data and reduces motion-induced signal variations in fMRI data, while preserving signal of interest and increasing the reproducibility of functional connectivity metrics. In addition, ICA-AROMA preserves the temporal non-artifactual time-series characteristics and limits the loss in tDoF, thereby increasing statistical power at both the subject- and the between-subject analysis level.
Introduction

Participant head motion during fMRI scanning can induce spurious temporal correlations between brain regions (Power et al., 2012; Van Dijk et al., 2012). Functional connectivity (FC) measures from resting state fMRI (rfMRI) data are especially vulnerable to the influence of such spurious correlations as signal of interest is related to the degree of temporal correlation between multiple voxel time series, and does not relate individual voxel time series to externally defined regressors. Motion-related fMRI artifacts can impact a wide range of FC measures (Satterthwaite et al., 2012; Yan et al., 2013) and can induce increases in local functional connectivity and decreases in long-distance functional connectivity (Power et al., 2012; Satterthwaite et al., 2012; Van Dijk et al., 2012). This effect is not due to motion per se, but an interaction between motion and the use of global signal regression (GSR) and equivalent regressors (Jo et al., 2013; Satterthwaite et al., 2013a). However, such spurious effects directly interfere with current hypotheses about local and long-distance functional connectivity in the context of neurodevelopment and Autism Spectrum Disorders (ASD; Courchesne and Pierce, (2005); Fair et al., (2007, 2008, 2009); Kelly et al., (2009); Power et al., (2010)). Structural between-subject differences in gross motion furthermore interfere with brain research into disorders which are associated with high motion (e.g. attention-deficit/hyperactivity disorder; ADHD). More-over, by reducing sensitivity towards effects of interest, the presence of motion-related artifacts might not only increase the probability of false-positive but also of false-negative results.

To minimize the impact of head motion in (r-)fMRI analyses, various strategies have been proposed that complement initial volume-realignment by addressing removal of residual motion-induced signal variance. The most commonly used strategy for removal of motion-related artifacts from fMRI data is extensive nuisance regression (Friston et al., 1996; Satterthwaite et al., 2013a; Yan et al., 2013). Here, a set of parameters derived by the volume-realignment algorithm is exploited to model and regress out signal variance associated with motion. More recently, it was proposed to specifically target motion-affected volumes, either by regressing them out (spike regression; (Lemieux et al., 2007; Satterthwaite et al., 2013)) or by fully removing them from the fMRI data (scrubbing; Power et al., 2012). While both spike regression and scrubbing remove motion-induced signal variations they have a significant impact on the temporal autocorrelation structure of fMRI data and can lead to a high and variable loss in temporal degrees of freedom (tDoF; Satterthwaite et al., 2013; Yan et al., 2013; Power et al., 2014; Pruim et al., 2015). Such variable loss in tDoF can induce between-group biases in cases where there is a between-group bias in gross head motion.
Alternative strategies for denoising fMRI data exploit Independent Component Analysis (ICA; Thomas et al., 2002; Kochiyama et al., 2005; De Martino et al., 2007; Perlberg et al., 2007; Tohka et al., 2008; Kundu et al., 2012; Bhaganagarapu et al., 2013; Rummel et al., 2013; Storti et al., 2013; Salimi-Khorshidi et al., 2014b). ICA is a data-driven approach that decomposes fMRI data into signal (e.g. functional networks) and structured noise (e.g. motion-induced variance and cardiac pulsation; Beckmann et al., 2005) components. Most ICA-based noise-removal strategies implement automatic classification of components representing noise. A recent example is ‘FMRIB’s ICA-based X-noiseifier’ (ICA-FIX) that implements noise component classification using an extensive set of features and a multi-level classifier (Salimi-Khorshidi et al., 2014). Yet, the complexity of such classifiers hampers their generalizability across datasets. Accordingly, these classifiers typically require re-training for every new dataset (for an exception see Bhaganagarapu et al., 2013). However, re-training is not trivial and entails manual component labeling in data of multiple participants that subsequently have to be excluded from further analyses.

To overcome drawbacks associated with current strategies for removal of secondary motion artifacts, we proposed an alternative strategy called ‘ICA-based Automatic Removal Of Motion Artifacts’, or ICA-AROMA (Pruim et al., 2015). ICA-AROMA identifies components representing motion-related artifacts by employing four theoretically motivated features embedded in a simple and robust classifier that avoids the need for classifier (re-)training across studies. The incorporated features evaluate the spatial structure of the component spatial maps with respect to overlaps with the edge of the brain and CSF within the brain. In addition, two temporal features evaluate the component time-course with respect to its tendency towards high-frequencies and its correlation with realignment parameters (RPs). Subsequent to their identification, components classified as motion-related are removed from the fMRI dataset by means of linear regression. Importantly, ICA-AROMA does not result in the removal of volumes but preserves the integrity of the fMRI time-courses. Accordingly, ICA-AROMA aims at preserving tDoF, increasing statistical power for any down-stream between-subject analysis.

Here, we provide an in-depth evaluation of ICA-AROMA using rfMRI data while comparing our strategy to alternative motion artifact removal strategies. Previous evaluations of strategies for motion artifact removal focused on their potential to remove motion artifacts (Power et al., 2012, 2014; Satterthwaite et al., 2013a; Yan et al., 2013), while the importance of retaining signal of interest is often disregarded. Importantly, we not only evaluate the techniques with respect to the ability to remove unwanted motion-induced signals of no interest in the data but
also characterize the ability to preserve the identification of resting-state networks as signals of interest. Here, we operationalized the latter aspect by evaluating the ability to clearly identify resting-state networks, as well as by assessing the reproducibility of resting-state networks across random splits of the included data. Moreover, we evaluated the loss in tDoF induced by the different strategies. An additional concern is that currently available evaluations of motion artifact removal strategies focus on healthy controls and lack replication in clinical samples. Yet, clinical samples are often most affected by motion-related artifacts, potentially biasing group comparisons as denoising strategies might work differently across sub-groups, thereby introducing biases in the cross-group comparison. To mitigate these concerns, we assessed motion artifact removal, preservation of signal of interest and lost tDoF in rfMRI-based functional connectivity analyses across four datasets including one comprising participants with ADHD.

Materials & methods

To evaluate ICA-AROMA we compared our strategy to a range of alternative strategies that aim to remove motion-related artifacts from rfMRI data. All strategies were applied after primary motion correction by means of volume-realignment (see fMRI data preprocessing below). We compared the following nine strategies:

1. ‘no MC’: no secondary motion correction
2. ‘6RP regression’: extensive nuisance regression, including 6 motion regressors
3. ‘24RP regression’: extensive nuisance regression, including 24 motion regressors
4. ‘spike regression’: 24RP regression including additional spike regressors (Lemieux et al., 2007; Satterthwaite et al., 2013a)
5. ‘scrubbing’: 24RP regression followed by scrubbing (Power et al., 2012)
6. ‘aCompCor’: 24RP regression additionally including a set of principle components derived from WM and CSF signals (Behzadi et al., 2007)
7. ‘SOCK’: Spatially Organized Component Klassifikator (Bhaganagarapu et al., 2013)
8. ‘ICA-FIX’: FMRIB’s ICA-based X-noiseifier (Salimi-Khorshidi et al., 2014)
Datasets

We conducted our statistical comparison in four different datasets. Table 1 provides an overview of the participant characteristics for each dataset. The first and second datasets are derived from the NeuroIMAGE project, a large project aimed at studying attention-deficit/hyperactivity disorder (ADHD) in adolescence (von Rhein et al., 2015). As a third and fourth evaluation we respectively included the Power2012 dataset and the healthy controls from the NYU Child Study Center ADHD200 dataset, both publically available through the International Neuroimaging Data-sharing Initiative (INDI; http://fcon_1000.projects.nitrc.org/).

Dataset 1 — ‘NI Controls’: 100 healthy controls selected from the NeuroIMAGE project. Participants were excluded if their data contained less than 125 volumes after scrubbing, as proposed by Power et al., (2012), or if they belonged to the 5% highest movers as determined by a motion summary score. As a motion summary score we defined the root mean square of the frame-wise displacement time-series (RMS-FD) derived from the realignment parameters (Jenkinson et al., 2002). Of note, the 30 randomly selected controls used for the
development of ICA-AROMA (Pruim et al., 2015) were not included in the current selection.

Dataset 2 — ‘NI ADHD’: To replicate the findings in a clinical sample we also selected 100 ADHD participants from the NeuroIMAGE project. The participants were pair-wise matched to the participants included in the NI Controls dataset, using the RMS-FD motion summary scores. As for the NI Controls, participants were not selected if their data contained less than 125 volumes after scrubbing.

Dataset 3 — ‘Power’: This dataset was used for the development of the motion scrubbing strategy (Power et al., 2012). Most participants were scanned in multiple short runs. We conducted preprocessing and first-level analyses separately for every run, and combined results for every participant at the second-level. Runs within the 5% highest RMS-FD and/or with less than 50 volumes left after scrubbing were excluded. In addition, we excluded participants with less than 125 volumes left after scrubbing combined across runs. These criteria excluded eight participants, leaving 69 participants for further analyses.

Dataset 4 — ‘NYU’: The NYU Child Study Center data is part of the ADHD200 sample in INDI (The ADHD-200 Consortium, 2012). We only included healthy controls from this dataset. The NYU data included shorter rfMRI scanning sessions (175 or 176 volumes) and might inherently be less affected by motion-induced artifacts. We included this dataset as motion-correction strategies have to generalize across datasets with varying imaging parameters and thus potentially varying levels of motion artifacts. Participants with-in the 5% highest RMS-FD or with less than 125 volumes left after scrubbing were excluded. As a result 92 participants (out of 110) were selected for further analyses.

**MRI acquisition**

Table 2 provides an overview of the MRI acquisition parameters for each dataset, highlighting variability in TR, TE, voxel size, type of MR-sequence, slice acquisition order, and field strength. For more specific information on the scanning parameters of the NeuroIMAGE datasets we refer to von Rhein et al. (2015). Details on the Power and NYU MR acquisition parameters can be found on the INDI website (http://fcon_1000.projects.nitrc.org). Importantly, the NeuroIMAGE data was collected at two different scanner locations and the Power dataset consisted of multiple cohorts acquired with varying MRI protocols.
fMRI data preprocessing

Preprocessing of the rfMRI data was performed using the FMRIB Software Library (FSLv5.0, available at http://www.fmrib.ox.ac.uk/fsl; Smith et al., (2004); Woolrich et al., (2009); Jenkinson et al., (2012)). Standard preprocessing steps included (1) removal of the first five volumes to allow for signal equilibration, (2) head movement correction by volume-realignment to the middle volume using MCFLIRT, (3) global 4D mean intensity normalization, (4) spatial smoothing (6 mm FWHM), and (5) high-pass filtering (cut-off frequency of 0.01 Hz). Additional preprocessing differed across the applied motion artifact removal strategies and will be discussed below. We co-registered rfMRI datasets to the participant’s structural image using affine boundary-based registration as implemented in FSL FLIRT (Jenkinson and Smith, 2001; Jenkinson et al., 2002; Greve and Fischl, 2009) and subsequently transformed them to MNI152 standard space with 4 mm isotropic resolution using non-linear registration through FSL FNIRT (Andersson et al., 2007).

In addition to these initial preprocessing steps, each motion artifact removal strategy required subsequent processing. Supplementary Figure 1 specifies the preprocessing pipelines for each of the nine strategies. In the no-MC strategy we applied nuisance regression using three nuisance regressors: white matter (WM) signal, CSF signal and a linear trend. White matter and CSF time-series were derived by determining the mean time-series over voxels within a conservative predefined WM and CSF mask (see Supplementary Figure 1). Importantly, nuisance regression was applied prior to high-pass filtering. In the 6RP regression strategy we additionally included six rigid body parameters, resulting from the volume-realignment, in the nuisance regression model. The 24RP regression strategy included a nuisance regression model including the six rigid body parameters, their derivatives, the squared version of the rigid body parameters, the squared derivatives, as well as the three standard nuisance regressors.

Table 2 - Resting-state fMRI acquisition characteristics of the datasets included in the study. NeuroIMAGE included both NI Controls and NI ADHD.

<table>
<thead>
<tr>
<th></th>
<th>NeuroIMAGE</th>
<th>Power</th>
<th>NYU</th>
</tr>
</thead>
<tbody>
<tr>
<td>No. of volumes per run</td>
<td>265/266</td>
<td>76-164</td>
<td>175-176</td>
</tr>
<tr>
<td>No. of runs</td>
<td>1</td>
<td>01-Jun</td>
<td>1</td>
</tr>
<tr>
<td>TR (s)</td>
<td>1.96</td>
<td>2.0-2.5</td>
<td>2.0</td>
</tr>
<tr>
<td>TE (ms)</td>
<td>40</td>
<td>27</td>
<td>15</td>
</tr>
<tr>
<td>Voxel size (mm)</td>
<td>3.5 x 3.5 x 3.0</td>
<td>4.0 x 4.0 x (4.0-6.8)</td>
<td>3.0 x 3.0 x 4.0</td>
</tr>
<tr>
<td>MR-sequence</td>
<td>GE-EPI</td>
<td>GE-EPI</td>
<td>ME-EPI</td>
</tr>
<tr>
<td>Slice acquisition</td>
<td>Ascending</td>
<td>Interleaved</td>
<td>Interleaved</td>
</tr>
<tr>
<td>B-field strength</td>
<td>1.5 T</td>
<td>3 T</td>
<td>3 T</td>
</tr>
</tbody>
</table>

GE: gradient-echo and ME: multi-echo.
Scrubbing and spike regression require labeling fMRI volumes associated with high motion. To that end we used the frame-wise displacement time-series (FD; Jenkinson et al., (2002)) and labeled volumes with FD > 0.2 mm, as well as its preceding and the following volume. A single spike regressor was constructed for each of these frames and added to the 24RP regression model in the case of spike regression. In the case of scrapping we deleted these frames from the residual fMRI data after 24RP regression and high-pass filtering.

aCompCor (Behzadi et al., 2007) is a nuisance regression strategy where realignment parameters are complemented by principle components derived from WM and CSF signals. As proposed by Muschelli et al., (2014), we constructed a nuisance model comprising the 24 realignment parameters complemented by the principle components that explained 50% of respectively the WM and CSF signals. We applied aCompCor before spatial smoothing and after temporal detrending.

We applied ICA-AROMA after spatial smoothing, and complemented the preprocessing with nuisance regression using WM, CSF, and a linear trend to regress out residual structured noise. After nuisance regression we applied high-pass filtering (Pruim et al., 2015). In contrast, ICA-FIX was applied after high-pass filtering and without subsequent nuisance regression (Salimi-Khorshidi et al., 2014). No preprocessing streamline was described for SOCK (Bhaganagarapu et al., 2013). SOCK implements a generic ICA-based classifier aimed at removing all types of structured noise, and additional nuisance regression was not specifically recommended. Therefore, we implemented SOCK analogous to ICA-FIX.

Note that for ICA-FIX it is recommended to re-train the classifier when applying it to a new dataset (Salimi-Khorshidi et al., 2014). Re-training presents a formidable challenge, particularly in the case of multi-site studies as either one would need to define a balanced single training dataset (incorporating data and therefore noise features from all sites) or implement multiple training runs (one per site), thereby potentially inducing ‘denoising bias’ in any subsequent comparison. Since our aim is to evaluate ICA-AROMA, which was specifically designed to avoid re-training, we did not re-train ICA-FIX but applied the ICA-FIX classifier as trained for application to a standard fMRI dataset (Salimi-Khorshidi et al., 2014).

To ensure that the results of the different preprocessing strategies are comparable we embedded them within a single, ‘minimal’ preprocessing steam. As such, we did not include extensions of WM and CSF signals (derivative or squared), GSR and/or band-pass filtering for any of the strategies. Naturally, the order of preprocessing steps could differ across strategies as required (e.g. aCompCor before spatial smoothing). Note that our embedding can differ from the processing pipeline
as used in each strategy's respective manuscript. As GSR and band-pass filtering are frequently considered, yet actively debated, within rfMRI research, we added a supplementary evaluation of these processing steps (see Supplementary materials).

Statistical analyses

Figure 1 outlines the different analyses that we conducted to assess the sensitivity and specificity of the nine motion artifact removal strategies. To that end we assessed the quality of motion artifact removal (sensitivity), and the preservation of signal of interest (specificity). In short, assessing the quality of the motion artifact removal focused on directly comparing participants with low motion to participants exhibiting high levels of movement during scanning. If secondary motion artifact removal strategies work, differences between participants exhibiting low and participants exhibiting high motion should be reduced to a minimum. In complementary analyses, we assessed whether the different strategies preserved signal of interest, i.e. potentially removed signal of interest along with motion-related artifacts. Both analyses were applied to different functional connectivity measures as described below. Additionally, we evaluated the strategies regarding their induced loss in tDoF.

Functional connectivity measures

Dual regression

For each participant we derived spatial maps of well-validated and commonly replicated resting-state networks (RSNs) using dual regression (Beckmann et al.,
ICA-AROMA evaluation

2009; Filippini et al., 2009). This analysis implements a multivariate spatial regression of a set of initial templates against the preprocessed fMRI data of every participant, and yields participant-specific time series for each template. Next, these time series are entered in a multivariate temporal regression against the same preprocessed data resulting in participant-level spatial representations, both parameter estimate (PE) and Z-statistical maps, of the initial templates. As templates, we used the 20 spatial maps as described by (Smith et al., 2009). These templates include 10 RSNs which were found to correspond well to networks involved in task-related processing. The other 10 templates are artifactual or represent more complex networks. Dual regression was performed using the full set of components to optimally model the data. We selected the 10 RSNs for use in subsequent analyses. These RSNs include the visual-medial (Vis-med), visual-occipital (Vis-occ), visual-lateral (Vis-lat), default mode (DMN), cerebellum, sensorimotor, auditory, executive control, right fronto-parietal (right FP) and left frontal-parietal (left FP) networks.

Dual regression was applied to each preprocessed rfMRI dataset. Of note, the Power dataset contained multiple runs for most participants. We applied dual regression to each single run. Subsequently, we derived participant-specific PE-maps by means of a within-subject fixed effects analysis, and Z-statistical maps by averaging the obtained Z-statistical spatial maps across runs.

Seed-based functional connectivity

We further evaluated the performance of the motion artifact removal strategies in the context of seed-based (SB) functional connectivity analyses. We included two analyses adapted from Power et al., (2012b). The first analysis comprised a whole brain seed-based regression using a medial parietal seed (MNI coordinates: x = −7 mm, y = −55 mm and z = 27 mm) to estimate a spatial map for the default mode network (DMN).

In addition we employed a seed-based correlation analysis using 264 ROIs (Dosenbach et al., 2010). The mean time-series of every seed was derived using a 10 mm sphere centered on the seed coordinates. All pair-wise correlation scores were converted to Z-scores by means of a Fisher r-to-Z transformation. In the Power dataset, correlation Z-scores were calculated for each run separately, and then averaged to obtain participant-specific fixed-effects estimates.

Assessing the quality of motion artifact removal

We assessed the ability of each strategy to remove motion artifacts by comparing the RSNs of participants exhibiting low head motion to RSNs of participants
exhibiting high amounts of head motion. To that end we divided each dataset into three subsamples: low, medium and high motion subsamples representing respectively 25%, 50% and 25% of each sample. The subdivision was guided by a single summary motion score (RMS-FD; root mean square of the frame-wise displacement time-series) for every participant.

To compare participants exhibiting low vs. high amounts of head motion we employed a between-group comparison within a General Linear Model (GLM), implemented through non-parametric permutation testing (5000 permutations). The three movement groups (low, medium, high) were included as separate group variables. Age and gender were included as covariates. Analyses on the NeuroIMAGE sample additionally included IQ and scan location as covariates. Contrasts of interest were 'low > high' and 'high > low'. Group-level t-statistical maps were thresholded using threshold-free cluster enhancement and corrected for multiple comparisons using family wise error (FWE) correction with p < 0.05. This comparison was made for each RSN separately (10 dual regression-based RSNs and 1 seed-based RSN). As a supplementary analysis, we replicated the group-based motion comparison using a dimensional approach by including RMS-FD as an effect of interest instead of separate group variables.

The effect of motion on the 264 × 264 SB-correlation matrices was assessed qualitatively by calculating the absolute correlation across participants of every seed-pair Z-score (264 × 264) with the motion summary score RMS-FD. Subsequently, we converted these correlation scores to Z-scores by means of a Fisher r-to-Z transformation and assessed the distribution of the resulting values. The goal of the respective strategies is to avoid spurious correlations related to motion artifact. Accordingly, in case the SB-correlations are unrelated to motion, the distribution of their correlations with the RMS-FD score is expected to approximate zero with a small standard deviation. We further evaluated the impact of motion on the Z-score correlation matrices using a categorical and dimensional approach analogous to the analyses on the RSN spatial maps described above with the exception that the t-statistical maps were not thresholded using threshold-free cluster enhancement. Results of these analyses are presented in the Supplementary materials.

**Testing the preservation of signal of interest**

We assessed to what extent each strategy preserved, or even in-creased, sensitivity towards signal of interest by evaluating the ability to identify and reproduce the estimated RSNs. RSN identifiability was defined in terms of a Z-score ratio between the mean absolute Z-score inside and outside a RSN-mask. This ratio was calculated
for every dual regression-based RSN at the participant-level, using the original templates thresholded at $|Z| > 2.3$ as masks. Similarly, for the seed-based DMN map, we calculated the ratio between the mean absolute Z-score inside and outside the DMN mask derived from the DMN template used in the dual regression analysis. By comparing signal within the RSN spatial map to noise outside the map, this score represents a signal to noise ratio of every participant-level RSN spatial map. As an example, a ratio of 1 would indicate that the RSN is indistinguishable from the surrounding voxels as both would have equal Z-scores. In contrast, a ratio $> 1$ indicates that the Z-scores within the mask were higher compared to the Z-scores of the surrounding voxels, suggesting increased RSN identifiability.

RSN reproducibility was investigated using split-half reproducibility. We randomly divided the total group of participants per sample into two equally sized groups. For both groups, we derived the average group-level spatial PE map across participants, for each of the 20 tem-plate ICs as estimated by dual regression. The group-level maps were masked to only include gray matter voxels using a MNI152 gray matter probability map thresholded at 50%. Subsequently, we determined the between-group spatial correlation for each spatial map, yielding a $20 \times 20$ correlation matrix. The correlation values on the diagonal of this matrix expressed the similarity (or reproducibility) of each IC between both groups (e.g., spatial correlation between the DMN of group 1 and the DMN of group 2). In contrast, the off diagonal correlations expressed the spatial correlation between non-matching IC tem-plates (e.g., the spatial correlation between the DMN of group 1 and the sensorimotor network of group 2). We used the off diagonal correlations as a null-distribution to convert the spatial correlations for the 10 matching RSNs to pseudo Z-scores. Analogously, we calculated the spatial correlation between the seed-based DMN group-level maps obtained for each group. We used the null distribution as estimated for the 20 dual regression-based maps to convert the between-group seed-based DMN spatial correlation to a pseudo Z-score. To obtain average Z-scores and standard deviations we conducted the reproducibility analyses for 500 random group splits.

In addition to RSN identifiability and reproducibility estimates, we used the $264 \times 264$ seed-based correlation matrix to assess preservation of the correlation structure in the data. We derived the mean Z-score correlation matrix across participants per dataset and assessed the results qualitatively. In addition, we tested the correlation structure more specifically by implementing a mixture modeling approach to partition the seed-pairs respectively into three categories: anti-correlated, uncorrelated and correlated seed-pairs. Motion artifact removal strategies that effectively remove motion-related noise but preserve signal of interest will amplify differences between the three categories and therefore improve
Chapter 2

the detection of (anti-)correlated seed-pairs. We therefore compared the mixture modeling results with respect to the results of the no-MC strategy to evaluate changes in sensitivity towards significant correlation induced by the different strategies.

Of note, it should be clear that we evaluate all techniques using both inherently seed-based approaches as well as dual-regression-based characterization of ICA-derived networks, thus avoiding potential bias with respect to the implemented strategies. In addition, despite the fact that the spatial templates used for our evaluation were derived from an ICA decomposition (Smith et al., 2009), our evaluation metrics will not be biased towards ‘good’ results for ICA-based motion artifact removal strategies. This is because we do not evaluate the spatial quality of individual representations of each template. Instead we focus on the similarity of the obtained representations between datasets (whether they are a good representation of the initial template or not).

Assessing the loss in temporal degrees of freedom

We evaluated the loss in tDoF associated with the different strategies to investigate its potential impact on statistical power and between-group bias related to a variable loss in tDoF. For every participant we determined the number of lost tDoF associated with each of the strategies by regarding every nuisance regressor, component and/or fMRI volume which was regressed out or removed from the data as a single tDoF. To obtain comparable results between participants we expressed the number of lost tDoF as a fraction of the total available tDoF which we defined as the total number of volumes within the fMRI time-series. Note that these common estimates of loss in tDoF and available tDoF ignore temporal autocorrelation present in the data. The actual number of tDoF is therefore likely to be lower than these estimates. Finally, temporal filtering will also result in lost tDoF and should ideally be accounted for. We employed high-pass filtering by conducting a local Gaussian-weighted line-fitting procedure as implemented in fslmaths. However, as temporal filtering is equal for all strategies included in our evaluation we did not incorporate lost tDoF induced by temporal filtering into our tDoF calculations.

Results

Quality of motion artifact removal

As illustrated in Figure 2, applying no specific correction for secondary motion artifacts (no-MC), yielded prominent differences between the low and high motion subsamples throughout the complete set of RSNs. A large portion of significant
effects persisted after preprocessing the data using RP-based regression strategies, aCompCor or SOCK. Spike regression, scrubbing, ICA-FIX and ICA-AROMA on the other hand reduced the differences between participants exhibiting low and high amounts of head motion to a minimum. These results were consistent across datasets except in the NYU dataset where the initial differences between low and high movers were less prominent but clearly more resistant towards removal by the different strategies (see Figure 2 and Supplementary Figure 2). Results of the dimensional group-level motion analysis were in line with the categorical results presented here (see Supplementary Figure 3).

The assessment of motion artifact removal by means of evaluating the correlation between RMS-FD and seed-based correlation functional connectivity estimates yielded comparable results. RP-based regression strategies, aCompCor and SOCK resulted in substantial correlation between seed-pair correlation Z-scores and RMS-FD. In contrast, spike regression, scrubbing, ICA-AROMA and to a lesser extent ICA-FIX decreased such correlation towards zero with small standard deviation across seed pairs (Figure 3, upper triangles and bottom right panel; see

**Figure 2** - Significant group-level differences between participants exhibiting low and high head motion (threshold-free cluster enhanced, p < 0.05, FWE-corrected). Left panel: Results of the between group-comparison in the NI Controls dataset, for all tested RSN spatial maps and preprocessing strategies. Each single map illustrates a RSN mask (gray) overlapped with a map of the significant between-group differences (green). The original 3D spatial maps are summed and binarized over all axial slices, i.e. a green pixel reflects that at least one voxel along that z-axis showed a significant residual motion-induced effect. See Supplementary Figure 2 for similar results for all datasets. Right panel: Percentage of voxels exhibiting a significant group-level difference in each of the included datasets. This percentage was calculated for each dataset across all 11 RSNs as shown in the left panel (columns). See Supplementary Figure 3 for similar results when employing a dimensional instead of categorical analysis, and for replication of these analyses on seed-based Z-score correlation matrices.
Supplementary Figure 3 illustrates the results of a categorical and dimensional analysis of the group-level effects of motion. Results confirmed the effects observed in Figure 3.

**Preservation of signal of interest**

RSN identifiability scores were highly consistent across motion artifact removal strategies, datasets and RSNs (Figure 4; see Supplementary Figure 5 for the results on individual datasets and RSNs). An important exception was ICA-FIX, which often resulted in decreased identifiability scores, especially in the NYU dataset. An identifiability score of 1 indicates no difference between Z-scores within and Z-scores outside the RSN mask, or, in other words, no contrast between signal and noise. Such decreased ratios suggest substantial removal of signal of interest during the artifact removal procedure. Note, however, that here we evaluate ICA-FIX without explicit re-training of the classifier, i.e. using the default classification criteria derived from a separate rfMRI study (Salimi-Khorshidi et al., 2014).

RSN reproducibility on the other hand, varied across strategies (Figure 4; see Supplementary Figure 5 for individual datasets and RSNs). Extensive nuisance regression strategies, scrubbing and spike regression resulted in the lowest...
ICA-AROMA evaluation

reproducibility scores, aCompCor exhibited slightly increased scores, while ICA-based strategies SOCK, ICA-FIX, and ICA-AROMA yielded the highest reproducibility scores. ICA-AROMA exhibited the most consistent pseudo Z-scores across the four datasets (see Supplementary Figure 5). Of note, it is evident from Figure 4 that all strategies resulted in high reproducibility Z-scores (i.e., the minimum observed pseudo Z-scores was 10.8). Yet, a clear distinction could be made between the RP-based and ICA-based motion artifact removal strategies.

Finally, all motion artifact removal strategies largely preserved the correlation structure observed between 264 ROIs (Figure 3, lower triangles). Supplementary Figure 4 illustrates the correlation structure for each dataset, highlighting that ICA-FIX completely destroyed the correlation structure in the NYU dataset, corroborating the low RSN identifiability observed for ICA-FIX in the NYU dataset and suggesting that dataset-specific classifier training would be necessary for acceptable denoising performance to be achieved. Mixture modeling analyses on the mean seed-based correlation matrices yielded increased and profoundly more consistent detection of correlated seed-pairs after applying ICA-AROMA compared to all alternative strategies (see Supplementary Figure 6 and Supplementary Table 1). Additionally detected connections were primarily located within clusters already detected when employing the no-MC strategy; therefore these additional connections likely reflect true-positives.

![Figure 4 - RSN identifiability and RSN reproducibility. RSN identifiability is quantified using a Z-score ratio expressing the Z-scores within a RSN relative to Z-scores outside the RSN. RSN reproducibility scores reflect the spatial correlation between group-level RSN maps (i.e. consistency) for random splits of the samples, normalized to pseudo Z-scores. Results are shown for all processing strategies in each dataset, combined across all 11 RSNs (10 dual regression + 1 seed-based).](image-url)
Loss in temporal degrees of freedom

The fractional loss in tDoF strongly differed between the nine strategies (see Figure 5). No secondary motion correction and 6RP regression consistently resulted in the lowest and least variable loss in tDoF whereas spike regression and scrubbing resulted in the highest and most variable loss in tDoF. For the ICA-based strategies, SOCK preserved most tDoF across all datasets whereas ICA-AROMA and ICA-FIX yielded comparable results across samples. However, in the Power and NYU datasets the loss in tDoF associated with ICA-AROMA was significantly lower compared to ICA-FIX (paired t-test, p < 0.01). Likewise, the difference between ICA-AROMA and 24RP regression was non-significant for the NI-samples (p = 0.38 and p = 0.12) and the Power dataset (p = 0.91), but significantly lower for the NYU datasets (p < 0.01).

Discussion

We evaluated ICA-AROMA and alternative strategies aiming at removal of motion artifacts from rfMRI data. Importantly, we not only focused on the ability of each strategy to remove motion-related artifacts, but additionally evaluated how well each strategy preserved signal of interest. Results were replicated across four datasets, including one clinical sample. ICA-AROMA performed equally well as
spike regression, scrubbing and ICA-FIX in removing motion-related artifacts. ICA-AROMA and ICA-FIX both yielded improved RSN reproducibility and decreased loss in tDoF compared to spike regression and scrubbing. However, without re-training the classifier, ICA-FIX decreased the level of signal of interest whereas ICA-AROMA successfully preserved signal of interest across all datasets.

Corroborating previous reports we observed that head motion during rfMRI scanning has a large impact on functional connectivity estimates (Power et al., 2012; Satterthwaite et al., 2012; Van Dijk et al., 2012; Yan et al., 2013). Across RSNs we observed widespread differences between participants exhibiting low and high amounts of head motion.

Nuisance regression including realignment parameters was unsuccessful in addressing such effects of secondary motion artifacts. The 6RP and 24RP regression strategies minimally reduced the impact of motion on RSNs and SB-correlation measures. This finding is consistent with previous results (Power et al., 2012; Satterthwaite et al., 2012; Van Dijk et al., 2012; Yan et al., 2013) and further highlights the limited applicability of realignment parameters to model the full complexity of secondary motion-induced variance (e.g. spin history effects). Extension of the nuisance model by principle components derived from WM and CSF as implemented in aCompCor had no benefit regarding the reduction of group-level motion effect. In contrast, spike regression, scrubbing, ICA-FIX and ICA-AROMA were successful in reducing significant effects of head motion. A notable exception was SOCK, which exhibited limited ability to reduce motion-related group-level differences and generally showed large amounts of variability with respect to its denoising ability.

Additionally we investigated how the different strategies affected group-level variability of RSN spatial maps. To that end we assessed RSN reproducibility, and found that all ICA-based strategies resulted in profoundly increased reproducibility scores relative to the alternative strategies or no additional motion denoising. This suggests that that ICA-based strategies remove structured noise from fMRI data more efficiently, decreasing group-level variability of RSN spatial maps. In contrast, nuisance regression, spike regression and scrubbing exhibited reproducibility scores that were similar to when no additional motion denoising was applied, suggesting that these artifact removal strategies are not only specific to motion artifact, but equally affect signal of interest. This suggestion is corroborated by evidence showing that fMRI data contained substantially higher levels of motion-related and generic noise, relative to signal of interest, after extensive nuisance regression and spike regression compared to ICA-AROMA (Pruim et al., 2015). Notably, whereas aCompCor did not have added value regarding the removal of group-level motion...
effects it did slightly increase RSN reproducibility compared to the nuisance regression strategies. This result suggests that the set of WM and CSF regressors derived by PCA particularly contributes to the removal of structured noise other than motion artifacts.

ICA-FIX increased RSN reproducibility and has been demonstrated to have the potential for accurate component classification and denoising of fMRI data (Griffanti et al., 2014; Salimi-Khorshidi et al., 2014). However, without re-training, we found that the ICA-FIX classifier was not sufficiently generalizable across datasets and resulted in decreased levels of signal of interest. In contrast, SOCK preserved signal of interest across datasets but at the cost of poor removal of motion-related noise. Such decreased sensitivity for SOCK has been previously documented (Sochat et al., 2014). The variability between ICA-based strategies directly reflects the trade-off that such strategies have to overcome regarding the sensitivity and specificity of the classifier that they implement. By implementing a small set of specific, standardized and theoretically motivated features, ICA-AROMA was designed to allow high sensitivity and specificity while achieving robust classification performance. Indeed, ICA-AROMA was the only ICA-based strategy within our evaluation that retained signal of interest while reducing motion artifacts across datasets.

Moreover, ICA-AROMA showed very consistent results across the four datasets. Most notably, ICA-AROMA resulted in consistent RSN reproducibility scores (see Figure 4) and detection of functional connections using seed-based analysis (see Supplementary Figure 6). Such consistency was achieved despite the large variance in MRI acquisition parameters across datasets. This not only demonstrates the applicability of ICA-AROMA to new datasets without re-training, but it additionally indicates its potential to improve the reliability of FC estimates. Functional connectivity metrics derived from fMRI data have already been shown to be reliable across participants, scan sequences, imaging sites, and time (Van De Ven et al., 2004; Damoiseaux et al., 2006; Shehzad et al., 2009; Biswal et al., 2010; Van Dijk et al., 2010; Zuo et al., 2010; Wisner et al., 2013). However, there is still considerable improvement possible in the reliability of these metrics (Wisner et al., 2013; Zuo and Xing, 2014). Although the impact of ICA-AROMA on within-participant test–retest reliability requires additional investigation, the increased between-participant reproducibility illustrates that ICA-AROMA can possibly facilitate such improved reliability. By improving the consistency of FC estimates over scanner-sites and datasets ICA-AROMA furthermore enhances the potential of multi-site studies and data-sharing initiatives (e.g. FCON1000, ADHD200, ABIDE) which typically suffer from variability across the total sample due to different scanner sites and MRI acquisition protocols. We furthermore showed that ICA-AROMA generalizes to clinical datasets such as those provided in the ADHD200
ICA-AROMA evaluation

(http://fcon_1000.projects.nitrc.org/indi/adhd200/) or ABIDE sample
(http://fcon_1000.projects.nitrc.org/indi/abide/) by replicating our findings in the NYU dataset and the NI ADHD dataset.

Despite controversy about their usage, global signal regression (GSR) and band-pass filtering are often considered in rfMRI research. While band-pass filtering exhibited some positive effects in reducing significant group-level motion-related effects, we observed no added benefit of GSR in addition to 24RP regression, scrubbing or spike-regression (see Supplementary Figure 7). Importantly, both GSR and specifically band-pass filtering decreased signal of interest as reflected by decreased RSN identifiability. These findings are not surprising when considering that the global signal is a superposition of both signal and noise components, and that higher frequencies contain signal of interest (Niazy et al., 2011; Liao et al., 2013; Kalcher et al., 2014).

We did not specifically evaluate the usefulness of combining multiple strategies that could potentially complement each other (e.g. 24RP-regression and SOCK) and did not include all currently available strategies for motion artifact removal, e.g. wavelet-despiking (Patel et al., 2014), SLOMOCO (Beall and Lowe, 2014), RDI (Spisák et al., 2014) or alternative ICA-based strategies (Thomas et al., 2002; Kochiyama et al., 2005; De Martino et al., 2007; Perlberg et al., 2007; Tohka et al., 2008; Kundu et al., 2012; Rummel et al., 2013; Storti et al., 2013; Sochat et al., 2014). In our evaluation we included a set of strategies that varied in their underlying principles (e.g. volume-removal, RP-model regression, component-regression), are currently most considered within the research field, are applicable across datasets, and are representative for the wide spectrum of available strategies.

Similar to the alternative strategies, ICA-AROMA did not achieve full removal of significant differences between participants exhibiting low versus high head motion (see Figure 2). The question remains whether these effects reflect motion artifact or neurobiological correlates which are directly (e.g. motor control) or indirectly (e.g. age) related to motion. Several studies suggest that such neurobiological correlates exist (Van Dijk et al., 2012; Couvy-Duchesne et al., 2014; Kong et al., 2014; Pujol et al., 2014; Zeng et al., 2014). In this regard, we note that the residual significant differences found within the resting-state networks appear to be consistent across datasets, comprising the cerebellum, sensorimotor, auditory and executive control networks. The cerebellum and sensorimotor network are involved in motor control, potentially suggesting that these findings might be related to neurobiological underpinnings of head motion. This hypothesis is particularly supported by the analysis of the NYU dataset which almost exclusively resulted in significant group-level effects within the four previously listed networks, and which
were more resistant to removal by the different strategies (disregarding the seed-
based results). These results are striking since the NYU dataset comprises
participants with profoundly lower levels of head motion (see Table 2). However,
these participants were scanned at a substantially earlier stage within their
neurodevelopmental trajectory (12.4 ± 3.1 years) compared to the participants in the
other datasets, providing support for the idea that the observed effects might be
related to neurodevelopment, explaining variance not captured by the age
covariates. Yet, additional research is required to reliably disentangle potential
neurobiological effects from motion-related artifacts. It is for instance important to
note that most denoising procedures discussed in the current manuscript implement
artifact removal through linear regression whereas head-motion is likely to induce
highly non-linear effects. Although such non-linear effects can be approximated by
a set of components in data-driven strategies, any residual motion-related group
effects can be related to inadequate modeling of the non-linear motion-related effects
and hence sub-optimal denoising. Similarly, volume-removal strategies depend on a
binary decision on a volume being affected by head motion and inherently will not
remove the full (non-linear) dynamics associated with motion artifacts.

Residual motion-related effects as discussed above are of particular importance
when considering group-level co-varying using summary motion scores. Though this
might prevent false positives, it might also reduce sensitivity towards any effect of
interest sharing variance with motion scores, e.g. age, gender, motion-related traits,
or neural activity (Van Dijk et al., 2012; Couvy-Duchesne et al., 2014; Kong et al.,
2014; Pujol et al., 2014; Zeng et al., 2014). In case residual differences between
participants exhibiting low and high head motion are attributable to artifacts, the
question remains whether and how such residual noise impacts future analyses. We
note that our group comparison compared two highly distinct samples comprising
the 25% lowest versus the 25% highest movers, without any overlap in motion
summary scores. Such extreme group differences are unlikely to appear in typical
fMRI re-search. Therefore we do not expect such residual noise to profoundly bias
typical fMRI between-group comparisons. However, with increasing sample sizes or
when testing continuous variables that potentially correlate with motion levels (e.g.
age or hyperactivity), investigators should interpret results carefully. To rule out
spurious effects related to residual motion artifacts, one could validate the results
in motion-matched groups (Satterthwaite et al., 2013b) or investigate the robustness
of the results when incorporating group-level motion covariates (Yan et al., 2013).
Conclusion

We provided an extensive evaluation of currently available strategies for motion-artifact removal from rfMRI data, by means of comparing our strategy, ICA-AROMA, to extensive nuisance regression utilizing 6 or 24 realignment parameters, spike regression, scrubbing, aCompCor, SOCK, and ICA-FIX. Our results indicated that scrubbing, spike-regression, ICA-FIX and ICA-AROMA minimized the impact of motion artifacts to a large extent, in contrast to extensive nuisance regression, aCompCor, and SOCK which fail at reducing these effects. Despite spike regression and scrubbing being comparable with respect to removing motion artifacts, ICA-FIX and ICA-AROMA resulted in increased reproducibility of resting-state networks while limiting the loss of tDoF and preserving the temporal autocorrelation structure. However, without re-training its classifier, ICA-FIX can have profound impact on signal of interest. ICA-AROMA on the other hand retained signal of interest and resulted in highly consistent results across functional connectivity metrics and datasets, endorsing its robustness and potential added value for multicenter studies.

Acknowledgments

The NeuroIMAGE project was supported by NWO Large Investment Grant 1750102007010, ZonMW Grant 60-60600-97-193, and NWO Brain and Cognition grants 433-09-242 and 056-13-015 to Dr. Buitelaar the EU FP7 grant TACTICS (grantno. 278948), and grants from Radboud University Medical Center, University Medical Center Groningen and Accare, and VU University Amsterdam. Dr. Mennes has received funding from the European Research Council under the European Union’s Seventh Framework Programme (FP7/2007-2013)/ERC grant agreement 327340. Dr. Beckmann is supported by the Netherlands Organisation for Scientific Research (NWO-Vidi 864-12-003) and we gratefully acknowledge the funding from the Wellcome Trust UK Strategic Award [098369/Z/12/Z].

Conflict of interest

The authors declare that they have no conflicts of interest.
References

Chapter 2


ICA-AROMA evaluation

2229.


Supplementary materials

Global signal regression and band-pass filtering

Method

We investigated global signal regression (GSR) and band-pass filtering regarding the removal of motion artifacts, RSN identifiability and reproducibility for RSN spatial maps derived by dual-regression as discussed previously in this chapter. We conducted the analysis on the neuroIMAGE control dataset and evaluated both techniques in addition to 24RP regression. Therefore we tested two variants: 24RP regression including GSR, and 24RP regression including both GSR and band-pass filtering (0.01Hz-0.1Hz). Global signal regression was implemented by adding the mean global signal as a regressor to the nuisance regressor model.

Additionally, we evaluated the preprocessing pipeline as proposed by Satterthwaite et al. (2013) which includes extensive nuisance regression using a model comprising of the 6 realignment parameters, WM, CSF and GS regressors, the derivatives of these nine regressors, and the squared terms of these 18 regressors. We complemented the Satterthwaite model by a linear trend regressor, giving a final model of 37 parameters (37par-model). We investigated this model for both spike regression (Satterthwaite et al., 2013) and scrubbing (Power et al., 2012). The proposed preprocessing pipeline by Satterthwaite et al., (2013) includes band-pass filtering. To evaluate the impact of GSR more specifically we evaluated two variants of the preprocessing pipeline including either high-pass (>0.01Hz) or band-pass filtering (0.01Hz-0.1Hz).

All strategies were compared with 24RP regression, spike regression and scrubbing as evaluated in the main manuscript (see section 2). In summary we investigated the following nine strategies:

24RP

1. '24RP': 24RP regression
2. '24RP & GSR': 24RP regression including GSR and high-pass filtering
3. '24RP & GSR & BP': 24RP regression including GSR and band-pass filtering

Spike Regression

4. 'Spike': Spike regression
5. 'Spike ext': Spike regression including the 37par-model
6. 'Spike ext & BP': Spike regression including the 37par-model and band-pass filtering

**Scrubbing**
7. 'Scrub': Scrubbing
8. 'Scrub ext': Scrubbing after nuisance regression with the 37par-model
9. 'Scrub ext & BP': Scrubbing after nuisance regression with the 37par-model and band-pass filtering

**Results**

Supplementary Figure 7 illustrates the findings on group-level motion effects, RSN identifiability and reproducibility. Global signal regression minimally impacted the findings on group-level motion effects. RSN reproducibility was increased whereas RSN identifiability was decreased, indicating that GSR removes both signal and structured noise. This is a direct reflection of the method as GSR regresses out the global signal which comprises a superposition of both signal and noise components. Band-pass filtering (combined with GSR) on the other hand reduced group-level effects to a minimum. However, band-pass filtering profoundly removes signal of interest as illustrated by a clear decrease of RSN identifiability and slight decrease of RSN reproducibility.

**References**


ICA-AROMA evaluation

**Supplementary Table 1 - Percentage seed-pairs detected as correlated after applying mixture modeling on the mean seed-pair correlation Z-scores averaged across participants within each dataset. Additionally, the results are presented after applying mixture modeling using a merged dataset including all participants across the four datasets. See Supplementary Figure 6 for the associated image.**

<table>
<thead>
<tr>
<th>Method</th>
<th>NI Controls</th>
<th>NI ADHD</th>
<th>Power</th>
<th>NYU</th>
<th>Merged dataset</th>
</tr>
</thead>
<tbody>
<tr>
<td>no MC</td>
<td>13.5</td>
<td>18.0</td>
<td>21.1</td>
<td>24.4</td>
<td>16.1</td>
</tr>
<tr>
<td>6RP regression</td>
<td>13.8</td>
<td>16.2</td>
<td>21.5</td>
<td>26.0</td>
<td>15.4</td>
</tr>
<tr>
<td>24RP regression</td>
<td>14.5</td>
<td>17.8</td>
<td>20.5</td>
<td>25.7</td>
<td>17.6</td>
</tr>
<tr>
<td>Spike regression</td>
<td>15.9</td>
<td>19.6</td>
<td>19.4</td>
<td>25.1</td>
<td>16.7</td>
</tr>
<tr>
<td>Scrubbing</td>
<td>16.0</td>
<td>19.5</td>
<td>20.2</td>
<td>25.3</td>
<td>17.3</td>
</tr>
<tr>
<td>aCompCor</td>
<td>16.5</td>
<td>17.3</td>
<td>19.7</td>
<td>23.2</td>
<td>18.1</td>
</tr>
<tr>
<td>SOCK</td>
<td>15.1</td>
<td>18.1</td>
<td>9.4</td>
<td>15.4</td>
<td>12.5</td>
</tr>
<tr>
<td>ICA-FIX</td>
<td>16.0</td>
<td>17.8</td>
<td>11.3</td>
<td>2.0</td>
<td>13.0</td>
</tr>
<tr>
<td>ICA-AROMA</td>
<td>25.7</td>
<td>27.0</td>
<td>23.6</td>
<td>24.9</td>
<td>24.6</td>
</tr>
</tbody>
</table>
Supplementary Figure 1 - Schematic overview of the fMRI preprocessing performed for the nine considered strategies for motion artifact removal. The green boxes at the top panels present variables required for the 2nd stage preprocessing in the lower panel. NR: nuisance regression; RPs: realignment parameters; WM: white matter; CSF: cerebrospinal fluid; lin: linear trend.
Supplementary Figure 2: Significant group-level differences between participants exhibiting low and high amounts of head motion, for all tested RSN spatial maps and preprocessing strategies (threshold-free cluster enhanced, p<0.05, FWE-corrected). Each single map illustrates a RSN mask (grey) underneath a map of the significant between-group differences (green). The original 3D spatial maps are summed and binarized over all axial slices; i.e. a green pixel reflects that at least one voxel along that z-axis showed a significant effect. Results are shown for each dataset, for every processing strategy and all 11 RSNs (10 dual regression + 1 seed-based).
Supplementary Figure 3: Percentage of voxels/seed-pairs exhibiting a significant group-level effect (threshold-free cluster enhanced, \(p<0.05\), FWE-corrected) after exploiting a categorical (left two images) and dimensional (right two images) approach on respectively the RSN spatial maps (upper images) and the seed-based Z-correlation maps (lower images). Results are shown for each dataset and for every processing strategy. The results for the RSN spatial maps comprise all 11 evaluated RSNs (10 dual regression + 1 seed-based) as described in the main text. Of note, threshold-free cluster enhancement was not applied to the correlation matrices.
Supplementary Figure 4 - Motion artifact removal and preservation of signal of interest in seed-based correlation functional connectivity analyses (264 ROIs). The upper triangles illustrate the correlation (Fisher r-to-z transformed) of each seed-pair correlation Z-scores with the motion summary score (RMS-FD). Depicted correlations were calculated across participants within each dataset. The boxplots in the lower-right panel of each dataset representation illustrate the distribution of these correlations for each processing strategy. The lower triangles in each panel illustrate the seed-pair Z-score correlation matrices averaged across participants within each dataset. Correlation Z-scores were transformed to percentiles to account for scaling differences between the different preprocessing procedures. This was done to enable assessing the correlation structure between the 264 seed ROIs rather than the strength of the obtained correlations. Results are shown for each dataset and all processing strategies.
Supplementary Figure 5 - RSN identifiability and RSN reproducibility for each of the included datasets. RSN identifiability is quantified using a Z-score ratio expressing the Z-scores within an RSN to Z-scores outside the RSN. RSN reproducibility scores reflect the spatial correlation between group-level RSN maps for random splits of the samples, normalized to pseudo Z-scores (see details in main text). Results are shown for each dataset, for every processing strategy and 11 RSNs (10 dual regression + 1 seed-based).
Supplementary Figure 6 - Sensitivity to signal of interest in seed-based correlation functional connectivity analyses (264 ROIs) investigated through detection of (anti-) correlated seed-pairs using mixture modeling. The lower triangles in each panel illustrate the seed-pair Z-score correlation matrices averaged across participants within each dataset. Correlation Z-scores were transformed to percentiles to account for scaling differences between the different preprocessing procedures. This was done to enable assessing the correlation structure between the 264 seed ROIs rather than the strength of the obtained correlations. Next we applied mixture modeling to each matrix to detect significant correlations. The upper triangles illustrate the additionally detected (red) or lost (blue) seed-pairs when compared to mixture modeling results on correlation matrices derived using the no-MC strategy. Results are shown for all processing strategies and each dataset, as well as across all participants from the four respective datasets. Supplementary Table 1 presents the fractions of seed-pair detected as correlated illustrated here.
Supplementary Figure 7 - Results of global signal regression (GSR) and band-pass filtering on group-level motion-effects, RSN identifiability and RSN reproducibility, within the NeuroIMAGE Controls dataset. The upper figure presents the significant group-level differences between participants exhibiting low and high amounts of head motion, for all tested RSN spatial maps and preprocessing strategies (threshold-free cluster enhanced, p<0.05, FWE-correction). The lower figures respectively present the results for RSN identifiability and reproducibility.
Chapter 3

Modelling categorical and dimensional mechanisms in mental disorders

In preparation as:
Abstract

Approaches for conceptualizing and investigating pathophysiology related to mental disorders are increasingly combining categorical and dimensional views. We discuss methodological challenges related to the typically close relationship (collinearity) of categorical and dimensional symptom measures and illustrate how current modelling approaches suffer from reduced sensitivity or interpretability. We describe a principled analysis framework that addresses these issues by first obtaining findings of interest with high sensitivity and subsequently disentangling categorical versus dimensional mechanisms using more stringent modelling. While these concepts are generically applicable for any standard linear modelling of behavioural or clinical variates, we here focus on imaging neuroscience data. As an example, we illustrate how to apply this framework to study functional brain architecture in attention-deficit/hyperactivity disorder using resting-state fMRI.
Trends box

- Understanding pathophysiology related to mental disorders should integrate categorical and dimensional views.
- Investigating the categorical-dimensional complexity associated with mental disorders is methodologically challenging due to the close relationship between categorical and dimensional symptom measures. Within a linear modelling framework this translates into collinearity between categorical and dimensional descriptor variables.
- Currently categorical and dimensional statistical models are typically evaluated separately or assessed within a single model including both types of variables. These approaches can suffer from decreased sensitivity or lack interpretability of the obtained results.

Glossary

**ADHD**: Attention-deficit/hyperactivity disorder; a neurodevelopmental disorder characterized by inattentive and/or hyperactive and impulsive behaviour

**Categorical mechanism**: A mechanism reflecting a systematic difference between multiple conditions

**Collinearity**: Property of a set of variables that indicates the linear dependency between the variables

**Dimensional mechanism**: A mechanism characterized as a continuum of features

**GLM**: General(ised) linear model (statistical linear regression model)

**Orthogonalization**: A mathematical procedure to transform two linearly dependent vectors into a set of linearly independent vectors which span the same space
The Diagnostic-Statistical Manual (DSM) is the most commonly used system to assess the presence of mental disorders (American Psychiatric Association, 2013). It characterizes a disorder using a cluster of behavioural symptoms that are determined by clinical methods including psychiatric interviews and questionnaires. This system is categorical by definition and assumes that there are discrete states where diagnosed individuals are systematically different from their healthy counterparts. However, it is increasingly appreciated that categorical behavioural measures (such as a diagnosis) often are a poor description of the underlying variates (i.e. pathophysiological mechanisms) (Hyman, 2007; Regier et al., 2009).

Accordingly, initiatives are arising to encourage inclusion of dimensional approaches for studying genetic, neural, and behavioural variates of mental disorders, and to identify bio-behavioural dimensions that cut across current diagnostic categories. Such approach was recently consolidated within the National Institute of Mental Health’s ‘Research Domain Criteria’ initiative (Insel et al., 2010; Morris and Cuthbert, 2012). However, the clinical utility of dimensional diagnostics is heavily debated since clinical decision making often comprises categorical decisions (e.g. start a specific intervention) (Coghill and Sonuga-Barke, 2012; Weinberger et al., 2015).

At the behavioural level most mental disorders seem to adhere to a dimensional structure (Coghill and Sonuga-Barke, 2012; Haslam et al., 2012). This, however, does not imply similar structures or mechanisms at other levels (e.g. genetic, neural or neurocognitive domain). An example is the case of attention-deficit/hyperactivity disorder (ADHD). ADHD has consistently been shown to adhere to a dimensional latent structure at the behavioural level (Haslam et al., 2006; Marcus and Barry, 2011; Asherson and Trzaskowski, 2015). However, neuroimaging studies that exploit categorical and dimensional measures to investigate neural variates of ADHD, suggested a hybrid categorical-dimensional view on its aetiology (Chabernaud et al., 2011; Elton et al., 2014). Therefore, in order to further our understanding of the pathophysiology of mental disorders we should not restrict research to either a categorical or dimensional implementation, but rather integrate categorical and dimensional conceptualisations and measurements, as common in most areas of medicine (Chabernaud et al., 2011; Rutter, 2011; Coghill and Sonuga-Barke, 2012).

Translating such categorical-dimensional views of mental disorders into statistical modelling approaches is methodologically challenging. This is due to the fact that we rely on categorical and dimensional behavioural measures as predictors.
in our statistical models. These measures are typically closely related (i.e. exhibit high statistical collinearity) as they represent alternative descriptors of the same behaviour. As a result, researchers typically resort to distinct statistical models to investigate categorical (e.g. DSM-diagnosis) and/or dimensional (e.g. impulsivity symptom scores) measures. Only few new studies are beginning to combine measures within single statistical models, while testing for unique contributions (Chabernaud et al., 2011; Elton et al., 2014, 2015).

Here, we will illustrate that using distinct statistical models reduces interpretability of the obtained findings, whereas using a single, unadjusted model reduces sensitivity to effects of interest. Accordingly, current analytical approaches for studying categorical and dimensional aspects of mental disorders are suboptimal and impede increasing efforts towards modelling the full pathophysiological complexity of mental disorders. Therefore, we propose an alternative methodological procedure which addresses these drawbacks. This procedure regards a principled statistical framework comprising of a single inferential step and subsequent characterization of the obtained findings using more stringent modelling. While this framework is generically applicable to any linear modelling of behavioural or clinical variates, we here focus on imaging neuroscience data and demonstrate its utility by investigating categorical and dimensional effects on functional neural connectivity in the context of ADHD.

**Modelling categorical-dimensional effects**

Typically, researchers investigate mental disorders using categorical designs where a patient group is compared to healthy controls. This approach implies a systematic difference between the two groups and aims to find categorical pathophysiological mechanisms related to the disorder. Conceptually, a categorical neurobiological mechanism can be viewed as reflecting an altered ‘state’ of the brain. This alteration can be caused by specific high-impact factors, i.e. pathological causes such as high penetrating genetic variants, extreme environmental influences as very premature birth, or traumatic disturbing life events. However, there are few if any empirical data supporting the concept of a ‘phase transition’ model of mental disorders characterized by an abrupt turning point between normal and abnormal behaviour and/or associated neurobiological mechanisms.

Alternatively, dimensional neurobiological mechanisms relate to accumulation of continuous underlying factors such as polygenic risks, or continued exposure to adverse situations (e.g. low socio-economic-status), which represent themselves
through e.g. decreasing levels of neurocognitive functioning. Although evident at the behavioural level, these accumulating factors are not taken into account in the diagnostic process. In contrast, current diagnostic categories are social constructs where symptom thresholds have been established by consensus, even in absence of sharp turning points. As a consequence, at the research end, investigators focus on genetic, cognitive or neural correlates of either a categorical DSM-diagnosis or a dimensional symptom severity score (Whelan et al., 2012). However, a full understanding of the pathophysiology of mental disorders requires an integrated analysis of categorical and dimensional effects (Coghill and Sonuga-Barke, 2012). Such an analysis should allow for examining, for example, whether the neural mechanisms underlying impulsive behaviour are similar across different categorical disorders such as ADHD, borderline personality disorders, etcetera.

**Limitations of current modelling approaches**

Figure 1 visualizes the variance of two variables of interest within a statistical general linear model (GLM, e.g. linear regression). This example includes a categorical and a corresponding dimensional measure. Since both measures are highly correlated (i.e. exhibit strong collinearity) the amount of variance they explain in the model overlaps. Specifically, the amount of overlap is proportional to the amount of shared variance between the two measures. Accordingly, the variance that each measure explains is partitioned into two portions: a non-overlapping portion corresponding to the variance uniquely explained by each measure, and a portion overlapping with the second measure representing the shared variance.

When investigating such related measures using separate models, each model evaluates the full variance of each measure. As a consequence, the results obtained using independent models have limited interpretability as the estimated effect for each measure is not unique and in fact might be driven by shared variance with the second measure. A further complication is the possibility that (collinear) measures can relate to opposite (i.e. negative versus positive) associations, which will additionally reduce sensitivity of the model.

Alternatively, when inferring upon two related measures jointly by including them both in a single GLM, the shared variance is ignored within the model estimation and only the unique variance associated with both measures is assessed (York, 2012; Mumford et al., 2015). Ensuing significant findings have high interpretability, as they are uniquely associated with the explanatory variable. However, a GLM including highly collinear variables will be less efficient as the shared variance will be disregarded and the variance associated with the parameter estimates will increase (York, 2012; Mumford et al., 2015). As such, a joint model
Categorical-dimensional modelling decreases the sensitivity and reliability of the analysis, increasing the rate of false negatives (i.e. type II error). Moreover, this model is conceptually suboptimal: the shared variance reflects the full behavioural phenotype and is therefore in fact most characteristic to the disorder, yet it is fully ignored in this model.

Selective orthogonalization

To counter the limitations specified above, GLMs can be adapted to specifically test the unique and shared variance of predictors. This is achieved through ‘orthogonalization’ of the predictors, allowing to assign shared variance between two predictors to either one of them (Mumford et al., 2015). Accordingly, by means of selective orthogonalization of predictors within a model one can evaluate different model variants to infer upon different portions of the total variance thus targeting different research questions. In practice, orthogonalization of predictor A with respect to predictor B removes all shared variance from predictor A, leaving only its residuals (which represent predictor A’s unique variance). This is achieved by regressing predictor B against predictor A (typically using an ordinary least squares regression) and subtracting appropriate portions of B from A. Note that the total variance explained by the model is not affected; orthogonalization only reallocates
variance within the model and therefore only effects the interpretation of the predictors.

Consider an example based on Figure 1. In this example, we evaluate a model including two predictors: one categorical and one dimensional measure. When we orthogonalize the categorical measure with respect to the dimensional measure (i.e. removing the shared variance from the categorical measure) the model infers upon the unique variance of the categorical measure and the full variance of the dimensional measure, while still modelling the full variance of the original model.

To further illustrate the utility of this approach we consider the case of ADHD which is generally characterized by two symptom domains: inattentive and hyperactive/impulsive symptoms. Accordingly, we can model the disorder using one categorical measure (i.e. the DSM-diagnosis) and two dimensional measures (i.e. the two symptom scores). Using selective orthogonalization we can then obtain six model variants, all visualized in Figure 2, which can be employed to answer different research questions.

- Models 1-3: These models test for the full variance associated with each measure. To obtain these models we, for each measure respectively, orthogonalized the two measures of no interest with the measure of interest. These models are closest to models that would evaluate a single categorical or dimensional measure in isolation, yet are potentially more sensitive due to decreased residual variance by including all variables into a single model. Models 1-3 are sensitive to effects of interest. However, as the full variance is modelled, the estimated effect of each measure is likely not uniquely associated with that measure only. The obtained findings are therefore non-specific and less interpretable (i.e. the effect might be driven by shared variance with the alternative measures). Yet, these models yield the highest sensitivity.

- Models 4-6: These models test for the unique variance of the categorical and dimensional measures. We obtained these models, for each of the variables respectively, by orthogonalizing the measure of interest with respect to the other two alternative measures. As previously noted, evaluating the three measures jointly within a single (non-orthogonalized) model will also only evaluate the unique variance. The only difference in testing for the unique variance using multiple orthogonalized models versus a single non-orthogonalized model is that the variance associated with the parameter estimates (i.e. reliability) will differ due to differences in collinearity of the model. Since the unique variances of the predictors only capture a minor portion of their full variance, these models are less sensitive to the effects of interest but produce very interpretable results compared to models 1-3.
These models are therefore suited for specifically disentangling categorical and dimensional effects rather than detecting any effect of interest associated with the disorder in the first place.

**Alternative framework**

To address the drawbacks of limited interpretability or sensitivity in currently employed analytical approaches we propose a statistical framework which utilizes orthogonalization (as shown in the models provided above) and comprises two subsequent steps.

First, effects of interest associated with the disorder are obtained with high sensitivity by evaluating those models that do not parcel out shared variance in the GLM estimation (e.g. models 1-3). Subsequently, the obtained significant effects are characterized more specifically by evaluating models that test for the association
with the unique variance of the predictors (e.g. models 4-6). As an example, the first step can entail the evaluation of model variants 1-3 in Figure 2. Alternatively, one could perform an F-test across the full non-orthogonalized model (Figure 2, top diagram). Subsequently, statistically significant findings (e.g. clusters of voxels, cognitive metrics, genetic variants etc.) are characterized by investigating the association of this effect with the unique variance of the predictors using models 4-6. Instead of evaluating models 4-6 separately, one can evaluate the three predictors in a single non-orthogonalized model which will also test for the unique variance of each predictor, at the cost of slightly increased parameter estimate variance due to collinearity of the model.

Note that, in practice, a categorical measure represents a simplification of a dimensional measure, which complicates the interpretation of categorical measures. Furthermore, it is important to realize that although this framework utilizes multiple regression models, it does not require additional multiple comparison correction compared to traditional analyses nor causes problems with double-dipping (Kriegeskorte et al., 2009). All six models explain the same total variance and are therefore equivalent with respect to signal versus noise considerations. Finally, models 4-6 employed in stage 2 are only performed on statistically significant results already obtained in stage 1 and used for interpretation only.

Example: application in ADHD

As an illustration of the principle, we conducted a study investigating functional connectivity in the context of ADHD. We aimed to disentangle effects related to 1) the categorical ADHD-diagnosis, 2) dimensional effects of inattentive symptoms, and 3) dimensional effects of hyperactivity/impulsivity symptoms. The correlation between these variables ranges from 0.70 to 0.76, indicating high collinearity. We investigated resting-state MRI scans from control (n=136, 48% male, age=17±3 years) and ADHD (n=179, 78% male, mean age=18±3 years) participants from the NeuroIMAGE project (von Rhein et al., 2015), which is a Dutch follow-up study of the International Multicenter ADHD Genetics (IMAGE) study (Rommelse et al., 2008; Nijmeijer et al., 2009; Müller et al., 2011a, 2011b). Since ADHD is associated with executive dysfunction (Barkley, 1997) we specifically investigated functional connectivity within a network involved in executive functioning which includes the anterior cingulate cortex (ACC), paracingulate and areas within the prefrontal cortex (PFC) (Beckmann et al., 2005). We refer to this network as the ‘executive control network’ (ECN) (Beckmann et al., 2005; Smith et al., 2009), but it is sometimes also referred to as the ‘salience network’ (Seeley et al., 2007). For each
Categorical-dimensional modelling

To illustrate the limitations of current modelling approaches and show how results can vary across different model set-ups we first evaluated all six models illustrated in Figure 2. We conducted non-parametric permutation testing \((n=5000)\), applied threshold-free cluster enhancement, corrected for family-wise error \((p<0.05)\) and included covariates for age, sex, and scanning site. As noted before, the results of models 1-3 are close to results that would have been obtained when using distinct categorical and/or dimensional models, whereas models 4-6 evaluate the specific association one would have obtained when evaluating the three measures using a single, joint model.

Figure 3 illustrates the results of this analysis. Although the various models generate comparable results there are significant differences. Models 1-3 result in a large significant effect related to ADHD-diagnosis and scores of inattention localised to PFC and ACC. The large extent of this cluster indicates the high sensitivity of these models. However, model 1-3 do not show effects related to

Figure 3 Significant results of testing for the full and unique variance of one categorical and two dimensional ADHD-related measures on functional connectivity within the executive control network (ECN). Specifically, this figure presents the results of evaluating models 1-6 illustrated in Figure 2. The Venn diagrams visualize the model set-up, in which the inferred upon variance is outlined in black. To obtain these results we applied the models using non-parametric permutation testing \((n=5000)\) on subject-level spatial maps of the ECN across 136 control and 179 ADHD subjects (threshold-free cluster enhanced, family-wise error corrected, \(p<0.05\)). Yellow: a significant positive effect, blue: a significant negative effect.
hyperactivity/impulsivity. Models 4-6 also result in an effect specific to the ADHD-diagnosis and measures of inattention. However, the clusters sizes reduced from 606 and 492 significant voxels to 21 and 57, indicating that the effects seen in model 1-3 are not unique to diagnosis or inattentiveness and/or that these models have decreased explanatory power. Moreover, a (large) significant negative effect of hyperactivity/impulsivity is now observed (model 6). This effect was not observed using model 3, most probably due to the high shared variance with the ADHD-diagnosis and inattentive measures with opposite direction. Evaluating models 4-6 therefore results in a considerably different interpretation compared to the findings of models 1-3, highlighting how results depend on how a GLM is designed.

Note that this is a single example to illustrate how sensitivity and interpretability can differ across different model designs. Next to an expected decreased effect of ADHD-diagnosis and inattention between models 1/4 and models 2/5, we observed an increased effect of hyperactivity between model 3/6. It is therefore most likely that functional connectivity within the ACC/PFC of the ECN has a complex relation with behaviour, possibly including opposite relationships with inattention and hyperactivity/impulsivity, which might be missed when not specifically testing for the unique associations of each measure.

**Alternative framework**

Next, we conducted our proposed two-stage framework. First we obtained effect(s) of interest by running a statistical test that incorporates the shared variance between the measures. To that end we defined a GLM including all three predictors as well as the previously described covariates, and ran an $F$-test across the predictors. As illustrated in Figure 4, the $F$-test yielded a large significant spatial cluster spanning ACC and PFC (peak voxel: MNI $x=9$, $y=57$, $z=6$; $F=13.7$, $p<0.001$). Next, we aimed to characterize this effect more specifically by testing for its association with the unique variance of our three predictors. To that end we first obtained a single connectivity score of this cluster for every participant by calculating the mean connectivity value across the voxels of the spatial cluster within every participant’s spatial map of the ECN. Using this score as a dependent variable we now evaluated models 4-6 to test for the specific association of this cluster with ADHD-diagnosis, inattentive and/or hyperactivity/impulsivity symptom scores. This analysis revealed that the cluster was associated with all three predictors, although the directionality of the results differed: While both an ADHD-diagnosis ($T=3.2$), and inattentive symptoms ($T=3.6$) were positively associated with ACC-ECN functional connectivity, an increase in hyperactivity/impulsivity symptoms was now associated with decreased ACC-ECN connectivity ($T=-4.3$; see Figure 4), reflecting opposite associations of inattention
and hyperactivity/impulsivity with ECN functional connectivity of the region which gets ‘masked’ when evaluating a highly collinear model.

Comparing these results with the results obtained in the previous section we can appreciate that this framework resulted in detecting a cluster as extensive as the clusters obtained using models 1-3, i.e. suggesting high sensitivity, while providing an interpretation of these effects as observed in models 4-6, i.e. accurate interpretation. The results of a more extensive analysis across multiple networks using our proposed analytical framework is described elsewhere (Pruim et al., 2017).

Concluding remarks

Research is moving towards integrated categorical-dimensional approaches for conceptualizing and investigating mental disorders (Chabernaud et al., 2011; Coghill and Sonuga-Barke, 2012). We discussed methodological challenges
associated with such analyses, due to statistical collinearity between categorical and dimensional measures, and explained limitations of currently employed approaches that aim to investigate these aspects. We proposed a strategy that aims to overcome these limitations by utilizing selective orthogonalization of measures after identifying effects of interest with high sensitivity. Key in this whole process is that researchers need to selectively analyse the portion of the variance inferred upon in their specific model.

In our current work, we focused on behavioural categorical and dimensional features. Yet, the discussed limitations and proposed framework extend to all kinds of research questions aimed at evaluating collinear features. For instance, investigating interaction between measures, evaluating a broad set of dimensional measures (e.g. multiple cognitive metrics), deciding between categorical and dimensional assessment of a phenotype (e.g. by testing the unique contribution of dimensional descriptors over and above categorical labels), investigating comorbidity of disorders (i.e. multiple categorical measures) or when moving to other domains (e.g. genetics by comparing a single genetic variant with respect to a polygenetic score). The concepts as explained in our current work therefore extend to a broad range of analyses conducted to investigate mental disorders, e.g. studying bio-behavioural dimensions in light of the ‘Research Domain Criteria’ initiative (Insel et al., 2010; Morris and Cuthbert, 2012). They are essential for optimal modelling and accurate interpretation of obtained results in research aiming to disentangle the complexity of the pathophysiology of mental disorders.

**Outstanding questions box**

- How to investigate a wider range of collinear clinical measures to more specifically disentangle mechanisms underlying a mental disorder? For instance, by evaluating a range of cognitive variables or investigating an extensive set of symptoms (e.g. 18 ADHD-related symptoms as described in the DSM) rather than limiting research to summary scores of behavioural domains. Such approaches will further increase collinearity of evaluated models and challenge the interpretation of obtained findings.

- How to take into account the absence of sharp turning points for delineating categorical disorders? One approach might be to perform sensitivity analyses of various symptom severity cut-off points for defining categories.
• How to detect (latent) categorical-dimensional mechanisms in a data-driven, bottom-up analysis? Most research relies on using phenotypic measures. It is questionable whether categorical mechanisms at the pathophysiological level align with behaviourally defined boundaries (e.g. DSM-diagnosis). Similarly, definitions and dimensional quantifications of behaviour do not necessarily align with dimensional pathophysiological processes. Data driven approaches are emerging to investigate such latent categorical or dimensional mechanisms (e.g. normative modelling (Marquand et al., 2016)).

• How should the concept of equifinality be incorporated within modelling procedures? Equifinality denotes that multiple independent pathophysiological pathways can ultimately lead to similar phenotypic presentations. As such, heterogeneity within subjects diagnosed with mental disorders is present at the phenotypic as well as at the pathophysiological level. Investigating the pathophysiology of mental disorders using a predictor from a specific domain (e.g. phenotypic, cognitive) might therefore miss underlying downstream pathophysiological heterogeneity.

• How should dimensional research findings be translated to clinical usage? Medicine generally uses dimensional scores to quantify severity of a disease, yet relies on categorical decisions (e.g. decision to treat or not). How research findings on categorical-dimensional underpinnings of mental disorders should be translated to clinical usage for diagnostics, prognostics, and individualized approaches to treatment remains an open question.

Acknowledgement

This work was supported by NWO Large Investment grant 1750102007010 (Dr Buitelaar), ZonMW Addiction: Risk Behaviour and Dependency Grant 60-60600-97-193 (Dr Buitelaar), NWO Brain & Cognition: an Integrative Approach grant 433-09-242 (Dr Buitelaar), NWO National Initiative Brain & Cognition 056-13-015 (Dr Buitelaar), the EU FP7 grants TACTICS (278948), IMAGEMEND (602450), MATRICS (603016) and AGGRESSOTYPE (602805), EU IMI grant EU-AIMS (115300), and grants from Radboudumc, University Medical Center Groningen, Accare, and VU University Amsterdam. Dr. Mennes has received funding from the European Research Council under the European Union's Seventh Framework
Chapter 3

Programme (FP7/2007-2013) grant agreement 327340 (BRAIN FINGERPRINT). Dr. Beckmann is supported by the Netherlands Organisation for Scientific Research (NWO-Vidi 864-12-003) and we gratefully acknowledge the funding from the Wellcome Trust UK Strategic Award [098369/Z/12/Z]. We acknowledge the Department of Pediatrics of the VU University Medical Center for the opportunity to use the mock scanner for preparation of our participants.

References


Supplementary materials

Participants

Participants were selected from a follow-up (2009-2012) of the Dutch part of the International Multicenter ADHD Genetics (IMAGE) study, performed between 2003-2006 (Rommelse et al., 2008; Nijmeijer et al., 2009; Müller et al., 2011a, 2011b). At first enrolment, 365 families with at least one child with combined subtype ADHD and at least one biological sibling (regardless of ADHD diagnosis) were recruited, in addition to 148 control families with at least one child, with no formal or suspected ADHD diagnosis in any of the first-degree family members. Recruitment for ADHD families was accomplished through ADHD probands attending outpatient clinics in the regions Amsterdam, Groningen, and Nijmegen (the Netherlands), as well as a VU University affiliated ADHD research institute. Control families were recruited through primary and high schools in the same geographical regions as the participating ADHD families. All family members, also those who did not participate in IMAGE, were invited for follow-up measurement with a mean follow-up period of 5.9 years (SD = .74). Follow-up rates were 78.4% for ADHD families and 80.4% for control families. In order to balance out the distribution of gender and age between the ADHD and healthy control groups, additional girls with ADHD (any subtype; N=37 families) and healthy control boys (N=34 families) were recruited for NeuroIMAGE. Inclusion criteria were the same for all participants, and largely consistent with IMAGE: participants had to be between 5-30 years, of European Caucasian descent, have an IQ ≥ 70, and no diagnosis of autism, epilepsy, general learning difficulties, brain disorders and known genetic disorders (such as Fragile X syndrome or Down syndrome). Since the data used in the current study was administered as part of an MRI protocol, participants were excluded if they were younger than 8 years or had any contraindication to MRI scanning (e.g. implanted metal or medical devices, or possible pregnancy). Including the newly recruited families, the complete NeuroIMAGE cohort comprised 323 ADHD families and 153 control families.

Within this project, diagnoses of ADHD and comorbid disorders was assessed by a trained professional using the Schedule for Affective Disorders and Schizophrenia for School-Age Children - Present and Lifetime Version (K-SADS)(Kaufman et al., 1997) and the Conners’ ADHD questionnaires (Conners et al., 1998). The Conners’ ADHD questionnaire furthermore provided summarized DSM-scores of both inattention (CPRS-I) and hyperactivity/impulsivity (CPRS-H) symptoms.
For our current study we only included participants of whom both a resting-state fMRI and structural MRI scan was available were considered (n=545). Ninety participants were excluded due to incomplete scanning (n=9), poor field-of-view coverage (n=23), incidental findings (n = 10), official exclusion criteria been put forward after inclusion (n=12), incomplete CPRS-questionnaire (n=16), and/or because they belonged to the 5% highest movers as determined by a motion summary score (i.e. motion outliers; n=27). We used the root mean squared of the frame-wise displacement time-series as a participant-level summary motion score (RMS-FD)(Jenkinson et al., 2002). Furthermore, we excluded healthy siblings of ADHD-diagnosed participants as well as participants with remitted ADHD (i.e. participants which received an ADHD diagnosis within the IMAGE study but not in the current follow-up NeuroIMAGE study).

Procedure

The current study was part of a comprehensive assessment protocol encompassing behavioural questionnaires, a diagnostic interview and several neurocognitive measures from all family members, and an extensive MRI scanning protocol in participating children. From participants whose genotypic information was missing during IMAGE, saliva was collected for DNA analysis. Testing was carried out either at the VU University Amsterdam and VU University Medical Centre, or at the Radboud University Nijmegen Medical Centre and Donders Institute for Brain, Cognition and Behaviour in Nijmegen. Participants were asked to withhold use of psychoactive drugs for 48 hours before measurement. During the testing day, participants were motivated with short breaks, and at the end of the day, children received a reward of €50,- and a copy of their MRI scan. Informed consent was signed by all participants (parents signed informed consent for participants under 12 years of age), and the study was approved by the ethical committee (Centrale Commissie Mensgebonden Onderzoek).

MRI acquisition

Data were acquired at two scanning locations on similar 1.5 Tesla Siemens scanners (Siemens Sonata at VU University Medical Centre in Amsterdam; Siemens Avanto at Donders Centre for Cognitive Neuroimaging in Nijmegen) using the same Siemens 8-channel head coil and identical scanning protocols. Anatomical images were obtained using an MPRAGE sequence (TR=2730 ms, TE=2.95 ms, T1=1000ms, flip angle=7, matrix size=256x256, FOV=256mm, 176 slices with 1mm isotropic voxels). Functional images during rest were obtained using a gradient echo echo-planar imaging (GE-EPI)
sequence (TR=1960 ms, TE=40 ms, FOV=224mm, 37 axial slices, flip angle=80, matrix size=64x64, in-plane resolution =3.5mm, slice thickness/gap=3.0mm/0.5mm). Participants were instructed to relax with their eyes open during the rfMRI scan.

**I**

**RF**MRI data preprocessing

Preprocessing of rfMRI data was carried out using tools from the FMRIB Software Library (FSL; [http://www.fmrib.ox.ac.uk/fsl](http://www.fmrib.ox.ac.uk/fsl)) (Smith et al., 2004; Woolrich et al., 2009; Jenkinson et al., 2012) and involved removal of the first five volumes to allow for signal equilibration, head movement correction by volume-realignment to the middle volume using MCFLIRT, global 4D mean intensity normalization, and spatial smoothing (6mm FWHM). Subsequently, we applied ICA-AROMA which concerns a data-driven strategy that effectively removes residual motion artifacts from fMRI data (Pruim et al., 2015a, 2015b). We completed preprocessing by nuisance regression, using mean white matter and CSF time-courses as well as linear trend as nuisance regressors, and temporal high-pass filtering (>0.01Hz).

For every participant we then transformed the preprocessed fMRI data to the participants’ structural image using a single pre-calculated affine boundary-based registration (FLIRT) (Jenkinson and Smith, 2001; Jenkinson et al., 2002). Subsequently, we registered the functional data to our NeuroIMAGE study template in MNI152 space (3mm isotropic resolution) using non-linear registration as implemented in FSL FNIRT (Andersson et al., 2007).

**Dual regression**

For each participant we derived spatial maps of well-validated and commonly replicated resting-state networks (RSNs) using dual regression (Beckmann et al., 2009; Filippini et al., 2009). This analysis implements a multivariate spatial regression of a set of initial templates against the preprocessed fMRI data of every participant, and yields participant-specific time series for each template. Next, these time series are entered in a multivariate temporal regression against the same preprocessed data resulting in participant-level spatial representations (parameter estimate maps) of the initial templates. As templates, we used the 20 spatial maps as described by (Smith et al., 2009). These templates include 10 RSNs which were found to correspond well to networks involved in task-related processing. The other 10 templates are artifactual or represent more complex networks. Dual regression was performed using the full set of components to optimally model the data. We selected the RSN representing the executive control network for our subsequent analyses.
Chapter 3

References


Categorical-dimensional modelling


Chapter 4

Bio-behavioural network correlates of categorical and dimensional attention-deficit/hyperactivity disorder

Under review as:
*Shared last authorship
Abstract

Attention-deficit/hyperactivity disorder (ADHD) is a heterogeneous neurodevelopmental disorder, putatively induced by dissociable dysfunctional biobehavioural pathways. Here, we aim to parse ADHD-related heterogeneity in its underlying neurobiology by investigating functional connectivity across multiple brain networks to 1) disentangle categorical diagnosis-related effects from dimensional behaviour-related effects, and 2) functionally map these neural correlates to genetic and neurocognitive measures. We identified functional connectivity abnormalities related to ADHD across 14 networks within a large resting-state fMRI dataset (n=409, age=17.5±3.3 years). We tested these abnormalities for their association with the categorical ADHD diagnosis, and with dimensional inattention and hyperactivity/impulsivity scores. Next, we evaluated the relationship of these findings with neurocognitive measures (working memory, response inhibition, reaction time variability (RTV), reward sensitivity) and dopamine neurotransmission-related genetic variants (in genes DAT1 and DRD4). Within the default mode network, we mainly observed categorical ADHD-related functional connectivity abnormalities, unrelated to genetic and neurocognitive measures. Clusters within the visual networks primarily related to dimensional scores of inattention and RTV, while findings within the sensorimotor networks were mainly linked to hyperactivity/impulsivity, and both reward sensitivity and working memory. Findings within cerebellum network and executive control network (ECN) related to both categorical and dimensional ADHD measures and were linked to response inhibition and RTV. Findings within the ECN moreover related to genetic variants in DAT1 and DRD4. This explorative study identified ADHD-related neural correlates across multiple functional networks, showing distinct categorical and dimensional mechanisms and their links to neurocognitive functioning and genetics.
Network correlates of ADHD

Introduction

Inter-individual differences are a hallmark of neurodevelopmental disorders such as Attention-deficit/hyperactivity disorder (ADHD) (Faraone et al., 2015). In ADHD, heterogeneity between diagnosed individuals is partly believed to originate from dissociable cognitive deficits and neural mechanisms (Coghill et al., 2014; Faraone et al., 2015). As an example, several large-scale brain networks have been associated with ADHD, comprising localized networks including visual and motor cortices as well as networks distributed across association cortex (e.g. the default mode and executive control network)(Castellanos and Proal, 2012; Mostert et al., 2016).

Most studies investigating pathophysiological mechanisms of ADHD rely on case-control study designs, testing for systematic (i.e. categorical) differences between cases and controls. However, there is increasing evidence that ADHD can also be understood as an ‘extreme’ on a continuum of typical functioning (i.e. dimensional attentive and hyperactive/impulsive traits)(Levy et al., 1997; Haslam et al., 2006, 2012; Marcus and Barry, 2011; Marcus et al., 2012; Asherson and Trzaskowski, 2015). Accordingly, new initiatives, such as RDoC (Insel et al., 2010; Morris and Cuthbert, 2012), seek to employ dimensional approaches to study the behavioural, neural, and genetic features of mental disorders. Interestingly, recent results endorse that the pathophysiology of ADHD is conceptualized by a complex interplay between categorical and dimensional mechanisms (Chabernaud et al., 2011; Elton et al., 2014; van Ewijk et al., 2014; Wu et al., 2017).

Accordingly, an advanced understanding of ADHD needs to address that pathophysiological mechanisms can be 1) categorical and/or dimensional, and 2) distributed across different brain networks. Here, we present a proof-of-concept study that documents categorical and dimensional effects of ADHD on functional brain architecture by evaluating multiple brain networks as derived from resting state functional magnetic resonance imaging (rfMRI) data.

To understand the interplay of categorical and dimensional aspects of ADHD, we need to integrate these aspects in one analysis. This poses methodological challenges due to the close relationship between categorical and dimensional measures, leading to highly collinear statistical models that suffer from decreased sensitivity (i.e. increased false negative results)(York, 2012; Mumford et al., 2015; Pruim et al., 2017). To address these drawbacks we employed an analytical framework that allows the identification of a set of biomarkers related to ADHD and subsequently characterizing these effects in terms of the distinct contribution of categorical and dimensional mechanisms (Pruim et al., 2017). In particular, we
distinguished between effects related to the categorical ADHD-diagnosis and those related to dimensional scores of inattentive and hyperactive/impulsive behaviour.

Dissociable bio-behavioral pathways underpinning ADHD will not be restricted to the neural and behavioral domain and can be expected to also relate to genetic variants and neurocognitive dysfunction. To further explore such bio-behavioural pathways, we related our brain network-findings to genetic and neurocognitive measures. We focused on genetic variants (variable number tandem repeats; VNTRs) in two genes related to dopaminergic neurotransmission (DAT1 coding for the dopamine transporter and DRD4 encoding the dopamine receptor D4) that have been implicated in ADHD and have survived in meta-analyses (Faraone et al., 2014; Gatt et al., 2015). At the neurocognitive level, we investigated core neurocognitive features in line with proposed dissociable neurocognitive pathways implicated in ADHD (Durston et al., 2011), including reward processing, reaction time variability, response inhibition, and working memory.

Materials/methods

Participants

We included participants from the NeuroIMAGE cohort (von Rhein et al., 2015b), which is a Dutch follow-up study of the International Multicenter ADHD Genetics (IMAGE) study (N. N J Rommelse et al., 2008; Nijmeijer et al., 2009; Müller et al., 2011a, 2011b). In NeuroIMAGE (on average 6 years after IMAGE), diagnostic, cognitive, MRI, and genetic data was acquired from ADHD and control participants as well as their siblings. Informed consent was signed by all participants (parents signed informed consent for participants under 12 years of age), and the study was approved by the ethical committee (Centrale Commissie Mensgebonden Onderzoek). Diagnosis of ADHD and comorbid disorders was assessed by a semi-structured diagnostic interview and Conners’ questionnaires (Conners et al., 1998b, 1999). For the interview, we used the Dutch version of the Parental Account of Children’s Symptoms (PACS)(Taylor et al., 1986; Taylor, 1991) in IMAGE and the Schedule for Affective Disorders and Schizophrenia for School-Age Children - Present and Lifetime Version (K-SADS) in NeuroIMAGE (Kaufman et al., 1997). We used the phenotypic information available across IMAGE and NeuroIMAGE to categorize all

Table 1 (next page) – Participant characteristics and association of these characteristics with diagnostic group and symptom scores. Significant between-group differences and association with symptom scores were assessed using analysis of variance (F-test), Chi-squared statistical test, or Pearson correlation (r). The 46 TDC participants used for group-ICA were not included in these statistical tests as they were excluded from all categorical/dimensional analyses presented in this manuscript (i.e. they are not part of the TDC diagnostic group
### Network correlates of ADHD

<table>
<thead>
<tr>
<th>Group</th>
<th>ICA</th>
<th>Statistical test</th>
</tr>
</thead>
<tbody>
<tr>
<td>TDC (n=46)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>TDC (n=90)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ADHD (n=179)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sibling (n=109)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Remitted ADHD (n=31)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

#### Symptom Scores

<table>
<thead>
<tr>
<th>Symptom</th>
<th>Total</th>
<th>Mean</th>
<th>SD</th>
<th>Mean</th>
<th>SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Inattentive symptoms</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hyperactive/impulsive symptoms</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

#### Neurocognitive measures

<table>
<thead>
<tr>
<th>Measure</th>
<th>Total</th>
<th>Mean</th>
<th>SD</th>
<th>Mean</th>
<th>SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Response inhibition (ms)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Working memory (%)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Reward sensitivity (ms)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Reaction timing variability</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

#### Genetic variants

<table>
<thead>
<tr>
<th>Variant</th>
<th>Total</th>
<th>Mean</th>
<th>SD</th>
<th>Mean</th>
<th>SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>DAT1 - 3UTR; 10R homozygote</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DRD4; long allele</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

#### Statistical analysis

- **p < 0.05, ** **p < 0.01**

---

1. Number of days (converted to years) on which any type of stimulants were prescribed. Obtained using the Conners Parent ADHD questionnaire (Conners et al., 1998a), standardized T-score, range 40–90.

SD: standard deviation, RMS-FD: root mean-squared framed displacement (head motion summary score) (Jenkinson et al., 2002), ODD-CD: oppositional defiant disorder – conduct disorder, SES: socio-economic status (see supplementary material), SD: standard deviation, TDC: total diagnostic categories, rem: remitted ADHD.
participants into four diagnostic groups: typically developing controls (TDC), ADHD (meeting criteria for ADHD at the two time points), unaffected siblings (Sibling), or remitted ADHD (rem-ADHD; i.e., an ADHD diagnosis at the time of IMAGE, but not in NeuroIMAGE). Moreover, Conners’ Parent ADHD questionnaires provided dimensional DSM-IV scores of both inattention and hyperactivity/impulsivity symptoms (Conners et al., 1998a). Importantly, we only included siblings and remitted ADHD participants to increase power and reduce dichotomy for dimensional analyses and defined them as separate groups (i.e. not assign them to TDC or ADHD) to avoid pollution of any categorical TDC versus ADHD effects in upcoming analyses (see below). See Supplementary material for a detailed description of the participants and study procedures.

Resting-state fMRI data

We selected participants from NeuroIMAGE who completed both a resting-state fMRI (rfMRI) and a structural MRI scan. After applying exclusion criteria (see Supplementary material) 455 participants were left, of which 46 TDCs were randomly selected to define template resting state networks (see below); the remaining 409 participants were used on our main analyses. The rfMRI data was preprocessed using a typical preprocessing pipeline, complemented with ICA-AROMA, an advanced strategy for identifying and removing residual motion artifacts from fMRI data (Pruim et al., 2015a, 2015b). All individual-level rfMRI images were normalized to a study-specific anatomical template in MNI152 standard space (3 mm isotropic). See Supplementary material for a detailed description of the MRI acquisition, participant exclusion, and rfMRI preprocessing.

Deriving functional brain networks

We used the preprocessed rfMRI data to investigate functional connectivity measures related to a set of functional brain networks (i.e. resting-state networks). In a first step, we derived these networks through independent components analysis (ICA) with MELODIC as implemented in FSL (version v5.0.6.) (Beckmann and Smith, 2004; Smith et al., 2004; Beckmann et al., 2009; Woolrich et al., 2009; Jenkinson et al., 2012). To this end, we applied group-ICA with automatic dimensionality estimation to temporally concatenated rfMRI data of 46 randomly selected TDC participants. The participants used to derive these networks were excluded from further analyses. All subsequent functional connectivity analyses, on the remaining 409 participants, used participant-level spatial representations of the networks identified by the group-ICA, using a multivariate regression strategy called dual regression (Beckmann et al., 2009; Filippini et al., 2009). These
participant-level maps represent the connectivity strength of every voxel with the network.

Categorical and dimensional functional connectivity analysis

To evaluate categorical and dimensional effects of ADHD on network connectivity, we used the obtained participant-specific spatial maps for each network as dependent variables and defined a standard, ordinary least squares (OLS) regression model (see Figure 1) including the following predictors: a main effect for the four diagnostic subgroups (TDC, ADHD, Siblings, rem-ADHD) and two additional regressors for the symptom scores. The model was completed with covariates for age, sex, and scan site. The categorical effect of an ADHD-diagnosis was assessed by evaluating the contrast between the main effects of the TDC and ADHD participants (i.e. ignoring siblings and rem-ADHD participants). We tested dimensional effects by respectively evaluating scores of inattention and hyperactivity/impulsivity across all 409 participants.
Importantly, the high association (i.e. high statistical collinearity) between the categorical and dimensional predictors (see Table 1) raised methodological challenges. Simply evaluating the three predictors within the single regression model as defined above (as for instance done by Chabernaud et al. (2012)(Chabernaud et al., 2011) and Elton et al. (2014)(Elton et al., 2014)) would have reduced sensitivity to any effect of interest, as the large portion of shared variance between the predictors would have been disregarded within the regression procedure(Mumford et al., 2015). In contrast, evaluating the three effects within separate models could have yielded findings that are in reality driven by shared variance with a non-modelled variable (i.e., by the overlapping areas, as illustrated in the variance visualization of the regression model in Figure 1), thus reducing the interpretability of the findings. To accommodate these concerns, we utilized a two-stage framework, in which different variations of the defined regression model were tested(Pruim et al., 2017). In short, in Step 1, we identified effects of interest by testing for the full variance of every single predictor separately, i.e. we tested the effect of a predictor, unadjusted for the other two predictors. Subsequently, in Step 2, we characterized these obtained effects of interest by testing for the unique contribution of every single predictor; i.e. test for the effect of every single predictor, adjusted for the other two predictors.

**Step 1: Identifying findings of interest (full variance modelling)**

First we evaluated three regression models that respectively modelled the full variance of each of our three predictors (i.e., ADHD-diagnosis, inattention score, and hyperactivity/impulsivity score). For each evaluated predictor, we obtained a full variance model by regressing the tested predictor from the other two predictors, and subsequently replacing these other predictors in the model with their respective residuals. As a result, the tested predictor was assigned all the variance in the model that was originally shared with the other two predictors. See Figure 1 for a visualization of the specific variance tested by the three different models. Every model was evaluated using non-parametric permutation testing (n=5000) as implemented in FSL randomise, applying threshold-free cluster enhancement, correcting for family-wise error (p<0.05), and only considering spatial clusters with a minimum cluster size of 8 voxels(Smith and Nichols, 2009; Winkler et al., 2014). This yielded a set of spatial clusters (n=27), which we refer to as ‘functional connectivity (FC) markers’. See Results for further details.

**Step 2: Characterizing the findings of interest (unique variance modelling)**

The precise association of the FC markers for categorical or dimensional mechanisms of ADHD as obtained in Step 1 is not evident, since we did not account for the shared variance between the three predictors. In Step 2, we addressed this
issue by disentangling the specific association of the FC markers with the categorical ADHD-diagnosis and the dimensional inattention and hyperactivity/impulsivity scores. To that end, we obtained participant-level scores for every FC marker. This score was defined as the mean connectivity value across the voxels of the spatial cluster within every participant’s network spatial map. For every FC marker, these scores were used as dependent variable to evaluate three models that respectively evaluated the unique variance of the three effects of interest. We obtained a unique variance model for every evaluated predictor by replacing the tested predictor by its residual after regressing out the other two predictors (i.e., the opposite procedure compared to step 1). This effectively resulted in the evaluated predictor only representing its unique variance, while all shared variance was modelled by the other predictors (see Figure 1). Again these models were evaluated using non-parametric permutation testing (n=5000, p<0.05). We conducted a post-hoc sensitivity analysis to verify that the observed effects were not related to socio-economic status (SES), oppositional-deviant disorder diagnosis (ODD), conduct disorder diagnosis (CD), stimulant medication use, or amount of head motion during the scanning session (see Supplementary material).

**Neurobiological pathway association**

We investigated our network-findings in relation to neurocognitive and genetic measures. Specifically, we tested for an association across participants of the FC marker scores, as obtained in Step 2 of our framework, with neurocognitive or genetic measures known to be associated with ADHD. We used non-parametric permutation testing (n=5000, p<0.05), including covariates for age, sex, and scan site.

**Association with neurocognitive functioning**

We mapped all significant FC markers to four neurocognitive domains/measures: response inhibition (stop-signal reaction time, SSRT), visuospatial working memory (VSWM; percentage correct responses), reward sensitivity (RW; reaction time difference between a reward versus non-reward cue), and reaction time variability (RTV) on a motor timing task (van Ewijk et al., 2013; Thissen et al., 2014; van Rooij et al., 2015; von Rhein et al., 2015a). SSRT, VSWM, RW, and RTV were available for respectively 219, 244, 208, and 373 participants, after excluding outliers (see Supplementary Material). Importantly, the direction of the association of the four measures with ADHD differs: SSRT and RTV are expected to be positively associated, whereas VSWM and RW are expected to be negatively associated to ADHD symptom severity or ADHD-diagnosis (van Ewijk et al., 2013; Thissen et al., 2014; van Rooij et al., 2015; von Rhein et al., 2015a).
Association with genetic variants.

An overview of genetic procedures in the IMAGE and NeuroIMAGE study can be found elsewhere (von Rhein et al., 2015b). We focused on VNTR variants in two genes related to dopaminergic neurotransmission that have been implicated in ADHD through meta-analyses (Faraone et al., 2014; Gatt et al., 2015). The first is a VNTR variant located in the third exon of the dopamine receptor D4 gene (DRD4), which has been linked to receptor activity (Asghari et al., 1995). The second variant is a VNTR located in the 3′-untranslated region (3′UTR) of the SLC6A3/DAT1 gene encoding the dopamine transporter. This variant may interfere with the expression of the transporter (VanNess et al., 2005). In accordance with current literature (Wu et al., 2012; Faraone et al., 2014; Pappa et al., 2015), and taking into account the average age of our participants, we defined the 10-repeat allele of DAT1 as the risk allele; for DRD4 we chose the wide definition for risk, defining the risk factor as the group of alleles with more than 4 repeats (long allele), to maximize power. Specifically, for DAT1 we compared 10R homozygotes (n=244) with the rest of the sample (n=159). In the case of DRD4, we compared risk-carriers (n=161) against homozygotes for the non-risk alleles (n=241). We restricted our evaluation to only those FC markers (n=8) that were located in the DMN-posterior or executive control network (ECN), since these networks have been associated with dopamine neurotransmission in previous research (Cole et al., 2013).

Results

Functional brain networks

The group-ICA analysis yielded 14 independent components that closely corresponded to functional networks identified in previous research (see Figure 2)(Damoiseaux et al., 2006; Seeley et al., 2007; Smith et al., 2009; Zuo et al., 2010; Power et al., 2011; Thomas Yeo et al., 2011). See Supplementary material and Supplementary Figures 1-6 for an extensive comparison of the spatial maps and nomenclature of our networks compared to networks previously reported by other studies.

Categorical & dimensional functional connectivity analysis

We tested all 14 functional networks for association between functional connectivity scores and ADHD-diagnosis, inattention, and hyperactivity/impulsivity scores by exploiting three separate full variance models. This analysis yielded 27 significant spatial clusters located throughout the DMN (anterior and posterior), ECN, visual-
medial, sensorimotor (medial and lateral), and cerebellum network (see Figure 3; Table 2).

Characterization of the 27 FC markers by relating participant-level scores of these markers to the unique variance of ADHD-diagnosis and the two dimensional scores (see section 2.4.2) yielded distinct associations across networks, as shown in Figure 3 and summarized in Table 2. For the DMN, three FC markers located within the posterior cingulate cortex (PCC) and prefrontal cortex of the posterior and anterior DMN subnetworks were specifically related to the categorical ADHD-diagnosis. In contrast, the FC marker located in the middle temporal gyrus of the DMN-anterior was positively related to hyperactivity/impulsivity scores. Similarly, markers located within the sensorimotor subnetworks were mainly positively related to hyperactivity/impulsivity, while increased scores of markers within the cerebellum and visual-medial networks were predominantly positively related to inattention symptoms. Finally, FC markers located within the ECN, mainly covering the dorsomedial prefrontal cortex (dmPFC; including parts of the dorsal...
anterior cingulate and paracingulate cortex) and posterior parts of the left striatum, were related to both categorical and (both positive and negative) dimensional effects. Post-hoc sensitivity analysis showed that none of the findings described above were related to ODD, CD, SES, stimulant medication use, or head motion (see Supplementary Table 1).

**Neurobiological pathway association**

Mapping the FC markers to neurocognitive measures yielded no results for markers within the DMN-posterior network. Within the ECN, high connectivity with dmPFC was associated with poorer response inhibition and higher RTV. Markers located within the cerebellum showed a similar association with both these neurocognitive measures, whereas markers located within the visual-medial network were only positively associated with RTV. In contrast, FC markers within the DMN-anterior and sensorimotor networks showed a broad association across all cognitive metrics except for response inhibition (see Figure 3; Table 2).

Our evaluation of the association of dopaminergic genetic variation (in DAT1 and DRD4) with the FC markers was restricted *a priori* to markers located within the DMN-posterior or ECN (see section 3.1). We identified weak associations between the DRD4 variant and decreased functional connectivity of the dmPFC, left putamen, and insula within the ECN (see Figure 3; Table 2). Connectivity of the left putamen/insula within this network was also weakly associated with DAT1 variation.
Figure 3 - The 27 significant spatial clusters (as also presented in Table 2) related to ADHD across functional networks and their association with behavioural, neurocognitive and genetic measures. The clusters were obtained by evaluating the full variance models presented in Figure 1 on all 14 functional networks (see Figure 2), whereas the unique categorical and dimensional behavioural effects as illustrated in the coloured table below the brain slices were obtained by evaluating unique variance models for each of the 27 significant clusters. * p<0.05, ** p<0.01

Table 2 (next page) – Significant spatial clusters related to ADHD across functional networks and their association with behavioural, neurocognitive and genetic measures. The clusters were obtained by evaluating the full variance models presented in Figure 1 on all 14 functional networks (see Figure 2), whereas the unique categorical and dimensional behavioural effects as illustrated in the current figure were obtained by evaluating unique variance models obtained in step 2. Cluster size is reported in number of 3 mm isotropic voxels.
<table>
<thead>
<tr>
<th>Region</th>
<th>DAT1 (df=409)</th>
<th>DAT1 (df=373)</th>
<th>DAT1 (df=403)</th>
<th>DAT1 (df=409)</th>
<th>DAT1 (df=373)</th>
<th>DAT1 (df=403)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sensorimotor</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Visual – motor</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DMN – motor</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DMN – executive</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DMN – DMN</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DMN – ACC</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DMN – PAC</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DMN – dACC</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DMN – PCC</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DMN – ACC</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DMN – PAC</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DMN – dACC</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DMN – PCC</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DMN – ACC</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DMN – PAC</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DMN – dACC</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DMN – PCC</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DMN – ACC</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DMN – PAC</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DMN – dACC</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DMN – PCC</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DMN – ACC</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DMN – PAC</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DMN – dACC</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DMN – PCC</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DMN – ACC</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DMN – PAC</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DMN – dACC</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DMN – PCC</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DMN – ACC</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DMN – PAC</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DMN – dACC</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DMN – PCC</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DMN – ACC</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DMN – PAC</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DMN – dACC</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DMN – PCC</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Discussion

We aimed to parse ADHD-related heterogeneity by investigating functional connectivity across multiple brain networks, disentangling categorical and dimensional effects, and mapping these neural correlates to genetic and neurocognitive measures. We disentangled such categorical and dimensional neurobiological underpinnings through application of specific regression models in a large sample of adolescents and young adults. Effects observed in the DMN were mainly related to the categorical definition of ADHD, while effects located within the visual-medial network mainly related to inattentive behaviour and effects within the sensorimotor network to hyperactive/impulsive behaviour. Effects within the ECN and cerebellum showed both categorical and dimensional mechanisms. We found specific and meaningful relationships between the spatial clusters and neurocognitive measures as well as DAT1 and DRD4 variants.

Corroborating and extending previous studies, we found aberrant functional connectivity of the posterior cingulate cortex (PCC) and frontal areas within the DMN (Castellanos et al., 2008; Fair et al., 2010; Qiu et al., 2011; Elton et al., 2014; Posner et al., 2014). These findings mainly comprised categorical relationships between functional connectivity and an ADHD diagnosis. We found that the frontal FC marker within the DMN was associated with multiple neurocognitive measures, whereas two PCC markers were unrelated to the neurocognitive measures, suggesting that subdivisions within the DMN play a dissociable role in the aetiology of ADHD. Specifically regarding PCC, its absent association with neurocognitive measures and the myriad of findings described in literature on PCC abnormalities across psychiatric disorders (Broyd et al., 2009), support the idea that abnormalities of this region are either directly related to ADHD, not influenced by cognitive risk factors for ADHD, or unspecific to ADHD and a consequence of accumulating remote pathological effects (Leech and Sharp, 2014).

In contrast to the categorical effects within the DMN, we identified a profound dimensional effect related to inattention within the visual network. The relevance of this network in the pathophysiology of ADHD is being increasingly acknowledged (Castellanos and Proal, 2012; Cortese et al., 2012). Especially, the regulation of visual function by attentional processes is considered an important remaining research area (Castellanos and Proal, 2012). More specifically, our significant finding is located within the posterior part of the paracingulate cortex (PAC) and extends to superior frontal gyrus and pre-supplementary motor area; regions related to visual attention shifting (Arrington et al., 2000; Shulman et al., 2009; Nelissen et al., 2013). A large meta-analysis across 55 functional MRI studies confirmed
hypoactivation of PAC within ADHD patients as one of the main findings (Cortese et al., 2012). The authors related this finding to the ECN, in their work referred to as the ventral attention network. In line with the framework proposed by Nigg and Casey (2005) (Nigg and Casey, 2005), they hypothesized that ECN hypoactivation underpins deficits in detecting environmental (ir)regularities. This in turn would lead to behavioural problems, when patients with ADHD are unable to modulate their behaviour in accordance to these environmental changes. Interestingly, visual areas are involved in maintaining or suppressing spatial attention to irrelevant stimuli (Capotosto et al., 2009; Shulman et al., 2009), and it has been proposed that hyperactivation within the visual network might act as a neural compensatory mechanism for impaired function of prefrontal cortex and ACC, areas associated with the ECN (Fassbender and Schweitzer, 2006). Given that our analysis specifically links aberrant PAC connectivity to the visual network instead of the ECN, we can now integrate these hypotheses by proposing that ADHD-related deficits in detecting environmental (ir)regularities might be related to visual attention processing as regulated by PAC, possibly as a compensatory mechanism for dysfunction of the ECN. This hypothesis is supported by our observation that the PAC FC marker strongly and specifically related to RTV, which is in agreement with the known relation between RTV and inattentive behaviour (Leth-Steensen et al., 2000; Tamm et al., 2012; Antonini et al., 2013; Kofler et al., 2013).

Increased functional connectivity within the two sensorimotor networks was predominantly related to increased hyperactivity/impulsivity symptoms and poorer outcome on a broad range of neurocognitive measures, most strongly on working memory. These findings relate to the notion that motor control impairments are associated with higher-order cognitive difficulties (Georgopoulos, 2000) and support the idea that hyperactivity might be a compensatory mechanism to cope with environmental demands related to cognitive/executive functioning (Sarver et al., 2015; Kofler et al., 2016). We furthermore observed that FC markers located within cerebellum were predominantly related to both categorical ADHD-diagnosis and dimensional scores of inattentive behaviour, and poorer outcomes on response inhibition and RTV. Given that both these neurocognitive functions involve (motor) timing processing, our findings support the hypothesized neurobiological pathway, in which dysfunction of cerebellum is associated with impaired (motor) timing processing (Durston et al., 2011).

Various clusters identified within the ECN were associated with both categorical and dimensional effects of ADHD. The associations of these markers with neurocognitive and genetic measures suggested two differential neurobiological pathways. First, a pathway in which DRD4 modulates functional connectivity of fronto-striatal connections within the ECN, affecting motor timing and inhibitory
control. Second, a pathway in which \textit{DAT1} is related to connectivity within putamen/insula, but unrelated to neurocognitive functioning. This differential localization of effects related to \textit{DAT1} and \textit{DRD4} genetic variants nicely corresponds to the expression profile of both genes (Ciliax et al., 1995; Primus et al., 1997; De La Garza and Madras, 2000) and is in line with an earlier report that grey matter volume of the striatum (caudate nucleus) and PFC were respectively associated with the \textit{DAT1} and \textit{DRD4} polymorphisms (Durston et al., 2005). Moreover, our observed associations correspond with two reviews on the relationship of genetics and neuropsychological measures (Nanda N J Rommelse et al., 2008; Kebir and Joober, 2011). Notably, we found opposite directionality of the relationship between the FC marker and scores of inattention and hyperactivity/impulsivity respectively. Since these markers were related to RTV, this provides new leads for research into the counterintuitive notion that RTV is positively related to ADHD but is negatively related to the \textit{DRD4} risk-variant (Durston et al., 2005; Kebir and Joober, 2011; Kofler et al., 2013).

In the current work, we characterized dissociable neurobiological underpinnings of ADHD-related behaviour. We presented an analysis strategy for disentangling categorical and dimensional disease mechanisms in ADHD and mapped our results to genetic and neurocognitive measures. Using this approach, we provided new insights which are supportive for a categorical-dimensional model of ADHD and refined current hypotheses on the aetiology of this disorder. Our results emphasize that ADHD should be investigated throughout multiple neural systems and by combining both categorical and dimensional models, rather than focusing on one or the other.

Financial disclosures

\textbf{Raimon H.R. Pruim} reports no competing interests.
\textbf{Christian F. Beckmann} reports not competing interest.
\textbf{Marianne Oldehinkel} reports no competing interests.
\textbf{Jaap Oosterlaan}
Dr. Oosterlaan has received an unrestricted investigator initiated research grant from Shire pharmaceuticals.
\textbf{Dirk J. Heslenfeld} reports no competing interests.
\textbf{Catharina A. Hartman} reports no competing interests.
\textbf{Pieter J. Hoekstra}
Dr. Hoekstra has received an unrestricted research grant from Shire and has been member of the advisory boards of Shire and Eli Lilly.
Stephen V. Faraone
Dr. Faraone has received consulting income, travel expenses and/or research support from Pfizer, Shire, Akili Interactive Labs, Otsuka Pharmaceutical, VAYA Pharma, SynapDx, and Alcobra Pharma and research support from the National Institutes of Health (NIH). His institution is seeking a patent for the use of sodium-hydrogen exchange inhibitors in the treatment of ADHD. He has also received consulting fees from, was on advisory boards for, or participated in continuing medical education programs sponsored by Shire, Alcobra, Otsuka, Ortho-McNeil-Janssen Pharmaceuticals, Novartis, Pfizer, and Eli Lilly. Dr Faraone receives royalties from books published by Guilford Press, Straight Talk About Your Child’s Mental Health, and Oxford University Press, Schizophrenia: The Facts.

Barbara Franke
Dr. Franke has received educational speaking fees from Merz and Shire.

Jan K. Buitelaar
Dr. Buitelaar has been a consultant to/member of advisory boards of and/or a speaker for Janssen-Cilag BV, Eli Lilly, Shire, Novartis, Roche, Lundbeck, Medice and Servier. He is not an employee of any of these companies, and not a stock shareholder of any of these companies.

Maarten Mennes reports no competing interests.

Acknowledgement

This work was supported by National Institutes of Health grant R01MH62873 (Dr Faraone), NWO Large Investment grant 1750102007010 (Dr Buitelaar), ZonMW Addiction: Risk Behaviour and Dependency Grant 60-60600-97-193 (Dr Buitelaar), NWO Brain & Cognition: an Integrative Approach grant 433-09-242 (Dr Buitelaar), NWO National Initiative Brain & Cognition 056-13-015 (Dr Buitelaar), the EU FP7 grant TACTICS (grant no. 278948), and grants from Radboudumc, University Medical Center Groningen, Accare, and VU University Amsterdam. Dr. Faraone is supported by the K.G. Jebsen Centre for Research on Neuropsychiatric Disorders, University of Bergen, Bergen, Norway, the European Community’s Seventh Framework Programme (FP7/2007-2013) under grant agreement n°602805 and NIMH grant R01MH094469. Barbara Franke’s research is supported by grants from the Netherlands Organization for Scientific Research (NWO), i.e. the NWO Brain & Cognition Excellence Program (grant 433-09-229) and a Vici grant from NWO (grant 016-130-669). Her research also received funding from the European Community’s Seventh Framework Programme (FP7/2007 – 2013) under grant agreements n° 602805 (Aggressotype) and n° 602450 (IMAGEMEND), and from the European Community’s Horizon 2020 Programme (H2020/2014 – 2020) under grant
agreement n° 643051 (MiND). In addition, her work is supported by a grant for the ENIGMA Consortium (grant number U54 EB020403) from the BD2K Initiative of a cross-NIH partnership. Dr. Mennes has received funding from the European Research Council under the European Union’s Seventh Framework Programme (FP7/2007-2013)/ERC grant agreement 327340 (BRAIN FINGERPRINT). Dr. Beckmann is supported by the Netherlands Organisation for Scientific Research (NWO-Vidi 864-12-003) and we gratefully acknowledge the funding from the Wellcome Trust UK Strategic Award [098369/Z/12/Z]. We acknowledge the Department of Pediatrics of the VU University Medical Center for the opportunity to use the mock scanner for preparation of our participants.

References


Chapter 4

attention-deficit/hyperactivity disorder. Biol Psychiatry 71, 434–42.


Supplementary materials

Materials and methods – Extensive

Participants

Participants were selected from a follow-up (2009-2012) of the Dutch part of the International Multicenter ADHD Genetics (IMAGE) study, performed between 2003-2006 (Rommelse et al., 2008; Nijmeijer et al., 2009; Müller et al., 2011a, 2011b). At first enrolment, 365 families with at least one child with combined subtype ADHD and at least one biological sibling (regardless of ADHD diagnosis) were recruited, in addition to 148 control families with at least one child, with no formal or suspected ADHD diagnosis in any of the first-degree family members. Recruitment for ADHD families was accomplished through ADHD probands attending outpatient clinics in the regions Amsterdam, Groningen, and Nijmegen (the Netherlands), as well as a VU University affiliated ADHD research institute. Control families were recruited through primary and high schools in the same geographical regions as the participating ADHD families. All family members, also those who did not participate in IMAGE, were invited for follow-up measurement with a mean follow-up period of 5.9 years (SD = 0.74). Follow-up rates were 78.4% for ADHD families and 80.4% for control families. In order to balance out the distribution of gender and age between the ADHD and healthy control groups, additional girls with ADHD (any subtype; N=37 families) and healthy control boys (N=34 families) were recruited for NeuroIMAGE. Inclusion criteria were the same for all participants, and largely consistent with IMAGE: participants had to be between 5-30 years, of European Caucasian descent, have an IQ ≥ 70, and no diagnosis of autism, epilepsy, general learning difficulties, brain disorders and known genetic disorders (such as Fragile X syndrome or Down syndrome). Since the data used in the current study included an MRI scanning session, participants were excluded if they were younger than 8 years or had any contraindication to MRI scanning (e.g. implanted metal or medical devices, or possible pregnancy). Including the newly recruited families, the complete NeuroIMAGE cohort comprised 323 ADHD families and 153 control families.

For our current study we only considered participants for whom both a resting-state fMRI and structural MRI scan was available (n=545). Ninety participants were excluded due to: incomplete scanning (n=9), poor field-of-view coverage (n=23), incidental findings (n = 10), official exclusion criteria been put forward after inclusion (n=12), incomplete CPRS-questionnaire (n=16), and/or because they belonged to the 5% highest movers as determined by a motion summary score (i.e. motion outliers; n=27). We used the root mean squared of the frame-wise
displacement time-series as a participant-level summary motion score (RMS-FD) (Jenkinson et al., 2002).

Procedure

The current study was part of a comprehensive assessment protocol encompassing behavioural questionnaires, a diagnostic interview and several neurocognitive measures from all family members, and an extensive MRI scanning protocol in participating children. From participants whose genotypic information was missing during IMAGE, saliva was collected for DNA analysis. Testing was carried out either at the VU University Amsterdam and VU University Medical Centre, or at the Radboud University Nijmegen Medical Centre and Donders Institute for Brain, Cognition and Behaviour in Nijmegen. Participants were asked to withhold use of psychoactive drugs for 48 hours before measurement. During the testing day, participants were motivated with short breaks, and at the end of the day, children received a reward of €50, and a copy of their MRI scan. Informed consent was signed by all participants (parents signed informed consent for participants under 12 years of age), and the study was approved by the ethical committee (Centrale Commissie Mensgebonden Onderzoek).

MRI acquisition

Data were acquired at two scanning locations on similar 1.5 Tesla Siemens scanners (Siemens Sonata at VU University Medical Centre in Amsterdam; Siemens Avanto at Donders Centre for Cognitive Neuroimaging in Nijmegen) using the same Siemens 8-channel head coil and identical scanning protocols. Anatomical images were obtained using an MPRAGE sequence (TR=2730 ms, TE=2.95 ms, T1=1000 ms, flip angle=7, matrix size=256x256, FOV=256 mm, 176 slices with 1 mm isotropic voxels). Functional images during rest were obtained using a gradient echo echo-planar imaging (GE-EPI) sequence (TR=1960 ms, TE=40 ms, FOV=224 mm, 37 axial slices, flip angle=80, matrix size=64x64, in-plane resolution=3.5 mm, slice thickness/gap=3.0 mm/0.5 mm). Participants were instructed to relax with their eyes open during the rfMRI scan.

(rf)MRI data preprocessing

Preprocessing of rfMRI data was carried out using tools from the FMRIB Software Library (FSL; http://www.fmrib.ox.ac.uk/fsl) (Smith et al., 2004; Woolrich et al., 2009; Jenkinson et al., 2012) and involved removal of the first five volumes to allow for signal equilibration, head movement correction by volume-realignment to the middle volume using MCFLIRT, global 4D mean intensity normalization, and spatial smoothing (6 mm FWHM). Subsequently, we applied ICA-AROMA which is a data-driven strategy that effectively identifies residual motion artifacts in fMRI.
data (Pruim et al., 2015a, 2015b). Identified motion artifacts were subsequently removed from the data using fsl_regfilt from FSL, which implements a multivariate regression. We complemented the nuisance regression, by using mean white matter and CSF time-courses as well as linear trend as nuisance regressors. Preprocessing was completed with the application of temporal high-pass filtering (>0.01Hz).

For every participant we then transformed the preprocessed fMRI data to the participants’ structural image using a single pre-calculated affine boundary-based registration (FLIRT) (Jenkinson and Smith, 2001; Jenkinson et al., 2002). Subsequently, we registered the functional data to our NeuroIMAGE study template in MNI152 space (3mm isotropic resolution) by applying the non-linear transformation between the participant’s high-resolution T1 image and the study template brain as obtained using FSL FNIRT (Andersson et al., 2007).

**Post-hoc sensitivity analysis**

We conducted post-hoc sensitivity analyses to verify that our observed effects, on the unique association of our functional connectivity markers with a categorical ADHD-diagnosis or dimensional symptom variables, were not related to socio-economic status (SES), oppositional-deviant disorder diagnosis (ODD), conduct disorder diagnosis (CD), stimulant medication use, or amount of head motion during the scanning session. To that end we replicated the analysis within the second stage of our categorical/dimensional modelling framework using additional covariates for SES, ODD-CD diagnosis, stimulant treatment duration and a motion summary score (RMS-FD), next to the already included covariates for age, sex and scanning site. As a measure for SES we used an average measure of ‘years of education’ from both parents, scaling from 0 (no formal education) to 17 (university) years of education (Buis, 2010). Measures on duration of stimulant medication use were obtained through pharmacy reports. Specifically, we evaluated the three ‘unique-variance’ models of stage two to the participant-level cluster score, using non-parametric permutation testing. Due to association of these covariates with ADHD diagnosis and/or symptoms they are expected to reduce sensitivity to the effects of interest. Therefore, we do not aim to replicate statistical significance of the effects but rather qualitatively investigate whether the effects are in agreement with the initial results.

As evident in Supplementary Table 1 we replicated our effects described in the main manuscript. With respect to the significance of the results, only the association of functional connectivity of the caudate nucleus / putamen with the ECN was not significantly related to scores of hyperactivity/impulsivity. However, this concerned a slightly reduced T-value from -1.7 (p=0.05) in the main analysis to -1.5 (p=0.08) in
the current analysis, possibly induces by reduces sensitivity. Moreover, we found the association of the right V + VI in cerebellum which was just above 0.05 threshold in the main analysis, now to be significant with $p=0.04$. 
### Supplementary Table 1

Post-hoc sensitivity analysis results. Within-network functional connectivity findings related to categorical or dimensional effects of ADHD, including covariates for ODD/CD-diagnosis, head motion, medication use and socioeconomic status (in addition to covariates for age, sex and scanner location). 

- **R**: right hemisphere, **L**: left hemisphere, **Post.**: posterior, **DMN**: default mode network, **ECN**: executive control network, **dACC**: dorsal anterior cingulate cortex, **PAC**: paracingulate cortex, **vPFC**: ventral prefrontal cortex, **MTG**: middle temporal gyrus, **PCC**: posterior cingulate cortex, **dmPFC**: dorsomedial PFC, **dlPFC**: dorsolateral PFC, **Caud. N.**: Caudate Nucleus, **WM**: white matter, **OFG**: occipital fusiform gyrus, **ACC**: anterior cingulate cortex.

<table>
<thead>
<tr>
<th>RSN/Brain Area</th>
<th>Association symptoms (unique variance)</th>
<th>T-value (p)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>ADHD&gt;TDC</td>
<td>Inattention</td>
</tr>
<tr>
<td>DMN – anterior</td>
<td></td>
<td></td>
</tr>
<tr>
<td>dACC, PAC, vPFC</td>
<td>4.9 (0.00)**</td>
<td>-0.0 (0.50)</td>
</tr>
<tr>
<td>MTG (R)</td>
<td>-0.6 (0.28)</td>
<td>0.0 (0.49)</td>
</tr>
<tr>
<td>DMN – posterior</td>
<td></td>
<td></td>
</tr>
<tr>
<td>PCC (R)</td>
<td>2.1 (0.02)*</td>
<td>-0.3 (0.37)</td>
</tr>
<tr>
<td>PCC (R)</td>
<td>4.2 (0.00)**</td>
<td>-1.1 (0.15)</td>
</tr>
<tr>
<td>ECN</td>
<td></td>
<td></td>
</tr>
<tr>
<td>dmPFC (PAC, dACC)</td>
<td>2.9 (0.00)**</td>
<td>2.6 (0.00)**</td>
</tr>
<tr>
<td>Caud. N./Putamen (L, Post.)</td>
<td>4.6 (0.00)**</td>
<td>0.7 (0.24)</td>
</tr>
<tr>
<td>Putamen/insula (L, Post.)</td>
<td>4.9 (0.00)**</td>
<td>-0.2 (0.42)</td>
</tr>
<tr>
<td>dmPFC, (PAC, dACC), dlPFC</td>
<td>1.9 (0.03)*</td>
<td>3.4 (0.00)**</td>
</tr>
<tr>
<td>WM (L)</td>
<td>0.8 (0.20)</td>
<td>3.0 (0.00)**</td>
</tr>
<tr>
<td>Left VI</td>
<td>0.4 (0.36)</td>
<td>-0.8 (0.20)</td>
</tr>
<tr>
<td>Visual – medial</td>
<td></td>
<td></td>
</tr>
<tr>
<td>PAC</td>
<td>-0.3 (0.38)</td>
<td>2.5 (0.01)**</td>
</tr>
<tr>
<td>Precentral gyrus (L)</td>
<td>1.0 (0.17)</td>
<td>1.2 (0.12)</td>
</tr>
<tr>
<td>PAC</td>
<td>-0.4 (0.36)</td>
<td>1.4 (0.08)</td>
</tr>
<tr>
<td>Sensorimotor – medial</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Precentral gyrus (R)</td>
<td>4.7 (0.00)**</td>
<td>-0.8 (0.22)</td>
</tr>
<tr>
<td>Precentral gyrus</td>
<td>0.5 (0.32)</td>
<td>-0.4 (0.33)</td>
</tr>
<tr>
<td>Precentral gyrus (L)</td>
<td>1.3 (0.10)</td>
<td>-1.0 (0.17)</td>
</tr>
<tr>
<td>ACC (R)</td>
<td>1.5 (0.07)</td>
<td>0.1 (0.48)</td>
</tr>
<tr>
<td>Precuneus (L)</td>
<td>0.6 (0.26)</td>
<td>-0.3 (0.37)</td>
</tr>
<tr>
<td>Sensorimotor – lateral</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Precentral gyrus (L)</td>
<td>1.3 (0.10)</td>
<td>-0.6 (0.25)</td>
</tr>
<tr>
<td>Precentral gyrus (L)</td>
<td>-1.2 (0.11)</td>
<td>0.3 (0.38)</td>
</tr>
<tr>
<td>Postcentral gyrus (L)</td>
<td>0.6 (0.29)</td>
<td>0.5 (0.29)</td>
</tr>
<tr>
<td>Cerebellum</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Right V + VI</td>
<td>2.9 (0.00)**</td>
<td>1.3 (0.10)</td>
</tr>
<tr>
<td>Cerebellum, OFG (L)</td>
<td>0.1 (0.47)</td>
<td>3.0 (0.00)**</td>
</tr>
<tr>
<td>WM (L)</td>
<td>-1.6 (0.05)</td>
<td>2.4 (0.01)*</td>
</tr>
<tr>
<td>Vermis VI. Left V. Right V</td>
<td>-1.4 (0.07)</td>
<td>1.9 (0.04)*</td>
</tr>
<tr>
<td>Right VI</td>
<td>0.1 (0.46)</td>
<td>0.6 (0.27)</td>
</tr>
<tr>
<td>Right I-IV</td>
<td>1.0 (0.17)</td>
<td>0.5 (0.32)</td>
</tr>
</tbody>
</table>
Functional network characterization and nomenclature

Multiple studies have been performed to identify and characterize (functional) brain networks, using both data-driven and more hypothesis-driven techniques (Beckmann et al., 2005; Dosenbach et al., 2007; Seeley et al., 2007; Smith et al., 2009; Power et al., 2011). Although networks obtained throughout these studies seem to be very consistent, the nomenclature of part of these networks is not consistent and often confusing in the current literature. This is partly due to the fact that it is still unclear to which extent these networks actually reflect similar networks or whether they have to be regarded as distinct (sub)networks.

The inconsistent nomenclature predominantly relates to two networks of brain regions located in frontal and parietal cortices, related to cognitive functioning (Seeley et al., 2007; Smith et al., 2009; Menon, 2011). The first network is anchored in the dorsolateral prefrontal cortex and lateral posterior parietal cortex, here referred to as the fronto-parietal network (FPN). The second network is anchored in the dorsal anterior cingulate cortex (DACC and anterior insula and here referred to as the executive control network (ECN)(Seeley et al., 2007; Smith et al., 2009; Menon, 2011). The FPN and ECN are functionally related and therefore confusingly similarly referred to in literature, but importantly concern distinct networks (Cole and Schneider, 2007; Smith et al., 2009; Menon, 2011).

Regarding the FPN, researchers typically find and discuss either a single non-lateralized (Power et al., 2011; Thomas Yeo et al., 2011) or two lateralized networks (Damoiseaux et al., 2006; Smith et al., 2009; Zuo et al., 2010; Menon, 2011). This network is also commonly referred to as the ‘central executive’ or ‘executive control’ network (Seeley et al., 2007; Menon, 2011) but should importantly not be confused with the insula/dACC-anchored network labeled as ‘executive control’ (ECN) in our current and other studies (Beckmann et al., 2005; Damoiseaux et al., 2006; Smith et al., 2009; Zuo et al., 2010). Moreover, the network we here refer to as ECN shows large spatial overlap with, and is commonly referred to as, the ventral attention, salience or cingulo-opercular network (Dosenbach et al., 2007; Seeley et al., 2007; Thomas Yeo et al., 2011; Kucyi et al., 2012; Uddin et al., 2013).

Next to these networks related to cognitive functioning, studies also find different subnetworks of for instance the default mode, visual and sensorimotor/auditory networks. Therefore, we will here discuss the networks identified within our main study by comparing these to networks already proposed
in previous literature and clarifying the nomenclature of these networks accordingly. To that end, we compared our 14 networks (see Supplementary Figure 1) to the seven networks obtained by Yeo et al. (2011), the ten networks obtained by Smith et al. (2009), the cingulo-opercular network proposed by Dosenbach et al. (2007), and the salience and central executive (also referred to as executive control) networks proposed by Seeley et al. (2007).
Supplementary Figure 1 - Overview of the 14 functional networks as obtained within our study by exploiting a group-ICA on the resting-state fMRI data of 46 typically developing controls (excluded from any further analyses in our study)
Method

The spatial network templates obtained by Yeo et al. and Smith et al. are publically available and used accordingly. To obtain a network map of the three other networks (i.e., cingulo-opercular, salience and central executive networks) we conducted seed-based regression analysis using the rfMRI data of the 46 participants also used to obtain the functional networks within our main study using a group-ICA.

To that end, we first derived participant-specific spatial maps of each of the three networks. To obtain the salience and central executive networks described by Seeley et al. (2007) we respectively defined a seed (6mm radius sphere) at the anterior insula (MNI: x=38, y=26, z=-1) and dorsolateral prefrontal cortex (MNI: x=44, y=36, z=20). To derive the cingulo-opercular network we defined a single seed-mask comprising the seven seeds (6mm radius sphere) corresponding to the brain regions reported to describe this network (Dosenbach et al., 2007). For every participant and every network we derived the average time-course within the seed-region, and subsequently used this time-course as a regressor (after demeaning and variance normalization) within an ordinary least squares procedure against the rfMRI data, providing a Z-statistical participant-specific network map. Finally, we derived the mean spatial map across all participants giving the final network map for each of the three networks.

All networks (except the binary networks by Yeo et al.) were thresholded using a mixture modeling procedure (p>0.95). Supplementary Figure 2-5 illustrates the comparison results, categorized by the network-labeling used in the main text of our manuscript. For visualization purposes, the colormap of every image was scaled, for every network specifically, to the first (minimum) and 99th (maximum) percentile of the full range of intensity values of the non-thresholded maps.

Results/discussion

Figure 2 presents the results of the network anchored in the dorsolateral prefrontal cortex and lateral posterior parietal cortex (i.e., FPN). This figure illustrates that the network commonly referred to as ‘central executive’ or ‘executive control’ (Seeley et al., 2007) highly relates to the (lateralized) FPNs obtained in other studies.

Figure 3 demonstrates the high spatial similarity between our defined executive control network with the ventral attention, cingulo-opercular and salience network. This results endorses that idea that these networks might in fact represent similar neural systems in which the variability is mostly due to methodological differences.; i.e. a regression procedure using a single predefined seed (Seeley et al.,
2007), a data-driven clustering of seed/nodes (Dosenbach et al., 2007; Power et al., 2010; Thomas Yeo et al., 2011), voxel-wise ICA-based procedures (Beckmann et al., 2005; Smith et al., 2009). For example, exploiting the anterior insula as a seed region to obtain the salience network will inherently result in a network-map with a (more) profound effect in this region compared to a network obtained in a data-driven fashion. Although most studies identify a single network anchored in the dACC and anterior insula, it has also been proposed that these networks are distinct (Power et al., 2011).

Figures 4-6 present the results of the default mode, visual, sensorimotor, auditory and dorsal attention network. These results show that the networks either largely overlap or represent subnetworks of the networks defined by Smith et al. (2009) and Yeo et al. (2011). Importantly, our results furthermore corroborate to the consistency of functional networks obtained across sites, datasets, and analysis procedures.

Supplementary Figure 2 - The spatial map of the fronto-parietal network (thresholded using mixture modeling procedure, p>0.95) as obtained in our study (Pruim) compared to maps obtained in previous studies (Dosenbach et al., 2007; Seeley et al., 2007; Smith et al., 2009; Thomas Yeo et al., 2011)
Network correlates of ADHD

Executive control network

Supplementary Figure 3 - The spatial map of the executive control network (thresholded using mixture modeling procedure, $p>0.95$) as obtained in our study (Pruim) compared to maps obtained in previous studies (Dosenbach et al., 2007; Seeley et al., 2007; Smith et al., 2009; Thomas Yeo et al., 2011).

Default mode network

Supplementary Figure 4 - The spatial map of the default mode network (thresholded using mixture modeling procedure, $p>0.95$) as obtained in our study (Pruim et al.) compared to maps obtained in previous studies (Smith et al., 2009; Thomas Yeo et al., 2011).
Supplementary Figure 5 - The spatial map of the visual and dorsal attention networks (thresholded using mixture modeling procedure, $p>0.95$) as obtained in our study (Pruim et al.) compared to maps obtained in previous studies (Smith et al., 2009; Thomas Yeo et al., 2011)
Supplementary Figure 6 - The spatial map of the sensorimotor, auditory and cerebellum networks (thresholded using mixture modeling procedure, \(p > 0.95\)) as obtained in our study (Pruim et al.) compared to maps obtained in previous studies (Smith et al., 2009; Thomas Yeo et al., 2011)
References


Buis, M.L., 2010. Inequality of Educational Outcome and Inequality of Educational Opportunity in the Netherlands during the 20th Century; Chapter 3: Scaling levels of education.


Summary and general discussion
In psychiatry there is a movement towards new conceptualizations of mental disorders. First of all, there is an increasing appreciation that diagnosed individuals represent the ‘end of a continuum’ on certain behavioral domains rather than a class of individuals with a systematic difference compared to ‘normal’ individuals (Coghill and Sonuga-Barke, 2012). Moreover, a ‘neural-systems’ view of mental disorders is emerging in which it is appreciated that dissociable neural deficits might underpin similar behavioral dysfunction, i.e. the existence of dissociable (categorical) subtypes of patients or (dimensional) dysfunctional bio-behavioral pathways (Makris et al., 2009; Menon, 2011). The resting-state fMRI (rfMRI) imaging protocol has large potential for investigating these aspects at the level of the functional architecture of the brain, by allowing to simultaneously investigate multiple (dissociable) neural networks.

Despite these promising developments in understanding the pathophysiological mechanism underlying mental disorders, the psychiatry research field is confronted with high heterogeneity within and between studies (Castellanos et al., 2009; Oldehinkel et al., 2013; Vargas et al., 2013; Mulders et al., 2015). This is a consequence of insufficient and/or suboptimal exploited methodology related to the neural-systems and categorical-dimensional conceptualization of mental disorders. Resting-state fMRI research is furthermore confronted with detrimental effects of structured noise in rfMRI data due to head movement of the participant during the scanning procedure. In the current thesis I developed new analytical methods to address these drawbacks and optimize research into the association between functional brain architecture and mental disorders. I applied the developed methodology to study the functional neural architecture of attention-deficit/hyperactivity disorder; a prime example of a disorder in which these drawbacks emerge. In the current section I will first summarize the work (developed methods and results) presented in this thesis and discuss them in relation to current literature. Next, I will discuss limitations of the methods I used and provide directions for future research.

**Developed methods and results**

**Denoising resting-state fMRI data**

Head motion is known to have a detrimental effect on functional MRI data (Friston et al., 1996). Such motion causes spatial misalignment of the acquired fMRI volumes (i.e. voxels do not represent the same brain area over time) and additional effects related to partial voluming, interpolation effects, magnetic field inhomogeneities, and spin history effects. These effects of head motion are typically accounted for by
spatially aligning the acquired fMRI volumes. Next, the motion parameters obtained by the realignment procedure are exploited as confound regressors in a linear regression procedure to regress out variance associated with secondary effects of head motion. However, in 2012 three studies were published, respectively by Van Dijk et al., (2012), Power et al. (2012) and Satterthwaite et al. (2012), which showed that functional connectivity metrics obtained from resting-state fMRI data are still heavily impacted by head motion despite this conventional method to remove motion artifacts. This elicited a lively debate in the rfMRI community, in which many subsequent studies were published, investigating various aspects of the origin or impact of head motion artifacts and/or proposing new methods for their removal (Wilke, 2012; Yan et al., 2013a, 2013b, Satterthwaite et al., 2013a, 2013b; Beall and Lowe, 2014; Couvy-Duchesne et al., 2014; Zeng et al., 2014; Kong et al., 2014; Muschelli et al., 2014; Patel et al., 2014; Pujol et al., 2014; Scheinost et al., 2014; Spisák et al., 2014). The method which gained most attention and was being increasingly used in the research field incorporated motion artifact removal by detecting high-motion time-frames (i.e. scan volumes) and subsequently rigorously deleting or regressing them out from the dataset, referred to as ‘scrubbing’ or ‘spike regression’ (Power et al., 2012; Satterthwaite et al., 2013a).

However, these methods highly depend on realignment parameters, destroy temporal autocorrelation, and induce a large and variable loss in degrees of freedom. Accordingly, in Chapter 1, I proposed and developed a new method for motion artifact removal called ‘ICA-AROMA’. This method utilized independent component analysis (ICA), a method that decomposes data into independent signal sources; in the case of rfMRI data these sources comprise BOLD-signal components (e.g. a function brain network) and structured noise components (e.g. motion artifact). Therefore, ICA offers to opportunity to detect motion artifacts in a data-driven fashion, which can then subsequently be removed from the data (Beckmann, 2012). However, the implementation of ICA for noise-removal is not straightforward as it requires to specifically identify which of the estimated components represent noise/motion, avoiding removal of components that represent signal of interest. Such identification of noise components can be performed manually (Kelly et al., 2010). Alternatively, one can develop an algorithm for automatic identification (i.e. a ‘classifier’) based on spatial-temporal characteristics of the components (‘features’) (e.g. ICA-FIX (Salimi-Khorshidi et al., 2014)). Whereas full manual assessment is subjective and labor intensive, automatic classification is challenging due to heterogeneity of component characteristics which complicates generalizability of the developed classifier across multiple datasets. To address the issue of generalizability, I carefully designed a small set of standardized, intuitive and robust features (using a small set of rfMRI scans), specifically focusing on identification of motion artifacts. ICA-AROMA classifies the components estimated
by ICA using these features and subsequently regresses out the variance associated with these components from the rfMRI using fsl_regfilt.

I validated ICA-AROMA by showing that it removes motion artifacts with high accuracy while preserving signal of interest, in both resting-state and task-based fMRI datasets, outperforming spike regression (Satterthwaite et al., 2013a) and conventional confound regression (Friston et al., 1996). The generalizability of ICA-AROMA towards task-based fMRI data, while being developed using resting-state fMRI data, provides a nice validation on the robustness of the method. After this initial validation I, in Chapter 2, extensively compared ICA-AROMA against a range of alternative methods that are currently used in the research field and evaluated the performance on motion artifact removal. I investigated functional connectivity metrics using both seed-based and dual-regression approaches and replicated the analysis across four different datasets. Importantly, current studies mainly focused on the extent to which motion artifacts have been removed while a thorough evaluation of the performance on preserving signal of interest is often overlooked (e.g. (Power et al., 2012)). Accordingly, in my research I additionally evaluated the loss in degrees of freedom and exploited novel metrics to evaluate the preservation of signal of interest and reproducibility of functional connectivity metrics. Notably, the awareness of these aspects is currently growing in the field. For instance, recent work by Bright and Murphy (2015) and Shirer et al. (2015) emphasizes the importance of thorough evaluation of signal of interest, and show that conventional methods remove such signal to a large extent, and an open science resource has been initiated for studying reliability and reproducibility in functional connectivity analysis (Zuo et al., 2014).

The evaluation of ICA-AROMA showed that it removes motion artifact to a same extent as scrubbing and spike regression, preserves signal of interest across datasets, increases reproducibility of functional connectivity metrics and has limited cost in lost degrees of freedom. Consistent with current literature, the findings illustrated that scrubbing and spike regression remove motion artifacts to a large extent (Power et al., 2012; Satterthwaite et al., 2013a). However, the increase of reproducibility with respect to conventional confound regression methods was only minimal compared to the increases obtained by ICA-based methods, while moreover having a high cost in the loss of degrees of freedom. The findings furthermore showed the drawback of limited generalizability of another ICA-based method called ICA-FIX, (Griffanti et al., 2014; Salimi-Khorshidi et al., 2014); stressing out the need for re-training of the ICA-FIX classifier when applying it to a new dataset, as already suggested by its developers (i.e. provide a set of manually labeled components obtained from a random subset of scans to allow the algorithm to ‘learn’ how to distinguish noise from signal in this particular dataset).
Conceptualization of mental disorders

It is increasingly appreciated that we should move towards integrated categorical-dimensional approaches for conceptualizing and investigating mental disorders (Chabernaud et al., 2011; Coghill and Sonuga-Barke, 2012). In this context, categorical effects comprise systematic differences between patients and controls, whereas dimensional effects reflect a continuum of a certain measure/phenotype in which patients might represent the end of that spectrum. As an example in the research field of ADHD, Chabernaud et al. (2011) and Elton et al. (2014) have investigated categorical and dimensional mechanisms related to this disorder and both found neural underpinnings associated with categorical and/or dimensional ADHD-related behavioral measures. However, in these research papers, and in current literature in general, there is little awareness of the methodological implications associated with these types of analyses. Chapter 3 aimed to raise such awareness by explaining that, due to high correlation between categorical and dimensional variables, some typically used models increase false positive rate while others suffer from reduced sensitivity (York, 2012; Mumford et al., 2015). To overcome these limitations, I proposed a modelling framework by utilizing selective orthogonalization of categorical and dimensional variables. As a proof of principle I applied this framework to investigate functional connectivity of the executive control network in ADHD. These results demonstrated how current modelling approaches indeed lead to reduced sensitivity and/or false interpretation of the obtained results. Our framework on the other hand provided a more refined characterization of the results while maintaining sensitivity to the effects of interest.

Next to categorical-dimensional views on mental disorders; mental disorders are also more and more believed to be multifactorial in its underlying pathology, affecting multiple independent neural systems (Makris et al., 2009; Menon, 2011). This aspect is known as ‘equifinality’, meaning different pathophysiological mechanism can result in similar behavioral symptoms. Resting-state fMRI allows to model different functional neural networks and therefore investigate this multifactorial aspect of mental disorders. However, to date researchers have typically employed univariate modeling procedures to model a specific network. By not acknowledging the multivariate nature of the data (i.e. simultaneous activation of multiple networks) these approaches are suboptimal and result in less specific networks (Smith et al., 2012). Moreover, in line with this limitation, researchers often define and investigated brain regions or networks using a ‘hard’ parcellation of networks in which every voxel is assigned to a single region/network. However, networks share anatomical infrastructure and brain regions can be involved in multiple networks, which complicates the interpretation of findings base on ‘hard’ parcellation. Fortunately, these drawbacks can be addressed by conducting
multivariate modeling of neural networks using ICA (which allows spatial overlap between networks/components) and dual regression (Beckmann et al., 2009; Filippini et al., 2009). Accordingly, we combined this method with the proposed methodology in this thesis into a single study to investigate functional connectivity in the context of ADHD. Specifically, we used ICA-AROMA to preprocess rfMRI data, from which we estimated a set of functional networks using dual regression, which were then investigated for categorical and dimensional mechanisms related to ADHD.

This functional connectivity study, see **Chapter 4**, identified a set of significant findings throughout different networks in which functional connectivity was associated with categorical and/or dimensional measures of ADHD; comprising the default mode, executive control, cerebellum, visual and motor networks. The anatomical locations of these findings are highly consistent with current literature in ADHD but the study provided a more detailed characterization of the findings by specifically relating them to dissociable networks and identifying its categorical-dimensional complexity.

A nice example regards the dimensional finding that increased functional connectivity of the paracingulate cortex within the visual–medial network was associated with inattentive behavior. Importantly, the exact same region was found to be significantly associated with ADHD in a neural systems meta-analysis of fMRI studies by Cortese et al. (2012). See Figure 1 for a comparison of their results with our finding presented in Chapter 4. However, this meta-analysis assigned all findings to specific neural systems using a ‘hard’ parcellation of the brain from a study by Thomas Yeo et al. (2011); in which the paracingulate cortex was assigned to the executive control network (ECN; in their work referred to as the ventral attention network). The authors interpreted this finding in line with a prominent hypothesis proposed by Nigg and Casey (2005), stating that ECN hypoactivation underpins deficits in detecting environmental (ir)regularities, leading to ADHD-related problems in modulating behavior according to these environmental changes. However, the ‘soft’ group-ICA parcellation used in this chapter showed that this region is not only involved in the ECN but also in the visual-medial network. Notably, it was specifically the functional connectivity of the paracingulate cortex within this visual network that showed the signification association with inattentive
behavior. Combined with the notions that visual areas can be involved in suppressing spatial attention to stimuli and might play a compensatory role for impaired functioning of brain regions within the ECN, this finding allowed refining the hypothesis on spatial and visual attention processing in ADHD. Specifically, Chapter 4 hypothesizes on the basis of these findings that ADHD-related deficits in detecting environmental (ir)regularities relate to visual attention processing regulated by the paracingulate cortex, possibly as a neural compensatory mechanism for dysfunction of the ECN. Importantly, these results stress out the importance of appropriate multivariate modeling of neural networks.

---

**Figure 1 - Comparison of the significant results located at the posterior part of the paracingulate cortex in respectively the neural systems' meta-analysis of Cortese et al. (2012) and neural systems' research conducted in Chapter 5 (i.e. Pruim et al.). The two left figures where adapted from the paper by Cortese et.al. (adapted from Cortese et al. (2012)). The upper figures show the significant clusters obtained in both studies. The lower left figure shows the (non-overlapping) network templates identified by Thomas Yeo et al. (2011) on the basis of which Cortese et.al. assigned their findings to the executive control network (ECN; in their work referred to as ventral attention network). The lower right figure illustrates the visual network identified in Chapter 5 using group-ICA (allowing spatial overlap between estimated networks), which showed involvement of the posterior paracingulate, next to it also being part of the estimated ECN. Importantly, the significant cluster (top right) regards the association of functional connectivity of paracingulate cortex within the visual network with ADHD (specifically: a dimensional association with inattentive behavior).**
Importantly, all the developed and proposed methodology in this thesis converges in these results, illustrating their (complementary) benefit for modeling functional brain architecture in mental disorders. First, effective cleaning of the rfMRI data using ICA-AROMA (Chapter 1 and 2) resulted in increased sensitivity such that we could derive very specific network templates (e.g. identify the paracingulate cortex as part of the visual network; see supplementary Figure 3 in Chapter 1 and Figure 1 in Chapter 4), and replicate the main finding of a meta-analysis of 55 fMRI studies. Second, due to the categorical-dimensional modeling framework (Chapter 3) we could characterize this finding as a dimensional mechanism related to inattentive behavior. And third, using multivariate network modeling (Chapter 4) we were able to associate this finding specifically to the visual network rather than the (expected) ECN which allowed more accurate interpretation and refinement of a hypothesis on visual-attention in ADHD.

For further validation of the findings and to explore genetic-neural-cognitive-behavioral pathways, the functional connectivity results in Chapter 4 were related to four ADHD-related cognitive measures (working memory, response inhibition, reaction time variability, reward sensitivity) and two dopamine neurotransmission-related genetic variants (in the genes DAT1 and DRD4) (Faraone et al., 2014). The results for example suggest two categorical-dimensional pathways within the ECN. The first, comprising a pathway in which the DRD4 gene modulates functional connectivity of fronto-striatal pathways within the ECN, affecting timing and inhibitory control. The second, comprising a pathway in which the DAT1 gene is related to connectivity within putamen/insula, unspecific to cognitive domains. Moreover, we found a categorical mechanism related to ADHD-diagnosis of functional connectivity of the PCC/precuneus within the default mode network; this is considered to be a key locus in ADHD pathology (Castellanos et al., 2008). Interestingly, this finding was unrelated to cognitive ADHD-related measures and the locus is found in many psychiatric disorders, suggesting that pathology/dysfunctioning of the PCC/precuneus is relevant yet unspecific to ADHD (Broyd et al., 2009; Leech and Sharp, 2014).

Accordingly, by using a neural systems approach, disentangling categorical-dimensional and relating the findings to genetic and cognitive measures of ADHD, Chapter 4 was able provide new insights into the neurobiological mechanisms (i.e. genetic-neural-cognitive-behavioral pathways) of ADHD.
Future directions

Denoising resting-state fMRI data

Although, as shown in Chapter 1 and 2, ICA-AROMA provides effective removal of motion artifacts, head motion can lead to signal loss that no denoising method will be able to recover. Accordingly, most important in reducing motion artifacts is to minimize the amount of head motion during the scanning procedure, for instance by minor fixation of the head using a piece of tape. Moreover, attempts have been made using real-time correction during the scanning procedure (i.e. prospective motion correction) (Maclaren et al., 2013), or the use of dual-echo and multi-echo sequences to estimate motion artifacts based on their differential dependency on echo time compared to BOLD signal (Kundu et al., 2012; Bright and Murphy, 2013). However, next to the attempts for removal of motion artifacts, researchers should keep in mind that motion most likely has a neural basis (Pujol et al., 2014; Zeng et al., 2014). This means that for instance BOLD fluctuation within the motor cortex can potentially be correlated with motion of the participant but should not be removed from the data.

The results presented in Chapter 2 and 3 do indicate the high utility of ICA for providing effective denoising of functional MRI data, as also shown in the work on the evaluation of ICA-FIX (Griffanti et al., 2014; Salimi-Khorshidi et al., 2014). ICA-FIX is a method for removing all types of structured noise but which requires re-training of the method when applying it to new datasets because of poor generalizability of the FIX classifier, as also shown in Chapter 3. I addressed the issue of generalizability by developing and tailoring an ICA-based method specifically towards motion artifacts using a small set of intuitive and standardized features. However, there exist a considerable amount of other types of structured noise, roughly comprising another 1/3 of the explained variance (see Chapter 2). In line with this thought, new ICA-based features and classifiers can be developed which specifically focus on other types of artifacts, possibly integrated with the ICA-AROMA classifier. Importantly, such a method could be applied at different stages of the preprocessing pipeline; e.g. ICA-based removal of scanner artifacts might be best performed prior to any preprocessing. Such features should specifically capture temporal and spatial characteristics of the targeted artifacts. Box 1 provides some ideas and suggestions on how such features could be designed.
Box 1 – Potential features for non-motion noise components

**MR-scanner artifacts**
Box 2 of the introduction of this thesis already discussed a few common MR-scanner artifacts such as spikes, ghosting and susceptibility artifacts. Spike artifacts typically appear as stripes across an fMRI image, ghost artifacts present as low-intensity shifted brain images along the central axis in the phase-encoding direction, while susceptibility artifacts appear as geometric distortions at tissue-air boundaries. These are all very specific spatial properties that could be quantified into features to automatically detect the components representing these artifacts. Such spatial features could for instance comprise 1) a spatial frequency feature to evaluate stripes, and/or a features assessing slice-specific intensities, in the component and detect spike artifacts, 2) the fraction of all voxel Z-values in the component represented inside versus outside the expected location of the ghost image and 3) the fraction of all voxel Z-values located within a priori defined air-tissue template mask in a reference space to detect susceptibility artifacts, analogous to the edge and CSF features in ICA-AROMA. Moreover, the power spectra of these artifacts will typically contain more high frequencies (i.e. >0.10 Hz) compared to signal components. In addition, ghost and susceptibility artifacts can interact with head motion, suggesting that a similar temporal feature as in ICA-AROMA can be used to assess the temporal correlation of the component time-series with a set of realignment parameters.

**Physiological noise**
Analogue, we can identify spatial-temporal characteristics of physiological artifacts which can be exploited for designing features. Temporally, artifacts related to cardiac and respiratory rates will mainly contain high frequencies, aliased into the low-frequency domain (Murphy et al., 2013). Spatially, the artifacts will be most profoundly located around large arteries, e.g. cerebral arteries around the brainstem. On the other hand, physiological noise related to veins typically exhibits low-frequencies in the power-spectrum and is located near large veins, e.g. sagittal and transverse sinus (Salimi-Khorshidi et al., 2014). Notably, new parallel imaging sequences (e.g. multiband imaging) allow to scan at much higher temporal resolution (i.e. shorter TR) which reduces the aliasing of cardiac and respiratory signals into the low-frequency domain, potentially increasing the performance of discriminating these artifacts from signal components (Griffanti et al., 2014). As an alternative to such data-driven estimation of the temporal dynamics of physiological noise, it has been proposed to record cardiac and respiratory rate during the scanning procedure to define physiological nuisance time-series, which can be regressed from the data (Glover et al., 2000; Birn et al., 2006). Possibly, such nuisance time-series can also be used to detect independent components related to physiological noise, e.g. by correlating the nuisance time-series with the component time-series. However, this will require obtaining cardiac/respiratory measurements and highly rely on the quality of the measurements, which might therefore not be the optimal method for the purpose of ICA-based denoising.
Modeling functional brain architecture

To study functional brain architecture using rfMRI data researchers assess temporal dependencies of spontaneous BOLD fluctuations between different brain areas (Fox and Raichle, 2007; referred to as functional connectivity (Friston, 1994). As explained in the introduction there are roughly two ways to assess functional connectivity, using seed-based correlation analysis (SCA) or ICA + dual regression. In this section I will discuss some limitations of these approaches and present recent advances in modeling functional brain architecture that are emerging to overcome these limitations.

For clarity of this discussion I will first present a general network modeling framework in which we define brain regions as ‘nodes’ and functional connections between these regions as ‘edges’. The modeling of functional brain architecture then consists of three steps: 1) define nodes, 2) extract a single time-series of every node, 3) assess connectivity between nodes. See Figure 2 for a schematic illustration of these steps. When fully parceling the brain into a large set of nodes and assessing all edges between them, the resulting whole-brain connectivity profile reflects the so-called the ‘functional connectome’ (Smith et al., 2013).

SCA and ICA + dual regression can be considered to be a variant of this framework. In typical SCA analysis, functional connectivity between a single node and all other voxels is determined, whereas in typical ICA + dual regression analysis functional connectivity of all voxels within a large-scale network is determined. Accordingly, both SCA and ICA + dual regression typically assess functional connectivity ‘within’ a network rather than estimating a whole-brain functional

Figure 2 - Schematic overview on the different steps in network modeling (adapted from (Smith et al., 2013))
connectome. Importantly, SCA can also be applied by defining a large set of anatomical nodes and determining the correlation between every pair of nodes to estimate a whole-brain functional connectome, following the framework as presented in Figure 2. ICA however, is typically employed at low-dimensionality to estimate approximately 10 large-scale functional networks (Beckmann et al., 2005; Damoiseaux et al., 2006; Smith et al., 2009). Although network time-courses obtained in the first stage of the dual regression procedure can be used to investigated the temporal dependency between the networks as a whole; this large-scale approach does not allow more refined assessment of temporal dependencies between more localized regions.

Importantly, advances have been made at each step in the framework presented in Figure 2. These advances include more accurate and biologically appropriate node definitions, more robust time-series extraction and more specific connectivity measures. Box 2 describes these developments in more detail. Optimizing such modeling will increase the robustness and reliability of estimated functional architecture and therefore improve sensitivity in our research towards detecting underlying pathophysiology of heterogeneous mental disorders.

Next to such improved whole-brain network modeling, there have also been additional developments regarding the characterization of large-scale networks. Up to now, large-scale networks are predominantly obtained using spatial ICA (Beckmann et al., 2005; De Luca et al., 2006; Smith et al., 2009). This method obtains a set of networks (spatial maps) that are maximally spatially independent. However, given the notion that large-scale neural dynamics might share anatomical infrastructure, it is questionable whether spatial independency is the most appropriate feature to identify distinct modes of brain activity (Friston and McKeown, 1998). An interesting direction is to further relax the constraint of spatial independency and identify independent spatiotemporal components using temporal ICA (Smith et al., 2012). Temporal ICA is not often considered in fMRI analyses due to poor temporal resolution, but new imaging sequences now open up new opportunities to move towards such more advanced temporal analyses (Griffanti et al., 2014). Research by Smith et al. has for instance identified several independent temporal components/modes with high spatial overlap; referred to as ‘functional modes’ (Smith et al., 2012). Interestingly, these functional modes share some relationship with resting-state networks (e.g. subdivisions of these networks) but are in fact quite different and therefore might have greater biological interpretation.
Box 2 – Optimization of modeling functional brain architecture

Define nodes/network
One of the main challenges in SCA analyses and network modeling based on nodes is that they highly depend on a priori definitions of nodes. Typically, in SCA analyses researchers define these regions as spheres located at anatomical coordinates or alternatively have been using anatomical regions obtained by anatomical parcellation of the brain (e.g. the AAL atlas). However, it is questionable whether anatomical and functional brain regions align; such inaccurate definitions of regions of interest can have considerable analytical impact (Fornito et al., 2013). Accordingly, alternative approaches have been proposed which aim to parcellate the brain into non-overlapping functional regions using resting-state fMRI data, for instance based on spectral clustering (Craddock et al., 2012) or instantaneous correlations (van Oort et al., 2016).

Importantly, these methods introduce ‘hard’ parcellations with clear-cut boundaries and where each voxel is assigned to a single parcel, whereas we known that voxel time-series are a mixture of multiple neural/artifactual signals (Smith et al., 2012; Fornito et al., 2013). A proposed alternative is the use of ‘soft’ parcellation by exploiting high-dimensionality ICA to identify localized nodes with coherent temporal activity, instead of large-scale networks obtained when employed ICA at low dimensionality (Kiviniemi et al., 2009; Smith et al., 2013). In contrast to ‘hard’ parcels these ‘soft’ parcels are represented by a weighted spatial map and can spatially overlap; improving robustness by accounting for potential overlap between brain regions, small misalignment between the group-level node and the individual node and modeling of artifactual sources.

Extract time-series
From every node, its representative time series needs to be extracted. In the case of hard parcels, the average time series across all voxels is most typically determined. Alternatively, the first eigenvariate can be extracted which represents the dominant ‘mode’ of varying signal intensity and might therefore be more robust against artifacts or misalignment of the parcel. In case of soft parcellation a multivariate spatial regression can be employed (equal to the first stage of the dual regression) to obtain its representative time-series.

Analogous to using the eigenvariate rather than the average time-series in case of hard parcels, it has been proposed to use ‘eigenregression’ time-series rather than first stage dual regression to estimate time-series of soft nodes, to improve robustness of the obtained time-series (Smith, 2014). Eigenregression obtains the first eigenvariate of the fMRI data, temporally variance normalized and weighted by the node’s spatial map, after spatially regressing out all other nodes. Although simulation has indicated that this might improve robustness of the estimated functional connectivity, this method has yet to be extensively evaluated (Smith, 2014).

Estimate connectivity
To assess functional connectivity between nodes, typically the Pearson correlation between node time-series is determined. Importantly, this measure cannot be exploited to assess causality or directionality of the connection/edge between nodes; referred to as ‘effective connectivity’ (Friston, 1994). Moreover, by defining ‘fixed’ spatial templates of functional networks and/or obtaining a single connectivity measure for
each edge, one assumes that functional connectivity is stationary (i.e. consistent over time). However, evidence is accumulating that the strength of functional connectivity might vary over time and is therefore non-stationary (Hutchison et al., 2013).

To investigate whether two nodes are directly connected or indirectly connected via another node, one can assess the partial correlation between the nodes’ time-series (Smith et al., 2011). Here, the correlation is assessed between the ‘unique’ time-series of the two nodes, obtained by regressing out the time-series of all other nodes from each of the two nodes’ time-series. Moreover, studies start to emerge which investigate causality and non-stationarity with such networks using resting-state fMRI data, which might hold promising insights on the functional architecture of the brain in the near future (Arnsten et al., 2010; Friston et al., 2011; Smith et al., 2011; Hutchison et al., 2013).

Multi-modal integration

In the current thesis I focused on the functional rather than structural architecture of the brain. However, brain structure and functioning is hypothesized to be intrinsically multimodal (Zatorre et al., 2012). Accordingly, structural and functional indices derived from different neuroimaging modalities are expected to be linked and share a high degree of common variance (Sui et al., 2014). Rather than investigating these modalities separately, effective modeling of such common variance might offer new opportunities for studying diseases and neurodevelopment (Calhoun et al., 2009; Groves et al., 2011; Sui et al., 2012). For example, data-driven multivariate modeling using a method referred to as linked-ICA has been shown to be able to detect such, biologically meaningful, multimodal patterns between imaging modalities (Groves et al., 2011; Douaud et al., 2014). Importantly, next to global patterns across modalities, this approach also identified localized and modality-specific components. The latter is important since we can expect certain biological mechanisms as well as artifacts to be more localized or related to specific imaging domains (e.g. distinct components related to grey and or white matter).

Till now, linked-ICA has been predominantly used for structural research by combining measures on for instance gray matter volume, cortical thickness, areal expansion and white matter characteristics (Groves et al., 2012; Douaud et al., 2014; Francx et al., 2016). As an example, Francx et al. (2016) exploited linked-ICA using the previously listed measures for an integrated analysis of gray and white matter alteration in ADHD using the NeuroIMAGE datasample which has also been used in the current thesis. This study identified two multimodal components that related to ADHD severity, suggesting that differential aspects of brain structure share underlying pathophysiology and illustrating the benefit of multimodal integration in brain research.
However, next to such anatomical/structural assessment, multimodal ICA can be generalized to other domains. For instance we know that functional and structural brain architecture are highly heritable (Thompson et al., 2013), i.e. associated with genetics. Accordingly, multimodal ICA can be exploited to investigate covarying patterns between functional and structural architectural features, and possibly genetic features to characterize function-structure-genetic associations in a explorative and data-driven fashion (Groves et al., 2011; Sui et al., 2012). This idea can be further extended by associating these multimodal covarying patterns with cognitive and demographic measures, either by post-hoc assessment or by including them as additional modalities in the multimodal ICA analysis. Next to ICA, alternative methods such as canonical correlation analysis and partial least squares can be exploited to investigate such cross-domain association analysis (Sui et al., 2012). The potential of such analyses have already been shown in several studies (Calhoun et al., 2009; Meda et al., 2010; Liu et al., 2012). A prime example being a recent study by Smith et al. (2015) where a strong mode of covariation across the population was found, using canonical correlation analysis, between the functional connectome and lifestyle, psychometric and demographic measures. Accordingly, similar analyses could be utilized to investigate such associations across domains in mental disorders (e.g. imaging, cognition and behavior).

**Conceptualization of mental disorders**

In the current thesis I proposed to model the heterogeneity of mental disorders (specifically ADHD) by incorporating the notion that distinct and independent neurobiological mechanisms can underpin a similar behavioral phenotype (Makris et al., 2009; Durston et al., 2011; Castellanos and Proal, 2012; Cortese et al., 2012; de Zeeuw et al., 2012) and might have distinct categorical-dimensional complexity (Chabernaud et al., 2011; Elton et al., 2014). Although this method advances upon two trends in current conceptualization of mental disorders it still has considerable limitations. Most importantly, the method is a ‘top-down’ approach that is dependent on behavioral variables (e.g. DSM-diagnosis) and therefore highly relies on the accuracy and validity of these measures. The interpretation of a categorical/dimensional measure might for instance depend on its context, e.g. a dimensional score of anxiety has a different interpretation when occurring in a context of psychotic episode or inattentive learning problems (Nigg, 2015). Moreover, the use of a categorical DSM-diagnosis variable is insufficient when we assume that behaviorally defined DSM-categories don’t align with underlying categorical mechanisms.

Accordingly, data-driven approaches are arising that aim to identify homogeneous subgroups (clusters) of patients (Fair et al., 2012; Karalunas et al.,
2014). Although these approaches can reveal interesting underlying heterogeneity, they have methodological difficulties, e.g. determining the optimal number of clusters, clustering method and selecting appropriate features (Marquand et al., 2016). Furthermore, by assuming the existence of discrete subgroups they don’t incorporate the notion that patients in fact might represent the extremes on a continuum. Moreover, similar to many classification studies, such approaches often search for subgroups within a DSM-defined clinical group and therefore still rely on behaviorally defined boundaries.

Investigating latent categorical-dimensional mechanisms without relying on (DSM-defined) behavioral boundaries will require new data-driven methodology. Dimensional mechanisms would ideally be identified by investigating normal variation in a population whereas categorical mechanism could be investigated in individuals that deviate from this normal distribution. Interestingly, Marquand et al. (2016) have proposed such a principled framework in which normative modeling is utilized to investigate relationships between symptom/behavioral dimensions and biological domains. First, a normative model is estimated that links clinical and biological variables using Gaussian process regression. Subsequently, the deviation from this normal distribution can be assessed for specific individuals. This allows to distinguish between individuals which have similar behavioral symptoms but are respectively located at the extreme of the normal distribution or can be considered outliers with a more individualized pattern of abnormality. This framework could be utilized to on the one hand identify (dimensional) bio-behavioral mechanism by evaluating the estimated normative distributions, while on the other hand the outliers can be post-hoc investigated for categorical effects. However, to reliably estimate the normative model, substantial amounts of data from a random (healthy) population are required; it is still a matter of debate whether such a normative model should be estimated on only healthy subjects or should also include clinical groups (Marquand et al., 2016).

Although these increasing efforts will provide more understanding of the neurobiology related to the healthy and clinical population, it still remains an open question how these insights should be incorporated into new nosology and be translated to clinical use. Importantly, for clinical purpose, findings on neurobiological underpinnings of mental disorders should be investigated for their association with treatment response and developmental trajectories (i.e. expected persistence, exacerbation or remittance of symptoms). Moreover, clinical practice will still heavily depend on categorical decision-making, e.g. a decision to start a treatment or not. Accordingly, when an individual presents him or herself to a clinician, assessing his/her deviation with respect to a normal distribution of multiple bio-behavioral dimensions might reveal to which specific neurobiological
mechanism(s) his/her problems might relate. This knowledge can inform towards more individualized diagnosis, treatment selection or prognosis. However, this will still require definitions of (multidimensional) cut-points. Accordingly, although there seems broad consensus on the added clinical value for (neuro)biological measures in psychiatry; how such biologically informed nosology and clinical decision making should be shaped is still a topic heavy debate (e.g. (Regier et al., 2009; Rutter, 2011; Cuthbert, 2014; Nigg, 2015; Weinberger et al., 2015)).

Concluding remark

In this thesis I have presented new methodology to optimize the modelling of neural networks for investigating mental disorders. By preprocesing rfMRI data using ICA-AROMA (Chapter 1) we can now more reliably model different functional networks in the absence of head-motion induced noise (Chapter 2) and investigate these networks for categorical-dimensional mechanisms using the group-level framework developed in Chapter 3. These advances have opened promising opportunities for unravelling the heterogeneous pathophysiological underpinning related to mental disorders such as ADHD (Chapter 4).

References

Bright, M.G., Murphy, K., 2013. Removing motion and physiological artifacts from intrinsic BOLD fluctuations using short echo data. Neuroimage 64, 526–537.
Summary and general discussion


Summary and general discussion

from resting-state fMRI time series. Neuroimage 95, 287–304.


van Oort, E.S.B., Mennes, M., Kumar, V.J., Grodd, W., Beckmann, C.F., 2016. Human brain parcellation using time courses of instantaneous correlations.


Appendix
Binnen de psychiatrie is er een beweging naar nieuwe conceptualisaties van psychiatrische stoornissen. Allereerst wordt er toenemende mate gedacht dat personen die gediagnosticeerd worden met een psychiatrische stoornis het ‘eind van een continuüm van normaal gedrag’ vertegenwoordigen in plaats van een groep individuen die systematisch verschilt van ‘normale’ personen (Coghill and Sonuga-Barke, 2012). Bovendien rijst er een ‘neurale-systemen’ visie op ten aanzien van psychiatrische stoornissen, waarbij wordt verondersteld dat er afzonderlijke neurale problemen onderliggend kunnen zijn aan eenzelfde soort disfunctionerend gedrag; met andere woorden, het bestaan van afzonderlijke (categorische) subtypes van patiënten of (dimensioneel) disfunctionerende biologische systemen (Makris et al., 2009; Menon, 2011). Een beeldvormingsprotocol genaamd resting-state fMRI (rfMRI) heeft grote potentie betreffende onderzoek naar deze aspecten op het vlak van de functionele architectuur van het brein, aangezien deze techniek het mogelijk maakt tegelijkertijd meerdere (afzonderlijke) neurale netwerken te bestuderen.

Ondanks deze veelbelovende ontwikkeling betreffende het leren begrijpen van pathofysiologisch mechanismen die onderliggend zijn aan psychiatrische stoornissen, wordt het psychiatrische onderzoekveld geconfronteerd met grote heterogeniteit binnen en tussen studies (Castellanos et al., 2009; Oldehinkel et al., 2013; Vargas et al., 2013; Mulders et al., 2015). Dit is het gevolg van ontoereikende en/of suboptimaal gebruik van methodologie betreffende de neurale-systemen en categorisch-dimensionele conceptualisatie van psychiatrische stoornissen. Resting-state fMRI onderzoek wordt bovendien geconfronteerd met schadelijke effecten van gestructureerde verstoringen in de rfMRI data die veroorzaakt wordt door hoofdbewegingen van de deelnemer tijdens de scanning-procedure. In dit proefschrift heb ik nieuwe analytische methoden ontwikkeld om deze nadelen te aanpakken en daarmee onderzoek naar de associatie tussen functionele brein architectuur en psychiatrische stoornissen te optimaliseren. Ik heb de ontwikkelde methodologie toegepast om de functionele brein architectuur van aandachtstekort-hyperactiviteitsstoornis (ADHD: attention-deficit/hyperactivity disorder) te onderzoeken; een klinkend voorbeeld van een stoornis waarin deze nadelen samenkomen.
Verwijderen van verstoringen in resting-state fMRI data

Het is al geruime tijd bekend dat hoofdbewegingen van een persoon tijdens het maken van een fMRI-scan tot grote verstoringen in de data kan leiden (Friston et al., 1996). Sindsdien wordt voor deze verstoringen gecorrigeerd door elk scan-volume te heroriënteren zodat alle volumes op een lijn liggen, en bovendien tijdsafhankelijke maten van deze heroriëntaties uit de data te filteren. Echter, in 2012 toonden drie publicaties, van Van Dijk et al., (2012), Power et al. (2012) en Satterthwaite et al. (2012), aan dat maten voor functionele connectiviteit verkregen uit rfMRI ondanks deze conventionele correcties nog sterk verstoord werden hoofdbewegingen. Dit leidde tot een levendig debat in de rfMRI wereld over de oorzaken en methodes voor de verwijdering van deze verstoringen (Wilke, 2012; Yan et al., 2013a, 2013b, Satterthwaite et al., 2013a, 2013b; Beall and Lowe, 2014; Couvy-Duchesne et al., 2014; Zeng et al., 2014; Kong et al., 2014; Muschelli et al., 2014; Patel et al., 2014; Pujol et al., 2014; Scheinost et al., 2014; Spisák et al., 2014). De methode die de meeste aandacht kreeg is genaamd ‘scrubbing’ waarbij scan-volumes, op tijdstippen dat het individu (te) veel heeft bewogen, uit de data worden verwijderd.

De ‘scrubbing’ methode is sterk afhankelijk van de eerdergenoemde heroriëntatie parameters, vernietigd autocorrelatie in de data, en zorgt voor een sterk en variabel verlies van vrijheidsgraden. Dit heeft sterk nadelige consequenties voor verdere statistische analyses op de data. Daartoe heb ik in Hoofstuk 1 van dit proefschrift een alternatieve methode ontwikkeld en voorgesteld, genaamd ‘ICA-AROMA’. Deze methode maakt gebruik van een mathematische techniek (ICA; independent component analysis) die aanneemt dat de data een mix is van verschillende bronnen en deze onderliggende bronnen kan karakteriseren. ICA-AROMA maakt gebruik van deze techniek door de ICA toe te passen op de fMRI-data, vervolgens van elke gedetecteerde bron te bepalen of deze bron ‘verstoring’ of ‘signaal’ reflecteert en de verstoringsbronnen vervolgens uit de data te filteren. Om een accuraat onderscheid te kunnen maken tussen ‘verstoring’ en ‘signaal’, die tevens valide is voor meerdere type datasets, heb ik vier robuuste maten ontworpen. Daarnaast heb ik in dit hoofdstuk een validatiestudie gedaan die aantoont dat ICA-AROMA de verstoringen met hoge nauwkeurigheid detecteert/verwijderd.

In Hoofstuk 2 heb ik vervolgens een evaluatiestudie verricht, waarbij ik ICA-AROMA heb vergeleken met een reeks andere methodes om verstoringen uit de data te filteren. In deze studie heb ik van elke methode beoordeeld, voor meerdere maten voor functionele connectiviteit, in hoeverre het betekenisvolle signaal wordt behouden, hoe reproduceerbaar het signaal is (maat voor betrouwbaarheid), de mate waarin er nog verstoringen aanwezig zijn en de mate van verlies in vrijheidsgraden.
Bovendien heb ik deze analyses gerepliceerd over vier verschillende datasets om zodoende de generaliseerbaarheid van de methoden over verschillende type datasets te beoordelen. In lijn met eerdere studies toonde deze evaluatie allereerst aan dat de conventionele methode, door de data enkel te filteren aan de hand van heroriëntatie parameters, inderdaad ondermaats is betreffende het verwijderen van verstoringen. Dit onderstreept nogmaals het belang van het ontwikkelen van nieuwe methodes. Met betrekking tot ICA-AROMA toonde de studie aan dat de methode vergelijkbaar is met de 'scrubbing' methodes betreffende de mate van het verwijderen van verstoringen. Echter, ICA-AROMA behoudt het betekenisvolle signaal in de data in grotere mate en leidt tot een minder groot/variabel verlies in vrijheidsgraden. Bovendien generaliseerde ICA-AROMA over verschillende datasets, wat aantoont dat de methode robuust en daarmee breed toepasbaar is.

**Conceptualisatie van psychiatrische stoornissen**

Binnen de psychiatrie wordt in toenemende mate gedacht dat we toe moeten naar categorisch-dimensionele benadering voor het conceptualiseren en onderzoeken van psychiatrische stoornissen (Chabernaud et al., 2011; Coghill and Sonuga-Barke, 2012). Waarbij categorische effecten systematische verschillen tussen patiënten en controles betreffen, en dimensionele effecten een continuüm van een bepaalde maat/fenotype representeert. Echter, in de huidige literatuur waarbij zowel categorische als dimensionele maten worden bestudeerd, lijkt er maar een beperkt bewustzijn te zijn betreffende de methodologische implicaties van het uitvoeren van dergelijke analyses. **Hoofdstuk 3** poogt dit bewustzijn te vergroten door toe te lichten dat, door de hoge correlatie tussen categorische en dimensionele maten, de typische gebruikte statistische modellen de kans op vals positieve bevindingen vergroot en de sensitiviteit (kans op correcte bevindingen) verminderd (York, 2012; Mumford et al., 2015). Om deze beperkingen te overkomen stel ik in dit hoofdstuk een alternatief statisch raamwerk voor waarbij gebruik wordt gemaakt van 'selectieve orthogonalisatie' van variabelen. Als voorbeeld heb ik het raamwerk toegepast bij een studie naar functionele connectiviteit binnen het 'executieve controle netwerk' bij ADHD. De resultaten toonden aan dat huidige benaderingen inderdaad leiden tot verminderde sensitiviteit en foutieve interpretatie van verkregen resultaten. Het nieuwe raamwerk daarentegen leidde tot een sterker verfijnde characterisatie van de resultaten en verhoogde sensitiviteit.

Naast de categorische-dimensionale benadering van psychiatrische stoornissen wordt er ook in toenemende mate gedacht dat stoornissen multifactorieel zijn in hun onderliggende pathologieën, waarbij meerdere onafhankelijke neurale systemen betrokken kunnen zijn (Makris et al., 2009; Menon, 2011). Verschillende pathofysiologische mechanismen kunnen hierbij leiden tot soortgelijke symptomen.
Appendix

Resting-state fMRI heeft de mogelijkheid dergelijke neurale systemen te onderzoeken.

**Functionele connectiviteit in ADHD**

ADHD is een klinkend voorbeeld waarbij alle bovengenoemde aspecten bijeenkomen. Er wordt verondersteld dat er categorische-dimensionele onderliggende neurale mechanismen zijn, dat er meerdere neurale netwerken bij betrokken zijn en de symptomen van hyperactiviteit veroorzaken verhoogde mate van verstoringen in de rfMRI data. Daartoe heb ik in het Hoofdstuk 4 van dit proefschrift alle ontwikkelde methodologie geïntegreerd en een studie uitgevoerd naar functionele connectiviteit in ADHD waarbij ik rfMRI, opgeschoond met ICA-AROMA, heb gebruikt om categorisch-dimensionele mechanismen te bestuderen in verschillende neurale systemen. Tevens heb ik de resultaten gerelateerd aan ADHD-gerelateerde cognitieve en genetische maten om zodoende de neurobiologie van ADHD verder in kaart te brengen. De studie resulteerde in een reeks significante bevindingen, verspreid over verschillende netwerken zoals het default mode, executieve controle, cerebellum, visuele en motor netwerk. De anatomische locaties van deze bevindingen komen overeen met eerdere bevindingen in de literatuur. Echter, de huidige studie voorziet in een meer gedetailleerde karakterisatie van deze bevindingen door deze te associëren met verschillende netwerken en de categorisch-dimensionele complexiteit te identificeren. Tevens waren enkele bevindingen specifiek geassocieerd met bepaalde cognitieve/genetische maten. Daartoe, door de neurale-systemen benadering, het uit elkaar trekken van categorisch-dimensionele effecten en het relateren van de bevindingen met cognitie en genetica heeft deze studie vernieuwde inzichten gegeven in de neurobiologische mechanismen van ADHD.

**Conclusie**

In dit proefschrift heb ik nieuwe methodologie gepresenteerd voor het optimaliseren van het modeleren van neurale netwerken voor het bestuderen van psychiatrische stoornissen. Door de rfMRI data voor te bewerken met ICA-AROMA (Hoofdstuk 1) kunnen we nu met grotere betrouwbaarheid functionele netwerken modeleren, zonder invloed van verstoringen door hoofdbewegingen (Hoofdstuk 2), en categorische-dimensionele mechanismen binnen deze netwerken bestuderen door gebruik te maken van een alternatief statistiek raamwerk (Hoofdstuk 3). Deze nieuwe methodologie geeft veelbelovende mogelijkheden voor het ontrafelen van heterogene pathofysiologie onderliggend aan psychiatrische stoornissen zoals ADHD (Hoofdstuk 4).
En toen.... het dankwoord. De bijdrage van naasten/betrokkenen betreft een dimensionele maat waarbij het niet meevalt om de verschillende type bijdragen te karakteriseren en hierbij signaal van ruis te onderscheiden betreffende de bijdrage aan het daadwerkelijke eindproduct. In dat opzicht verschilt het dankwoord dan ook geenszins van dit gehele proefschrift waarbij ik categorische-dimensionele mechanismen heb proberen te ontrafelen in verschillende neurale systemen terwijl die verrekte hoofdbewegingen voor ruis zorgden. Dus ik heb niet de illusie in deze enkele pagina’s een volledige en accurate decompositie te kunnen geven van de bijdragen aan dit proefschrift en vervolgens over elk aspect mijn dank uit een te zetten. Echter, ik zal een poging wagen, hopende daarmee zoveel mogelijk variantie te verklaren in deze enkele pagina’s. En ja, voor de Bêta’s onder jullie, ik heb geprobeerd deze verklaarde variantie zo hoog mogelijk te krijgen door de bijdragen te kwantificeren in verschillende maten, deze te normaliseren en hier een PCA op los te laten om zodoende de eerste paar componenten te selecteren en dit dankwoord zo effectief mogelijk te kunnen schrijven. Ik vrees echter dat het verstandiger is om deze Bêta-kant even te parkeren en me te beperken tot de wondere wereld van de lettertjes.

Een component die aanzienlijk veel variantie verklaart en daarmee, wat mij betreft, als eerste principle component genoemd mag worden is Maarten. Wat een gedrevenheid en geduld. Het valt niet mee signaal van ruis te onderscheiden als Raimon elke week met een stapel papieren vol resultaten aan je bureau staat, gelukkig heeft Dropbox gaandeweg de bossen gered! Toch kon ik altijd bij je binnenstappen, was je altijd behulpzaam, constructief, enorm effectief (en overzichtelijk) in het tot stand brengen van je scripts en wist je me bij tijd en wijle een tik op de vingers te geven en analyse-restricties op te leggen. Ook je bijdragen en correcties aan de teksten waren zeer leerzaam. Bovendien waren de figuren aanzienlijk minder aligned als ij er niet was geweest. Serieux, je bijdrage is niet te onderschatten en ik ben je daar oprecht dankbaar voor.

Jan, respect voor hoe je ondanks de drukte alle projecten georganiseerd weet te houden, richting aan de onderzoeken weet te geven en onderzoekers enthousiasmeert. Hierbij heb je enerzijds genoeg creatieve vrijheid gegeven maar wist je bovendien met je kennis het onderzoek kritisch te benaderen, te concretiseren, goed te kaderen naar het huidige onderzoeksveld en met creatieve nieuwe invalshoeken te komen. Ik heb respect voor hoe je dit alles weet te bewerkstelligen, dank voor de mogelijkheden die je hierin hebt gegeven!
Christian, without a doubt, you taught me a lot. Although it might sound a bit vague and straightforward, but in my opinion I learned a lot from you on how to properly view and handle data, and actually do something useful and truly valid with it. How to properly model data and come up with sensible measures, critically analyse proposed methods/measures, and make sure the research clearly increases the signal to noise ratio in literature. I realize that I use many of these concepts and ways of thinking in my analyses, reading and reviewing nowadays. Especially when developing ICA-AROMA we have had many meetings and you always took the time to explain me some core concepts. I am truly grateful for that. Importantly, you did all of this with a smile and a funny note. Probably without you even knowing, you really put my working memory to the test every time my head was full of questions when entering your office but we first had to discuss pony's, cupcakes, etc. Thanks again for how approachable and helpful you have been.

Next to my (co-)promotors I would of course like to thank all of my colleagues at the Donders Institute. I won’t go by all of them by name. But a special thanks to the members of the SIN-group; the group has expanded during the years up to what is a very efficient and high-quality group. It was very interesting to follow the discussions in the group and get to know everybody’s way of thinking, I learned a lot from that! Also a big thanks to the members of the NeuroIMAGE-group, with whom we had the weekly WIP-meetings. Everybody had their own style and research approaches which were often very interesting to hear about; it’s seems like everybody had quite an own perspective on ADHD and associated problems/research which was inspiring. And last but definitely not least, also a big thanks to all the colleagues at the ICT and administration, you all do a crucial and wonderful job in keeping everything and everybody aligned and prevent the institute from collapsing, thanks a lot!

Verder wil ik graag mijn vrienden en familie in het algemeen bedanken. Jullie weten dat het energie heeft gekost om dit alles voor elkaar te krijgen, vooral het afronden van het proefschrift. Met name mijn lieve en betrokken familie (Chantal, Ilse, Jeffrey, Ma, Pa, Wim; lekker politiekcorrect op alfabetische volgorde :P) heeft door de jaren heen soms moeten lijden onder mijn klagen, mijn te lange verstopte houding achter de laptop, en soms iets te enthousiaste en uitgebreide verhalen over het onderzoek. Ik zal jullie niet allemaal bij langs gaan, maar natuurlijk ook dank aan de vrienden die om me heen staan voor de TG-uitjes, klaverjasavonden, Finse ontspanning, etc., Echter wil ik Matthijs toch bij naam noemen. Jij hebt overduidelijk veel te verdragen gehad in dit hele traject; als ik bijvoorbeeld ‘even wat jobs moet submitten’ en vervolgens urenlang van de wereld was. Jij hebt me
Ik zal me ervan weerhouden reclame te maken. Maar het betreft een dankwoord en dan kom ik er toch niet onderuit om ook de fabrikant van mijn hardloopschoenen/kleding te bedanken, net als een onlinedienst voor het kijken van films/series on demand. Beide hebben namelijk voor aanzienlijke en belangrijke ontspanning gezorgd waarbij de eerste tevens een niet te onderschatten bron van inspiratie heeft gefaciliteerd.

Dan, last but not least. Een woord van dank aan de inspirerende en enthousiasmerende docenten. Van basisschool, tot middelbare school tot universiteit. Ik lees dit niet vaak in het dankwoord, maar voor mij een enorme factor in het bewerkstelligen van een loopbaan in de wetenschap en mijns inziens van een niet te onderschatten belang voor het onderwijs in het algemeen. Er zijn enkele momenten en docenten geweest die met hun persoonlijke aandacht, betrokkenheid, toegankelijkheid en bovenal een passie voor wetenschap en extra inzet zeer stimulerend en enthousiasmerend voor mij zijn geweest. Daarbij denk ik bijvoorbeeld aan Peter de Wilde die een masterclass organiseerde, die mij vrijelijk een modeleeropdracht liet ontwerpen en waarmee ik zeer spontane en interessante gesprekken kon voeren. En Arjan Pruim (ja, familie van, maar nee, zeker geen bias!) die als enthousiast Natuurkunde-docent het vak continu vernieuwend en aantrekkelijke wist te houden (al is natuurkunde an sich natuurlijk al fascinerend genoeg). Als ik terugdenk aan de middelbareschooltijd is dit wat het eerste in me opkomt, ik kon me helemaal verliezen in de extra programma’s die je opzette zoals het modelleren van het klimaat (och wat heb ik lopen spelen met die modellen) en project moderne natuurkunde (hoe fascinerend is het om thuis helemaal uit te mogen vogelen hoe die ‘deeltje-in-een-doosje’ modellen toch in elkaar zitten). Het zijn dergelijke docenten die je in het onderwijs nodig hebt om vuurtjes bij studenten aan te wakkeren zodat zij hun passies ontdekken en ontwikkelen.

Dan is hiermee een eind gekomen aan het dankwoord. Aan het begin dacht ik “pfff, het dankwoord, moet ik nu echt ongemakkelijk iedereen bij langs om ze te bedanken?”. Maar het schrijven van dit stuk liet me wel degelijk realiseren hoeveel mensen er om me heen staan, die allen enerzijds of anderzijds van belang zijn geweest voor het vormen van mij als persoon en onderzoeker. Ik hoop dat dit dankwoord enigszins recht heeft gedaan aan de oprechte dank die ik naar jullie allen toe heb. Dank ;)

Acknowledgement/dankwoord
Appendix
Raimon Henry Ruben Pruim was born in Oldebroek on the 25th of September in 1987. In 2006, he graduated from high school at the Lambert Franckens College in Elburg, finishing course profiles in Nature & Health and Nature & Technique. Given his interest in mathematics, physics and medicine he decided to start a bachelor in Technical Medicine at the University of Twente. He continued his education my studying the master Technical Medicine, where he specialized in Robotics & Imaging and gained a particular interest for computational modelling and brain imaging. This interest was further increased after participating in a Toolkit on Neuroscience. In line with these interests he completed a graduation project at the Donders Institute on functional connectivity in ADHD. He graduated cum laude as a Technical Physician in 2012. Following his graduation work he started a PhD at the Donders Institute at the Statistical Imaging Neuroscience group to work on methodology related to fMRI image processing and psychiatry. Here, he developed a method for removal of motion artefacts from fMRI data and worked on categorical-dimensional modelling of mental disorders, applied to functional connectivity research in attention-deficiency/hyperactivity disorder. Subsequently, he worked as a postdoc on data-driven multimodal neuroimaging analyses in a large population study at the Radiology and Medical Informatics departments at the ErasmusMC. Now, he will start a position as Scientist Innovator at the department of Intelligent Imaging at the Netherlands Organisation for applied scientific research (TNO) where he will be working in the field of computer vision and deep learning.
Donders Graduate School for Cognitive Neuroscience

For a successful research Institute, it is vital to train the next generation of young scientists. To achieve this goal, the Donders Institute for Brain, Cognition and Behaviour established the Donders Graduate School for Cognitive Neuroscience (DGCN), which was officially recognised as a national graduate school in 2009. The Graduate School covers training at both Master’s and PhD level and provides an excellent educational context fully aligned with the research programme of the Donders Institute.

The school successfully attracts highly talented national and international students in biology, physics, psycholinguistics, psychology, behavioral science, medicine and related disciplines. Selective admission and assessment centers guarantee the enrolment of the best and most motivated students.

The DGCN tracks the career of PhD graduates carefully. More than 50% of PhD alumni show a continuation in academia with postdoc positions at top institutes worldwide, e.g. Stanford University, University of Oxford, University of Cambridge, UCL London, MPI Leipzig, Hanyang University in South Korea, NTNU Norway, University of Illinois, North Western University, Northeastern University in Boston, ETH Zürich, University of Vienna etc.. Positions outside academia spread among the following sectors: specialists in a medical environment, mainly in genetics, geriatrics, psychiatry and neurology. Specialists in a psychological environment, e.g. as specialist in neuropsychology, psychological diagnostics or therapy. Positions in higher education as coordinators or lecturers. A smaller percentage enters business as research consultants, analysts or head of research and development. Fewer graduates stay in a research environment as lab coordinators, technical support or policy advisors. Upcoming possibilities are positions in the IT sector and management position in pharmaceutical industry. In general, the PhDs graduates almost invariably continue with high-quality positions that play an important role in our knowledge economy.

For more information on the DGCN as well as past and upcoming defenses please visit: http://www.ru.nl/donders/graduate-school/phd/