Clinical Applications of real-time fMRI neurofeedback training – Premises, Promises, and Pitfalls

David Marc Anton Mehler

This dissertation is submitted for the degree of Doctor of Philosophy

at Cardiff University Brain Research Imaging Centre (CUBRIC)

School of Psychology, Cardiff University, Wales, UK

December 2018

"Science is an ongoing race between our inventing ways to fool ourselves, and our inventing ways to avoid fooling ourselves." -- Saul Perlmutter

"At the heart of science is an essential balance between two seemingly contradictory attitudes--an openness to new ideas, no matter how bizarre or counterintuitive they may be, and the most ruthless skeptical scrutiny of all ideas, old and new. This is how deep truths are winnowed from deep nonsense." -- Carl Sagan

STATEMENT OF PUBLICATIONS

Work presented in this thesis have been accepted for peer-reviewed publications.

Parts of Chapter 1.2 have been published as a review article on power in neuroscience in the *Journal of Neurophysiology* and are based on a book chapter about the replication crisis and open science commissioned by Elsevier:

Algermissen, J.*, **Mehler, D.M.A.* (equal contribution)**, 2018. May the power be with you: Are there highly powered studies in neuroscience, and how can we get more of them? J. Neurophysiol. 2114–2117. doi:10.1152/jn.00765.2017

Mehler, D., 2019. The replication crisis: Is brain imaging next? Chapter 11 in: The dark side of neuroimaging, eds. Amir Raz and Robert Thibault.

Parts of Chapter 2 and 3.1 have been published as a review article on clinical neurofeedback in Psychiatry in the journal *Encephale* (Arns et al. 2017):

Arns, M., Batail, J.-M., Bioulac, S., Congedo, M., Daudet, C., Drapier, D., Fovet, T., Jardri, R., Le-Van-Quyen, M., Lotte, F., **Mehler, D.**, Micoulaud-Franchi, J.-A., Purper Ouakil, D., Vialatte, F., 2017. Neurofeedback: One of today's techniques in psychiatry? Encephale. 43, 135–145. doi:10.1016/j.encep.2016.11.003

Parts of Chapters 4 and 5 are accepted for publication in *Neuropsychopharmacology*:

Mehler, D.M.A., Sokunbi, M.O., Habes, I., Barawi, K., Subramanian, L., Range, M., Evans, J., Hood, K., Lührs, M., Keedwell, P., Goebel, R., Linden, D.E.J., 2018. Targeting the affective brain—a randomized controlled trial of real-time fMRI neurofeedback in patients with depression. Neuropsychopharmacology. doi:10.1038/s41386-018-0126-5

Parts of Chapter 6 are accepted for publication in *NeuroImage*:

Mehler, D.M., Williams, A.N., Krause, F., Lührs, M., Wise, R.G., Turner, D.L., Linden, D.E., Whittaker, J.R., 2019. The BOLD response in primary motor cortex and supplementary motor area during kinesthetic motor imagery based graded fMRI neurofeedback. Neuroimage 184, 36–44. doi:10.1016/j.neuroimage.2018.09.007

Parts of Chapter 7 are adapted from a preregistration protocol available on the Open Science Framework:

Mehler, D.M.A., Williams, A.N., Whittaker, J.R., Krause, F., Lührs, M., Turner D.L., Linden, D.E.J. Gradual real-time fMRI neurofeedback training of motor imagery in middle cerebral artery stroke patients, 2017a. Preregistration and analysis plan available: https://osf.io/qnsv7/

Parts of Chapter 8 have informed an experimental protocol for a trial which has been submitted for journal publication:

Pritchard, A.W., Cooke, A., MacLeod C.A., Bracewell, M., Martin-Forbes, P., **Mehler, D.M.A.**, Linden, D.E.J., Lawrence, C.L., Hindle, J. ENact-PD: EEG Neurofeedback to improve motor function in Parkinson's (under review). Protocol preregistration: http://www.isrctn.com/ISRCTN16783092

Parts of Chapters 1.2 are currently submitted for separate journal publications:

Allen, C.D.A.*, **Mehler, D.M.A.* (equal contribution), 2018**. Open Science challenges, benefits and tips in early career and beyond (under review). PsyArXiv preprints. doi.org/10.31234/OSF.IO/3CZYT.

Parts of the data presented in Chapters 4 to 6 have further been published as conference proceedings of the *International Society of Cerebral Blood Flow and Metabolism (ISCBFM)*, the *Organization for Human Brain Mapping* and the 3rd real-time functional imaging and neurofeedback conference. Other parts of Chapters 1, 3, 4, and 9 are in preparation for separate journal publications. The work presented in this thesis has received the 2017 Young Investigator Award by ISCBFM (Chapters 6 and 7) and the Cardiff University School of Psychology 2nd year PhD prize (Chapters 4 and 5).

SUMMARY

Neurofeedback training represents a form of biofeedback training with a history of over 50 years. During neurofeedback training participants aim to gain control over a feedback signal that represents the activity of a brain region or network of interest. As such, it holds promise for clinical translation as an add-on treatment for psychiatric and neurological conditions. Yet, currently available evidence for its therapeutic efficacy remains limited.

Originally provided based on cortical signals measured with electroencephalography (EEG), methodological developments have allowed providing neurofeedback based on (cortical and subcortical) brain signals acquired from functional magnetic resonance imaging (fMRI). The aim of this thesis was to test the feasibility and clinical efficacy of fMRI neurofeedback (fMRI-NF) training in a psychiatric population and to develop protocols that allow translating the technique to motor rehabilitation. Specifically, this thesis summarises the clinical and neuroimaging results from a randomised controlled trial conducted in patients suffering from depression. Depression represents a leading cause of disability in adults and epidemiological data indicates that up to one third of patients remain depressed after treatment.

Another focus was the development of a motor imagery-based fMRI-NF protocol in healthy participants. This work has informed a proof-of-concept study for motor rehabilitation in stroke survivors, for which the methodology was preregistered on a public platform before data collection started to increase transparency. The thesis aims to address problematic research practices that have been attributed to the replication crisis in many areas of science, including a clear separation of planned and exploratory hypotheses and the use and adaptation of alternative statistical methods. A review chapter discusses potential electrophysiological target signatures for EEG-NF to improve motor symptoms in Parkinson's disease patients. The thesis concludes with a discussion of current premises, promises, and pitfalls in clinical applications of neurofeedback training and considerations for clinical trials development.

ACKNOWLEDGEMENTS

The last four years have been an incredible journey academically, geographically and personally. I would like to thank my supervisors Prof. David Linden and Prof. Richard Wise for giving me the possibility to pursue a PhD at Cardiff University and follow my research interests. I am in particular thankful to David for his continued support, guidance, and encouragement that allowed me to pursue my intellectual interests and form valuable collaborations that were crucial for the work presented in this thesis and beyond.

I am indebted to my closest collaborators in Cardiff, Maastricht, and Bangor, many of whom have become good friends, for lending me their support and expertise: Dr. Joseph Whittaker, Michael Lührs, Dr. Florian Krause, Johannes Algermissen, Dr. Chris Allen, Dr. James Kolasinski, Dr. Angharad Williams, Stefanie Kunas and Aaron Pritchard – it was a great pleasure working with you! I am also deeply thankful to the patients who took part in our research. I thank my colleagues Dr. Mark Postans, Dr. Carl Hodgetts, Dr. Thomas Lancaster, Jude Harrison, Dr. David McGonigle, Dr. John Evans, Prof. Duncan Turner, Prof. Rainer Goebel and Prof. Derek Jones; my mentors Prof. Mike Owen, Prof. Konrad Kording, Dr. Richard Morey and Prof. Chris Chambers; and the hosts of a recent talk series that I gave in the U.S. Prof. Xiaofu He, Mona Rosenke, Prof. Kevin Weiner, Prof. Russell Poldrack, Dr. Mark Mikkelsen, Prof. Nick Puts, Dr. Michal Ramot and Dr. Martin Herbert for their support. Through science communication I've rediscovered my passion writing, and so I'd also like to give thanks to my colleagues at the 'OHBM Comm Committee' Kevin, Nils, Niko and Randy.

Und zuletzt ein besonderer Dank an Familie und Freunde. Danke Mama, dass Du mich immer in meinen Vorhaben voll unterstuetzt und ermutigt hast. Und fuer das Selbstvertrauen und die Neugierde, die Du in mir genaehrt hast – sie haben mich diesen Weg erst gehen lassen. Hoffwn ddiolch i fy mhartner Rhian am ei chariad a chefnogaeth anhygoel dros y flwyddyn diwethaf. Hoffwn hefyd ddiolch i'w theulu – Mair, Tom a Lowri – gyda chi mae Cymru'n teimlo cymaint mwy fel ail gartref! Destek için teşekkür ederim babam, Samet, Talha ve Şafak! Ich danke meinen engsten und vertrautesten Freunden Judith, Raphael, JP, Sophie, Martin and Markus fuer Eure Geduld, sowie das Teilen von Erfolgen und Misserfolgen! Ich danke Nobby & Heinz-Udo fuer Eure Unterstuetzung von Kleinauf! Thank you Schlottmanns, you've inspired me! Lastly, cheers to the 'The Deli' crowd: from 'passing through' to 'break on through', you rock!

DATA COLLECTION AND FUNDING

My research and dissemination activities were supported by Health and Care Research Wales (HS/14/20, grant holder Prof. David Linden) and the BRAINTRAIN consortium, which is supported by the European Commission, under the Health Cooperation Work Programme of the 7th Framework Programme, under the Grant Agreement n° 602186.

The randomised controlled trial presented in Chapters 4 and 5 was supported by the Medical Research Council (grant G 1100629, grand holder Prof. David Linden) and a studentship to Dr. Isabelle Habes from the National Institute for Social Care and Health Research (HS/10/25, grant holder Prof. David Linden). The sampling plan for this trial was designed by The South East Wales Trials Unit (now the Centre for Trials Research). The trial was designed and planned mainly by Prof. David Linden and Dr. Isabelle Habes, the setup was piloted by Dr. Isabelle Habes, Dr. John Evans and Dr. Moses Sokunbi, the data collection was mainly performed by Dr. Isabelle Habes, Dr. Moses Sokunbi and Ms. Kali Barawi. Patient assessment was performed by Prof. David Linden and Dr. Leena Subramanian. I completed the pre-processing of the neuroimaging data and analysed all data presented in this thesis.

Dr. Florian Krause, Michael Lührs, Dr. John Evans and Dr. Cyril Charron helped piloting the setup for the neurofeedback study presented in Chapter 7. Patient recruitment was supported by Ms Bethan Morse and Health and Care Research Wales.

Lastly, I would like to thank Kayla Friedman and Malcolm Morgan of the Centre for Sustainable Development, University of Cambridge, UK for producing the Microsoft Word thesis template that I have adapted to produce this dissertation.

CONTENTS

1 Introduction 15

1.1 ROBUST SCIENCE 17

1.1.1 Replication crisis and questionable research practices 18

1.1.2 Some remedies for (more) robust science 25

1.2 CLINICAL CONDITIONS OF INTEREST 31

1.2.1 Unipolar depression 31

1.2.2 Stroke 43

1.2.3 Parkinson's disease 49

2 Review of clinical real-time fmri neurofeedback applications 57

2.1.1 Principles of learning and transfer 58

2.1.2 Clinical applications of fMRI-NF 60

2.1.3 fMRI-NF in Major Depressive Disorder 62

2.1.4 fMRI-NF for motor rehabilitation in stroke 70

2.1.5 fMRI-NF training in PD 76

2.2 RESEARCH QUESTIONS AND AIMS 79

3 Methods 82

3.1 BASIC PRINCIPLES OF ELECTROENCEPHALOGRAPHY (EEG) 83

3.2 BASIC PRINCIPLES OF FUNCTIONAL MRI 84

3.3 FMRI DATA PRE-PROCESSING AND ANALYSIS 86

3.3.1 fMRI processing 86

3.3.2 fMRI analysis 88

3.3.3 Real-time fMRI setup and processing 91

3.4 STATISTICAL APPROACHES TO HYPOTHESIS TESTING 92

3.4.1 Frequentist hypothesis testing 92

3.4.2 Bayesian hypothesis testing 98

3.4.3 Randomised Clinical Trials 105

3.4.4 The CONSORT statement 108

3.4.5 Trial registration 109

3.4.6 Missing data and data analysis 110

3.5 PREREGISTRATION AND REGISTERED REPORTS 110

4 Clinical effects of fMRI-NF training in Depression 113

4.1 INTRODUCTION 114
4.2 METHODS 115
4.2.1 Patient recruitment 115
4.2.2 Randomisation and masking 115
4.2.3 Trial design 116
4.2.4 Primary (registered) clinical outcome measure 116
4.2.5 Secondary (registered) clinical outcome measure 116
4.2.6 Other (registered) outcome measure 117
4.2.7 Power calculation 117
4.2.8 Statistical tools and planned analyses 117
4.2.9 Exploratory analyses 119
4.3 RESULTS 123
4.3.1 Recruitment, randomisation and termination 123
4.3.2 Attrition, baseline characteristics, and completion rates 125
4.3.3 Primary (registered) clinical outcome measure 126
4.3.4 Secondary (registered) clinical outcome measures 127
4.3.5 Other (registered) outcome measures 130
4.3.6 Exploratory analyses 132
4.4 DISCUSSION 136
5 Neuroimaging Correlates of Neurofeedback Training in Depression 143
5.1 INTRODUCTION 144
5.2 METHODS 145
5.2.1 Participants 145
5.2.2 MR image acquisition 145
5.2.3 Physiological recordings 145
5.2.4 fMRI-NF setup 145
5.2.5 Mental imagery instructions 147
5.2.6 Offline fMRI analysis 147
5.2.7 Comparisons between size of target regions 147
5.2.8 Region of interest analysis 148
5.2.9 Exploratory analyses 148
5.3 RESULTS 152
5.3.1 Target region size 152
5.3.2 Region of interest analyses 153

5.3.3 Exploratory analyses 154

5.4 DISCUSSION 160

6 Motor imagery based neurofeedback of primary motor cortex and supplementary motor area 167

6.1 INTRODUCTION 168

6.2 METHODS 170

6.2.1 Participants 170

6.2.2 Motor imagery questionnaire 171

6.2.3 Instruction for localiser and neurofeedback training 171

6.2.4 MRI data acquisition 172

6.2.5 fMRI-NF setup 172

6.2.6 Physiological recordings 173

6.2.7 fMRI training session 173

6.2.8 Target region selection and feedback normalisation 174

6.2.9 Online fMRI analysis, neurofeedback calculation and presentation 175

6.2.10 Post-training questionnaire 176

6.2.11 Offline fMRI analysis 176

6.2.12 Power calculation 177

6.2.13 Planned statistical analyses 177

6.2.14 Exploratory analyses 177

6.3 RESULTS – PLANNED ANALYSES 179

6.3.1 Motor imagery questionnaire 179

6.3.2 Percent signal changes in target ROIs 179

6.3.3 Physiological measurements 181

6.3.4 Post-training questionnaire 182

6.4 RESULTS – EXPLORATORY ANALYSES 182

6.5 DISCUSSION 184

7 Graded fMRI-NF training of motor imagery in middle cerebral artery stroke patients – a preregistered proof-of-Concept study 191

7.1 INTRODUCTION (PREREGISTERED) 192

7.2 METHODS (PREREGISTERED) 194

7.2.1 Design and materials 194

7.2.2 Procedure 196

7.2.3 Analysis plan 201

7.3 CHANGES APPLIED TO PREREGISTERED METHODOLOGY 203
7.4 RESULTS 205
7.4.1 Recruitment 205
7.4.2 Patient demographics 206
7.4.3 Preregistered analyses 207
7.4.4 Exploratory analyses 211
7.5 DISCUSSION 214
8 EEG-NF for Parkinson's disease – A Translational Chapter 222
8.1.1 EEG-NF in PD and potential EEG target markers 223
8.1.2 Conclusion and trial design 228
9 General Discussion 230
10 Appendix A-C 250
10.1 APPENDIX A 250
10.2 APPENDIX B 252
10.3 APPENDIX C 257
11 Bibliography 259

LIST OF TABLES

TABLE 1.1 QUESTIONABLE RESEARCH PRACTICES (QRPS). 19
TABLE 1.2 BINOMIAL TEST OF NULL FINDING PREVALENCE 27
TABLE 1.3 HOEHN AND YAHR SCALE STAGING 52
TABLE 2.4 REPORTED MAIN OUTCOME MEASURES FOR FMRI-NF TRIALS 66
TABLE 3.1 EVIDENCE CATEGORIES FOR BAYES FACTOR 103
TABLE 3.2 CLINICAL TRIAL PHASES FOR DRUG DEVELOPMENT 106
TABLE 4.1 DEMOGRAPHICAL AND CLINICAL BASELINE DATA 126
TABLE 4.2 GROUP MEAN DIFFERENCE ESTIMATES 128
TABLE 4.3 CHANGES OVER TIME IN RAW SCORES 129
TABLE 4.4 T-TEST FOR POMS DEPRESSION AND POMS TMD 131
TABLE 4.5 DESCRIPTIVE STATISTICS OF IMPROVEMENT - TMD 132
TABLE 4.6 DESCRIPTIVE STATISTICS OF IMPROVEMENT - DEPRESSION 132
TABLE 5.1 REGRESSION COEFFICIENTS FOR SELF-EFFICACY MEASURES. 156
TABLE 6.1 CORRELATIONS BETWEEN PHYSIOLOGICAL PARAMETERS AND MOTOR IMAGERY TASK PREDICTOR 181
TABLE 7.1: POWER CALCULATION 196
TABLE 7.2 PATIENT'S MOTOR IMPAIRMENT AND HEAD MOTION 207
TABLE 7.3 TEST OF NORMALITY FOR H1_A ON SUBJECT LEVEL 209
TABLE 7.4 TEST OF NORMALITY FOR H1_B ON SUBJECT LEVEL 209
TABLE 7.5 FREQUENTIST ONE SAMPLE T-TESTS (H1_A) - SUBJECT LEVEL 210
TABLE 7.6 BAYESIAN ONE-SAMPLE T-TESTS (H1_A) - SUBJECT LEVEL 210
TABLE 7.7 FREQUENTIST PAIRED T-TESTS (H1_B) - SUBJECT LEVEL 211
TABLE 7.8 BAYESIAN PAIRED T-TESTS (H1_B) - SUBJECT LEVEL 211
TABLE 7.9 SIDE OF DOMINANCE AND STROKE, SELF-RATED MOTOR IMAGERY SUM SCORES FOR VISUAL IMAGERY AND KINAESTHETIC IMAGERY 213s

LIST OF FIGURES

FIGURE 1.1 DISTRIBUTION OF SAMPLE ESTIMATES OF A SMALL EFFECT EITHER IN LARGE STUDIES OR IN SMALL STUDIES 23
FIGURE 1.2 BAYESIAN BINOMINAL TEST FOR OBSERVED PROPORTION OF NULL FINDINGS 28
FIGURE 1.3 SCHEMATIC DIAGRAM OF BASAL GANGLIA CIRCUITRY 50
FIGURE 2.1 OVERVIEW OF FMRI-NF STUDIES CONDUCTED IN DEPRESSED 65
FIGURE 3.1 SCHEMATIC REAL-TIME NEUROFEEDBACK SETUP 92
FIGURE 3.2 EXAMPLE POWER CALCULATION 95
FIGURE 3.3 EXAMPLE EQUIVALENCE TESTS 98
FIGURE 3.4 EXAMPLE CALCULATION OF THE SAVAGE-DICKEY RATIO 101
FIGURE 3.5 BAYESIAN SEQUENTIAL ANALYSIS 104
FIGURE 3.6 REGISTERED REPORTS (RRS) WORKFLOW DIAGRAM. 111
FIGURE 4.1 CONSORT PATIENT FLOW DIAGRAM. 124
FIGURE 4.2 CHANGE IN PRIMARY OUTCOME MEASURE HDRS 127
FIGURE 4.3 MEAN SCORES OVER SESSIONS FOR POMS 131
FIGURE 4.4 ROBUST REGRESSION ANALYSIS BETWEEN CHANGE IN SELF-EFFICACY AND HDRS-17 AT PRIMARY ENDPOINT. 134
FIGURE 5.1 BOXPLOTS OF NUMBER OF VOXELS 152
FIGURE 5.2 PROBABILITY MAP OF TARGET REGIONS 153
FIGURE 5.3 AVERAGE T-VALUE OF TARGET ROIS 154
FIGURE 5.4 CORRELATION BETWEEN UP-REGULATION AND HDRS-17 155
FIGURE 5.5 WHOLE BRAIN ANALYSIS 157
FIGURE 6.1 EXPERIMENTAL SETUP 174
FIGURE 6.2 BOLD PERCENT SIGNAL CHANGES (PSC) IN THE TARGET ROIS 180
FIGURE 7.1 EXAMPLE SEQUENCE OF CONDITIONS 200
FIGURE 7.2 CONSORT FLOW DIAGRAM 206
FIGURE 7.3 BAYESIAN T-TEST AND SENSITIVITY ANALYSIS FOR H1_A 208
FIGURE 7.4 BAYESIAN T-TEST AND SENSITIVITY ANALYSIS FOR H1_B 209

List of Abbreviations and Acronyms

ANOVA – Analysis of variance

ANCOVA – Analysis of covariance

BF – Bayes factor

BOLD – Blood level oxygenation dependent

BSS – Bayesian sequential sampling

CI – Confidence interval

CONSORT – Consolidated standards of reporting trials

CST – Corticospinal tract

DBS – Deep brain stimulation

EEG – Electroencephalography

ERD – Event related desynchronisation

fMRI – functional magnetic resonance imaging

fNIRS – functional near-infrared spectroscopy

GLM – General linear model

HDRS – Hamilton depression rating scale

IDEAL – Innovation, development, exploration, assessment and long-term studies

LFP – Local field potential

NFE – fMRI neurofeedback of areas involved in processing positive emotions

NFS – fMRI neurofeedback of areas involved in processing visual scenes

M1 – Primary motor cortex

MCA – Middle cerebral artery

MCID – Minimal clinically important difference

PD – Parkinson's disease

PoC – Proof-of-concept

POMS – Profile of mood states

RCT – Randomised controlled trial

SESOI – Smallest effect size of interest

SMA – Supplementary motor area

STN – Subthalamic nucleus

TIDieR – Template for Intervention and Replication

TMS –Transcranial magnetic stimulation

LIST OF APPENDICES

FIGURE A1 CONSORT CHECKLIST 250
TABLE A1 CHANGES OVER TIME IN RAW MEAN SCORES AND STANDARDISED EFFECT SIZES BOTH INCLUDING 251
FIGURE B1 WHOLE BRAIN PROBABILITY MAP OF TARGET REGION SELECTION FOR BOTH TREATMENT GROUPS. 252
TABLE B1 WHOLE BRAIN ANALYSIS. BRAIN ACTIVITY DURING NEUROFEEDBACK BLOCKS COMPARED TO REST 253
TABLE B2 WHOLE BRAIN ANALYSIS. ACTIVITY DURING NEUROFEEDBACK BLOCKS FOR GROUP CONTRAST 254
TABLE B3 WHOLE BRAIN ANALYSIS. WITHIN FACTOR (SESSION) 255
FIGURE C1 EXAMPLE ANATOMICAL IMAGE WITH ARTEFACT 257
FIGURE C2 VIEW OF CONTROL INTERFACE OF TURBO-BRAINVOYAGER 258

1 INTRODUCTION

This Chapter serves as an introduction to the clinical conditions that will be addressed in the context of neurofeedback training in this thesis. The chapter commences with a brief meta-research review about replicability and reproducibility in the biomedical sciences to introduce some criteria for assessing the evidential value of published literature. Specifically, the replication crisis and research practices that have led to it are discussed, followed by an overview of developments that aim to address problematic research practices and that can increase the robustness of scientific research. The chapter continues with a review of the main clinical characteristics, accepted and experimental treatment options for the mood disorder unipolar depression, as well as upper limb rehabilitation in stroke and Parkinson's disease (PD). Further, real-time fMRI neurofeedback (fMRI-NF) studies that have been conducted for these conditions will be reviewed. The chapter will close with a presentation of remaining research questions and a short outlook how these will be addressed in this thesis.

1.1 Robust Science

"Reproducibility, rigor, transparency, and independent verification are cornerstones of the scientific method" (McNutt, 2014). The Scientific method describes a set of techniques used in investigating phenomena to acquire new or correct previous knowledge. The concept of falsifiability describes the ability to test and reject statements, e.g. in the form of hypotheses, and it is thus considered a central element of scientific discovery (Popper, 2005). Falsifiability relies on reproducibility and replicability to allow comparability between results and thereby justify valid rejection of hypotheses. This thesis aims to highlight developments in the scientific community to increase and maintain falsifiability and show where these have been realised in the presented work. Seemingly pervasive failure to replicate previous research questions the robustness of findings reported in many scientific disciplines including clinical medicine (Moonesinghe et al., 2007) and psychology (Open Science Collaboration, 2015) and requires developing a better understanding of underlying reasons. Also known as the "replication crisis", recent meta-research has shown that a substantial amount of published work fails to replicate (or reproduce). Besides reproducibility, replicability forms a prerequisite for robust science. Although the concepts of reproducibility and replicability are related and both are central to scientific practices, they describe different aspects: whereas reproducibility describes the capacity to re-calculate results within an existing data set and existing methods, replicability describes the chance that other scientists will find a consistent result in a new data set but using a highly similar design (Patil et al., 2016). A lack of transparent reporting of methodology, valid use of reliable statistical methods and selective publications of outcome variables have been identified as some main reasons for these crises (Chambers, 2017; Munafò et al., 2017). The following paragraphs summarise the replication crisis and highlight some underlying causes (Chapter 1.1.1) and describe developments that have been introduced to improve the status quo (Chapter 1.1.2). These developments include 1) reporting guidelines such as the CONSORT statement (Chapter 3.4.4) and trial registration (Chapter 3.4.5) for RCTs, and 2) preregistration and registered reports (Chapter 3.5). Although the principles of falsification are key for the definition of robust science as presented in this thesis, it should be mentioned that other philosophies of science have put less emphasis on this principle. In fact, the philosopher Paul Feyerabend for instance critiqued the call for falsifiability as a methodological monism that may obstruct scientific breakthroughs. He referred to case studies of major discoveries in the history of science, including Galileo Galilei's contributions to the heliocentric model.

Feyerabend pointed out how practices often deviated from the rational scientific method proposed by Popper and other advocates of critical rationalism (Feyerabend, 1993).

1.1.1 Replication crisis and questionable research practices

Many scientific disciplines including cancer biology (Begley and Ellis, 2012; Willingham et al., 2017), psychology (Open Science Collaboration, 2015), clinical medicine (Moonesinghe et al., 2007), computational science (Peng, 2011), and lately also Artificial Intelligence (Hutson, 2018) are facing a replication crisis: a substantial amount of published work fails to replicate. The underlying reasons for the replication crisis are numerous. The high prevalence of underpowered studies in biomedical research and neuroscience is one major factor (Button et al., 2013; Dumas-Mallet et al., 2017; Nord et al., 2017; Nosek et al., 2012). For neuroscience, Button et al.'s (2013) reported an alarmingly low median power of only 0.21, which means that an effect could only be found once in five times it was investigated. However, their analysis was performed as a meta-analysis on all meta-analyses published in neuroscience in 2011 (N = 49), if all studies stemmed from the same population of studies. In response, Nord et al. tested if these studied stemmed from different underlying populations using Gaussian mixture modelling (GMM), a technique that tries to recover separate normal distributions that in sum produce the observed distribution (Nord et al., 2017). For instance, if a data set featured many low and a few highly powered studies, a median merely reports that (at least) 50% of these studies feature low power. In contrast, GMM can infer that a distinct subset of highly powered studies exists and hence allows a more nuanced interpretation of the data. Nord et al.'s analysis revealed that over 70% of studies featured low or very low power of less than 0.5 (i.e., less than the chance level of landing heads or tails in a coin toss). However, their data also suggested that over 10% of studies appeared highly powered (>0.8). Further, the authors discovered remarkable differences between subfields with very low power (<0.2) for genetic association studies and relatively high power (>0.9) in neurochemistry. Similar results were reported for the literature in the biomedical sciences by another study that investigated power in 660 studies that were included in meta-analysis between 2008 and 2012 (Dumas-Mallet et al., 2017). The authors found that 50% of studies had very low statistical power (mean power <0.2). Splitting results by methodology, they found that purely cognitive/behavioural studies had high power (median power 0.93), genetic studies very low power (median power 0.08), and brain imaging low to medium, but still inadequate power (median power 0.27). Lastly, also with regards to clinical trials, underpowered studies dominate the literature. One study estimated that the

proportion of Cochrane meta-analyses that consist exclusively of underpowered studies (defined as power < 0.5) was as high as 70% (Turner et al., 2013).

Two conclusions can be drawn from a literature that is plagued by underpowered studies: 1) If the majority of published research was initially underpowered to detect meaningful effect sizes, published findings that report significant findings are likely predominantly false positives (Ioannidis, 2005; Simmons et al., 2011). Indeed, recent estimates suggest more than 50% of published findings in neuroscience are likely *false positives* (Szucs and Ioannidis, 2017), compromising the evidence that can be drawn from literature reviews, which are traditionally regarded as the gold standard for evidence based medicine (Ault, 2003). In other words, treatments that are reported to work may not work reliably, genes that are reported to contribute to a phenotype may contribute little, and conditions that are reported to matter for cognitive processes may only play a marginal role.

2) Underpowered studies are less likely to detect an effect, yet, the literature is largely dominated by positive findings (Cristea and Ioannidis, 2018; Fanelli, 2012). Further, while the prevalence of positive findings has decreased for clinical trials since trial registration was introduced (Kaplan and Irvin, 2015), meta-research suggests that negative findings largely remain unreported. Two mechanisms likely drive the high significance rates despite low average power. a) *researcher bias*— e.g. "Hypothesizing After the Results are Known" (Kerr, 1998), "Selected Hypothesized Areas after Results are Known" (Poldrack et al., 2017) and other forms of post-hoc selection of dependent and control variables, "P-hacking" (Simmons et al., 2011), as well as selective submission of positive results, also called *file drawer effect* (Rosenthal, 1979); b) *publication bias*—significant results are more likely accepted for publication than nonsignificant findings (Easterbrook et al., 1991). Table 1.1 provides a more detailed overview of these and other questionable research practices and external constraints (Wicherts et al., 2016).

P-Hacking	Statistical tests and data selection criteria are modified based on the outcome of previous tests to reach a desired statistical outcome. Such questionable practices include selectively discarding data, e.g. by

	applying post-hoc exclusion criteria, and changing statistical tests. Besides data analysis, also data acquisition can be instrumentalised for p-hacking, e.g. by employing a flexible sampling plan with a stopping rule that is based on the (uncorrected) test statistic (Simmons et al., 2011).
HARKing	Short for "Hypothesizing After the Results are Known" (Kerr, 1998). Post-hoc interpretations are being falsely declared as planned hypothesis that were formulated prior to data collection (i.e. they declared as *a priori*).
SHARKing	Short for "Selecting hypothesized areas after results are known" (Poldrack et al., 2017). This questionable research practice represents a special form of HARKing and affects neuroimaging studies where regions of interest (ROIs) are being analysed. Such ROI analyses are falsely declared as planned hypothesis tests, while they are truly based on a previous (unreported) exploratory analysis.
Double-dipping	Using the same dataset for selection and selective analysis (Kriegeskorte et al., 2009). SHARKing can involve double-dipping, e.g. when the same data set is used for hypothesis generation and hypothesis confirmation. Other examples include classification studies where training and test data sets are not clearly

	separated, and classification accuracy is being inflated by fitting correlated noise.
Publication bias / File-drawer effect	Studies with statistically significant results are more likely to be published than studies with negative or inconclusive results (Scargle, 2000). As a result, authors are less incentivized to compose articles about studies that contain negative or null findings (file-drawer effect).
Improper use of statistical methods	Improper use of statistical methods likely represents the most active area of debate within the sciences and spans a wide range of questionable practices. In neuroimaging, it refers in particular to inflated false-positive rates , but also inflated false-negative rates due to in-appropriate correction for multiple comparison (Eklund et al., 2016; Kessler et al., 2017). Other examples include the fallacy of inferring the absence of an effect from not observing evidence for an effect (and hence, from merely inconclusive data; Lakens et al., 2018), and interpreting single group differences without demonstrating formal interactions (Gelman and Stern, 2006; Nieuwenhuis et al., 2011).

Table 1.1 Questionable research practices (QRPs). Modified from (Lorenz, 2017).

Within current reward structures, researchers are incentivised to *find* significant results and rapidly publish in high impact journals (Chambers, 2017; Higginson and Munafò, 2016; Munafò et al., 2017). Research fields that work with high dimensional data—such as produced by brain signals—require complex pre-processing and data analysis pipelines. These usually

involve numerous pre-processing and data analysis steps, which often result in many ways of analysing such data. In consequence, different analysis pipelines can lead to vastly different analysis outcomes and interpretations (Carp, 2012a). Questionable research practices (QRPs) have been investigated for different neuroscience fields. For functional neuroimaging, Carp (2012b) demonstrated how exhaustive combinations of possible pre-processing and data analysis steps results in several thousand unique analysis pipelines. Their results varied remarkably with regards to brain activation strength, location, and extent. Moreover, out of a random sample of more than 240 fMRI studies, more than 80% of studies featured unique analysis pipelines, which often lacked sufficient detail in justifying processing and analysis decisions (Carp, 2012b). For event-related potentials (ERPs) in electrophysiology, Luck and Gaspelin (2017) demonstrated how the common practice of first selecting time windows based on a test statistic (e.g. the grand average) and then comparing conditions on the very same statistic may yield statistically significant, but hardly replicable results. For non-invasive brain stimulation, Héroux et al. (2017) investigated the prevalence of questionable research practices among researchers who work with brain stimulation techniques. In their survey, the authors found that a high proportion admitted committing questionable research practices such as selective reporting of outcomes and adjusting statistical analyses to reach significant results. As we would expect, when researchers tweak analyses to reach significant results, small or non-existent effects become inflated and appear more reliable in the literature than they really are (Héroux et al., 2017).

Low powered studies, which mostly result from small sample sizes, in combination with *publication bias* – positive findings are rewarded and more likely to be published compared to null findings (Franco et al., 2014; Rosenthal, 1979) – lead in particular to inflated effect sizes. Small studies usually contain higher levels of noise and are thus less likely to find an effect. However, if a statistically significant result is found with a small sample, some researchers tend to believe that it must reflect a truly large effect ("what does not kill my effect makes it stronger"; Loken and Gelman 2017). This belief is misleading because the increased noise in small studies renders effect size estimates imprecise and increases their variability (compare shape and means of distributions in Figure 1.1). In fact, significant estimates are often *inflated*, i.e. much larger than the *true* effect sizes (Loken and Gelman, 2017). High power estimates can occur either with a) large samples that can detect small to moderate, realistic effect sizes,

or b) with small samples that can only detect large effect sizes—which likely represent inflated estimates of small effects. This follows from three assumptions:

1) Only few effects that are not trivial and hence worthwhile investigating with scientific instruments are truly large, but many are small to moderate. For instance the average effect size for fMRI neuroimaging (across different task paradigms) has been estimated as a Cohen's d (*d*) of about 0.5 (Poldrack et al., 2017), and for behavioural Psychology around $d = 0.2 – 0.3$ (Open Science Collaboration, 2015).

2) As shown in Figure 1.1, small effects can only become significant in typical small samples if they are inflated (Loken and Gelman, 2017).

3) Significant effects are more likely to be published (publication bias). This effect is also illustrated in Figure 1.1. Small sample sizes result in larger variability and hence a broader distribution (see red distribution) compared to large sample sizes (see green distribution). Assuming publication bias and a true effect size of $d = 0.30$, small studies with significant results overestimate the true effect more than large studies with significant results.

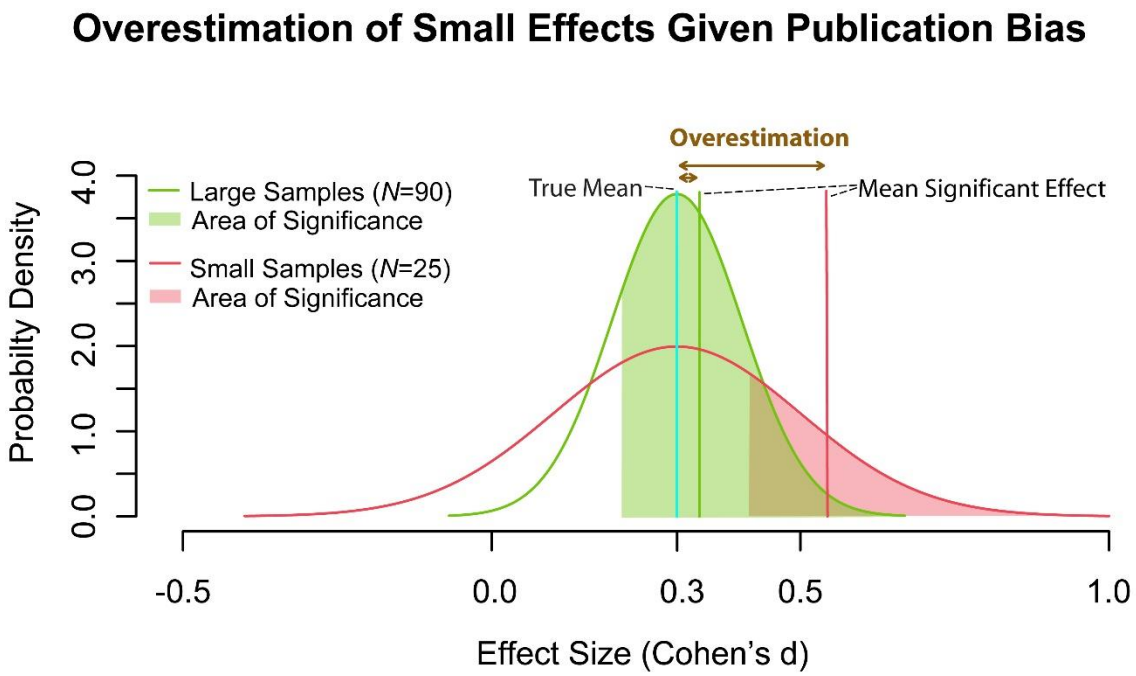

Figure 1.1 Distribution of sample estimates of a small effect either in large studies (green distribution) or in small studies (red distribution). Shaded areas indicate reported

effects if only significant results are reported (publication bias). When there is a true effect of *d* = 0.30 (cyan vertical line), most studies (80%) with large samples will detect it and yield a significant result for effect size estimates > 0.21 (shaded in green). In contrast, studies with small samples can only detect it for effect size estimates > 0.42, and thus only a small fraction (30%) will detect the effect (shaded in red). In the presence of strong publication bias, small-sample studies only get published when they yield a significant result. Such studies will always overestimate the true effect (indicated by the lack of an overlap between the red shaded area and the cyan vertical line) and will do so to a greater extent than large published studies (see difference between green and red vertical line). The following parameters were used to create the figure: The small sample size (N = 25) is based on 0.30 power to detect an effect of Cohen's d = 0.30. Power of 0.30 is equivalent to the median power in neuroscience found by Nord et al. after excluding null results from meta-analyses. The large sample size (N = 90) is based on a hypothetical statistical power of 0.80, which is a value that is often recommended. Shown are results for a one-sample two-sided t-test at an alpha level of 0.05. Adapted from (Algermissen and Mehler, 2018).

One remarkable demonstration of the effect of *publication bias* has been provided by (Kühberger et al., 2014) who correlated reported sample sizes with estimated effect sizes for a random sample of 1,000 psychological articles. While these two measures should in theory be independent, the authors found a moderate negative relationship (r =−.45), indicating that in particular small samples report high effect sizes. It is thus likely that also recent studies that reviewed power in neuroscience (Button et al., 2013; Dumas-Mallet et al., 2017; Nord et al., 2017) reported findings which were likely inflated because these meta-analyses were based on published weighted mean effect sizes. Thereby, these studies relied on empirical effect sizes that were subjected to QRPs and publication bias. Hence, the true power of included studies was likely even less to detect realistic effect sizes (Open Science Collaboration, 2015; Poldrack et al., 2017).

In the following paragraph some remedies for QRPs will be presented, including ways to reinforce better reporting standards and approaches that address *researcher bias* and *publication bias*. Moreover, the publishing formats preregistration and registered reports (RRs) will be highlighted. These require separating planned from exploratory hypotheses and hence protect against various above described QRPs (Chambers, 2015; Munafò et al., 2017).

1.1.2 Some remedies for (more) robust science

A review of the replication crisis and underlying biases demonstrates that transparent documentation and reporting is pivotal for robust science. With regards to RCTs, inconsistencies and a lack of sufficient reporting of key design details such as randomisation (Altman and Doré, 1990; Schulz et al., 1994), sample size (Pocock et al., 1987), power calculations (Moher et al., 1994) and outcome variables (Pocock et al., 1987) have motivated the development of structured reporting guidelines (Andrew, 1994). Introduced in 1996 and since then continuously refined (Begg et al., 1996; Moher et al., 2010, 2001), the Consolidated Standards of Reporting Trials (CONSORT) statement contains a checklist for details about the trial design (e.g. blinding and randomization) and results such as attrition rates, main clinical outcome and exploratory analyses. The CONSORT checklist forms a requirement for trial manuscript submissions to medical journals. Moreover, trials are required to be registered before start of data collection in a public record, and provide information about outcome measures, trial dates and planned sample sizes.

First meta-analyses indicate that these measures have improved the quality of trials reporting. For instance, the introduction of CONSORT was linked to improved reporting for some measures, but reporting standards require further improvements (Turner et al., 2012). Moreover, the proportion of null findings has remarkably increased by about 50% (Kaplan and Irvin, 2015). However, despite ubiquitous registration of clinical trials and a rise in reported null findings among published studies, a significant proportion of registered trials remains unpublished. For instance, one study estimated that almost 1/3 out of a set of 585 large registered trials with more than 500 participants were not published within 5 years of study completion (Jones et al., 2013), indicating substantial publication bias. By focusing exclusively on large, laborious trials, the authors limited the likelihood that studies were not published due to a mere lack of interest of the involved scientists (thus minimising a file-drawer effect). The omission to report trials of this size represents a major loss of information because these studies were likely sufficiently powered to provide evidence for the absence of an effect (Munafò and Neill, 2016; Walker and Nowacki, 2011) . Taken together, while CONSORT in combination with trial registration allows to rule out certain questionable practices including a) switching of primary and secondary outcomes, b) unjustified changes of trial duration and c) arbitrary stopping rules in data collection, RCTs are not safeguarded against the *file drawer effect*

(Rosenthal, 1979), or *publication bias* (Easterbrook et al., 1991). Both factors therefore remain an issue, imperil the robustness of findings, and likely contributed to the replication crisis encountered in medicine (Ioannidis, 2005; Moonesinghe et al., 2007).

How can clinical trials and neuroscience become more reliable? The trial registration practice has inspired the new publishing formats preregistration and registered reports (RR) (Chambers, 2013). RRs are peer-reviewed before data collection based on the study's rationale, design, power calculation and analysis plan, and may result in an acceptance in principle (Chapter 2.5). Importantly, RRs distinguish between planned and exploratory hypotheses and hence reduce. Hence, RRs can effectively reduce or even eliminate several biases *researcher bias* and *publication bias*.

With regards to CONSORT, future versions may benefit by incorporating elements of RRs that are currently not included:

- Peer-review of power calculations: this would likely improve the quality of adequate sample size estimations. Although *a priori* defined sample sizes are required, the underlying effect sizes and exact power calculations are usually not reported, which bears the risk that trials are statistically underpowered.
- *A priori* declaration of analysis plans: Trial registrations usually do not contain a detailed description of analysis plans, which leaves many "researcher degrees of freedom" in data preprocessing and analysis, in particular for complex interventions that involve digital signals (e.g. neuroimaging).
- In principle acceptance: While trial registrations offer a public record for trial protocols and outcome that allows conducting meta-science studies, a large proportion of research remains unreported and is thus lost for the public (Jones et al., 2013).

Altogether, these mechanisms should protect research against QRPs and consequently result in a higher proportion of non-significant findings reported in the literature. Indeed, a preliminary analysis conducted on the prevalence of reported null findings in published registered reports listed in the registered report database (The Registered Reports Steering Committee, May 2017) suggested that a substantially larger proportion than estimates of 5-20% of null findings that have been traditionally reported in the literature (Cristea and Ioannidis, 2018; Fanelli,

2012). When compared to a relatively liberal test value of 30%, classical and Bayesian binomial tests indicated that published preregistered studies feature a substantially larger proportion of null findings (Table 1.2, Figure 1.2). These results are in line with reports from clinical medicine, where required study preregistrations was associated with a rise of reported of null findings (Kaplan and Irvin, 2015). Since RRs give scientists a guarantee to publish work as a peer-reviewed study irrespective of significance status of results, null findings do not end up in a file drawer and researchers gain a publication. Taken together, this data supports the idea that RRs might help alleviate publication bias towards positive findings.

						95% Confidence Interval	
	Counts	Total	Proportion	p	VS-MPR*	Lower	Upper
Null findings	67	95	0.705	< .001	2.173e +13	0.603	0.794

Table 1.2 Binomial Test of null finding prevalence in a sample of preregistered published studies (May 2017). Vovk-Sellke Maximum p -Ratio: Based on the p -value, the maximum possible odds in favour of H_1 over H_0 equals 1/(-e p log(p)) for $p \leq .37$ (Sellke, Bayarri, & Berger, 2001). Proportions tested against a value 0.3, which represents a conservative estimate for the proportion of null findings among published literature. Analyses were carried out based on preregistered studies published prior to May 2017 which were listed in a public repository on the Open Science Framework (https://www.zotero.org/groups/479248/osf/items/collectionKey/VKXUAZM7).
Specifically, prespecified independent hypotheses were identified in the preregistration document, compared with reported results, and coded as either a null finding or a positive finding.

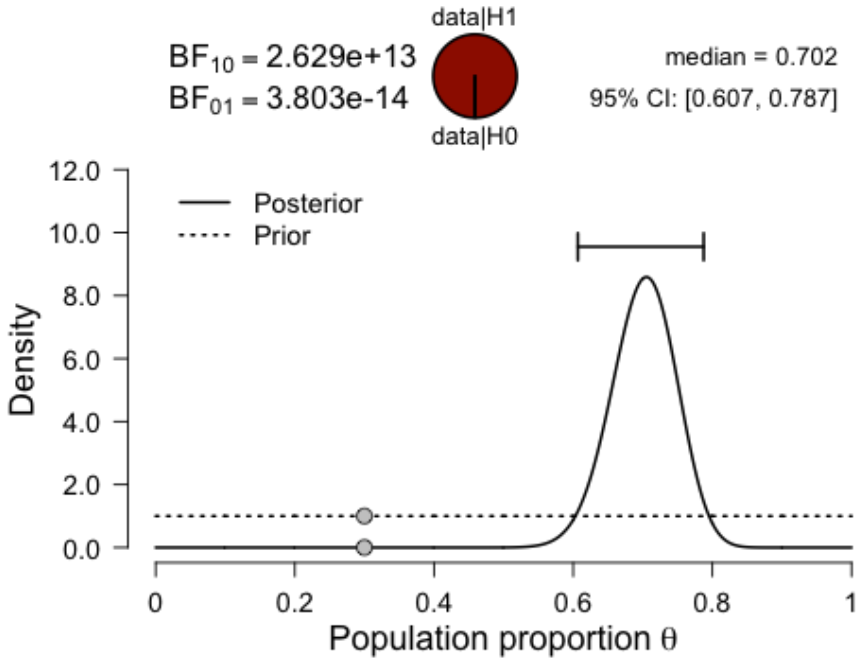

Figure 1.2 Bayesian binominal test based on observed proportion of null findings (~0.7) tested against a level comparable to the upper limit of the proportion of null findings reported in the literature (0.3; (Cristea and Ioannidis, 2018; Fanelli, 2012)).

Taken together, the introduction of peer-reviewed RRs at medical journals represents a desirable development that is in line with the principles of trial registration and CONSORT. Although guidelines for statistical analysis plans do not exist yet for trial registrations, first recommendations for their conceptualisation have recently been formulated (Gamble et al., 2017). Considering the relatively short time span of about 2 years that it took between first recommendations (Andrew, 1994) for structured reporting of RCTs and the introduction of a first template (Begg et al., 1996) in the form of CONSORT, analysis plans may become part of trial registrations in the near future. Open science practices that can improve transparency, replicability and reproducibility are thus likely becoming increasingly important in clinical research. Indeed, the International Committee of Medical Journal Editors (Taichman et al., 2017), whose recommendations are binding for most leading biomedical journals, has recently issued a data-sharing statement policy. According to this consensus statement, clinical trials submitted for publication must include a data sharing statement as of July 2018 and new trials beginning to enrol patients from 2019 are required to include a data sharing plan.

How can neuroscientists solve their power problem? First, they can improve their power calculations (Chapter 3.4.1.2), which form an essential part in the peer review process of RRs (Chapter 3.5). To do so, researchers should calculate power *before* data collection and specify their *smallest effect size of interest* (SESOI; Lakens et al. 2018). They should neither rely on effect sizes reported in the literature, which are often inflated, nor on effect size estimates from small-sample pilot studies, which vary largely (Figure 1.1, red distribution) and might thus severely underestimate the sample size required for adequate power. In contrast, SESOIs (Chapter 3.4.1.5) require that researchers specify the smallest effect size they consider worthwhile investigating. SESOIs may vary between different fields and hypotheses. For instance, for translational and clinical research minimal clinically important differences (MCIDs) define how large a change on an outcome variable (e.g. a clinical scale) should be to be considered worthwhile detecting (Mcglothlin and Lewis, 2014). Taken together, researchers who work with SESOIs are more likely to conduct adequately powered studies. A related, albeit independent aspect is the use of appropriate statistics, which is key to draw valid inferences: this includes controlling type-I error rates by appropriate correction for multiple testing (Asendorpf et al., 2013; Eklund et al., 2016), avoiding fallacies in the interpretation of interactions (Nieuwenhuis et al., 2011), and employing statistical techniques that allow providing evidence for the absence of an effect (Dienes, 2014; Lakens, 2017). Defining SESOIs before data collection allows conducting reliable equivalence or non-inferiority tests (Lakens et al., 2018; Walker and Nowacki, 2011) and can inform prior distributions for Bayesian hypothesis testing (Dienes, 2014). Both techniques allow to provide evidence for a negligible, or null effect (Chapters 3.4.1.5 and 3.4.2).

Second, novel real-time imaging and optimisation techniques provide additional scope to optimise data acquisition and minimises *researcher bias*. Specifically, experiments that involve real-time data analysis, such as fMRI-NF studies, can effectively limit researcher's degrees of freedom, because their pipelines require pre-specifying crucial parameters:

- Online pre-processing parameters (e.g. spatial smoothing kernel's and motion correction) are predefined in the real time analysis protocol, pre-determining substantial parts of the analytic pre-processing space that researchers can justify for offline analyses (Carp, 2012a).

- Data analysis parameters and region of interests are usually localised in an independent scan (functional localiser) before the training starts. Hence, the exact coordinates and number of voxels comprising regions of interest are predefined. Hence, real-time experiments implicitly predefine their main neural outcome measures, which may help avoiding forms of double-dipping (Kriegeskorte et al., 2009), HARKing (Kerr, 1998), and SHARKing (Poldrack et al., 2017).
- Other experimental parameters that pertain the brain imaging data, for instance the way the feedback is calculated based on model estimates or raw percent signal changes, are also prespecified. Hence, additional degrees of freedom that arise during the analysis are limited, which can avoid forms of p-hacking (Wicherts et al., 2016).

These principles are largely applicable for neurofeedback studies that use other modalities such as Electroencephalography (EEG) and functional near-infrared spectroscopy (fNIRS). Hence, neurofeedback studies in general implicitly preregister most of their analysis pipeline, and they are thus in particular suited for study preregistrations and the publication of training protocols (Cox et al., 2016). However, to ascertain compliance with prespecified real-time analysis pipelines, it is mandatory that either real-time data is subsequently used for offline data analysis, or that offline data analysis based on raw acquired data replicates online-analyses using identical parameter settings.

1.2 Clinical conditions of interest

1.2.1 Unipolar depression

1.2.1.1 Main clinical characteristics and pathophysiology

The main clinical features of depressive disorder are persistent low mood, which is often accompanied by anxiety, loss of pleasure (anhedonia), lack of energy (lethargy), and symptoms of insomnia (Geddes et al., 2012). Depression is diagnosed based on detailed medical interviews such as the Mini-International Neuropsychiatric Interview (MINI) or the Structured Diagnostic Interview for DSM-5 (SKID) that are conducted by trained healthcare professionals (Pettersson et al., 2018; Tolin et al., 2018). For clinical purposes, however, severity is currently staged based on the 10th revision of the International Statistical Classification of Diseases and Related Health Problems (ICD-10). The ICD-10 criteria for depression distinguish between key symptoms including persistent low mood, loss or interest (anhedonia) and fatigue, as well as associated symptoms such as disturbed sleep, poor concentration and feelings of guilt. According to the number of symptoms, the ICD-10 classifies patients as not depressed (fewer than two symptoms), mild depression (two or three symptoms), moderate depression (four or more symptoms), and severe depression that additionally features feelings of worthlessness, guilt, and often suicidal thoughts (Geddes et al., 2012).

In fact, the risk of suicidality is increased such that about 10% of patients eventually commit suicide (Geddes et al., 2012). Therefore, a careful examination of individual symptoms is required to grade the severity of depression and evaluate suicidal risks. For research purposes, additional questionnaires such as the clinician rated Hamilton Depression Rating Scale (HDRS) and the self-reported Beck Depression Inventory (BDI) are used to assess depression severity and monitor changes in response to treatment. One defining character of depression is that it occurs in episodes that consists of relapses and remission in a substantial proportion of patients (Andrews, 2001). Thresholds have been defined that help assessing treatment success; on the HDRS-17 a response is usually defined as a reduction of >50%, remission is defined as an HDRS-17 score <=7, and a relapse entails an HDRS-17 score >=14.

The presentation of symptoms between different patients is heterogeneous (Chen et al., 2000; Geddes et al., 2012). Indeed, different subtypes of depression have been suggested based on constellations of symptoms with the aim to stratify patients and identify more effective

treatment strategies. For instance, melancholic depression is characterised by pervasive anhedonia, depressed mood that is not reactive to changes circumstances and psychomotor retardation. In contrast, atypical depression is characterised by the absence of symptoms that define melancholic depression, a reactive mood, and fatigue. Depression subtype classifications have been explored as means of identifying more effective psychopharmacological treatment strategies for patients (Joyce and Paykel, 1989). However, the prognostic value of depression subtypes remains unclear. One longitudinal population-based study suggested that rather depression severity than depression subtypes is indicative for the course of depression (Chen et al., 2000). This notion was corroborated by a multi-centre pharmacological trial that found no prognostic value for melancholic, atypical and anxious depression subtypes regarding treatment response to commonly used psychopharmacological medication (Uher et al., 2011). Of interest, symptom based subtyping may still be informative with regards to interactions with other co-morbidities such as cardiovascular disease (Baune et al., 2012). With regards to treatment stratification in depression, one criticism is that the neurobiological basis of phenotypes still largely remains unclear, which has incited the exploration of neurophysiological markers (Drysdale et al., 2016, but see also Dinga et al., 2018).

Epidemiological predictions suggest that depressive disorder will represent the largest contributor to disability globally by 2020 (World Health Organization, 2017). The lifetime prevalence of unipolar depression ranges between 15 and 20% for developed countries (Geddes et al., 2012; Kessler et al., 2003). However, given that most epidemiological studies estimate prevalence retrospectively and rely on responses from patients and may thus be affected by response biases, these numbers may underestimate true prevalence rates. This conjecture has been corroborated by data from a prospective epidemiological study that found a nearly 20% prevalence of depression over a period of only 13 years (Patten, 2009). The authors noted that these estimates were nearly twice as high as lifetime prevalence of major depressive episodes reported during the same time interval. Hence, available literature that estimated life time prevalence of depression likely underestimated the true ratios to a remarkable extent, likely because many cases remained unreported. With regards to gender ratios, women are more likely to be diagnosed with depression (2:1). To which extent this gender difference mainly results from biopsychological factors or the way female and male patients present symptoms differently or may seek (psychiatric) consultation to a different degree, remains an active field of research (Geddes et al., 2012).

Depressive disorder can result from an interaction between stressful life events and so called constitutional factors, which include genetic predispositions (Wray et al., 2018), childhood experiences and cognitive coping strategies (Maciejewski et al., 2000). Epidemiological data suggest that the occurrence of a severely stressful lifetime event increases the odds to suffer from depression within the following 6 months by a factor of six (Geddes et al., 2012). This observation highlights the relevance of life experiences in the development of depressive disorder. Studies in clinical psychology suggest that the effect of stressful, or even traumatising experiences on an individual's mental well-being depends on how individuals process stressful life experiences. Whereas negative cognitive biases on the one hand can increase the likelihood that a negative experience will trigger the beginning of a depressive episode, cognitive coping strategies on the other hand can help preventing the onset and relapse. Predisposing psychological factors include a bias to remembering life events or experiences that were associated with negative emotions (Geddes et al., 2012) and cognitive distortions (Disner et al., 2011) such as the tendency to overgeneralize from a singular negative experience to other experiences (negativism bias). Cognitive coping strategies that can reduce an individual's vulnerability include self-efficacy and certain personality trait factors such as low levels of neuroticism (Mulder, 2002). Self-efficacy describes an individual's perceived capacity to cope with challenges (Bandura, 1982) and may thus protect individuals from developing depression. For instance, one population-based longitudinal study with over 2,800 respondents found that self-efficacy mediates the vulnerability to develop depressive symptoms following stressful life events (Maciejewski et al., 2000). High levels of self-efficacy also reduce the impact of depression on engaging in detrimental behaviour such as nicotine abuse (Minnix et al., 2011).

Unipolar depression is recognised as a pathological entity, it should be noted, however, that depressive episodes also occur, together with other symptoms such as mania, in patients suffering from bipolar disorder (Geddes et al., 2012). The putative neurobiological aetiology of depression is complex and involves various neurotransmitter and neurohormonal systems, including the monoamines serotonin, noradrenaline and dopamine, as well as non-monoamines such as glucorticoids (Berton and Nestler, 2006; Nestler et al., 2002). Depression is increasingly viewed as a mood disorder that results from maladaptive, stress-induced neuroplastic changes that involve brain regions of the limbic system, as well as the prefrontal cortex (Krishnan and Nestler, 2008). Based on invasive electrophysiological studies as well as non-invasive neuroimaging studies conducted in humans, neural circuit based models have

been developed and linked to psychopathological patterns observed in depressed patients (Beck, 2008; Disner et al., 2011; Mayberg, 2003).

Replicable neural correlates associated with recovery from depression have motivated the development of translational applications, including invasive and non-invasive brain circuit based intervention methods such as deep brain stimulation (DBS) and fMRI-NF (Delaloye and Holtzheimer, 2014; Linden, 2014). For instance, brodmann area 25 (BA25) as part of the subcallosal cingulate has been found to indicate whether patients likely do respond to pharmacological (see Chapter 1.2.1.2.1) or psychological (see Chapter 1.2.1.2.2) treatment (Dunlop and Mayberg, 2014). Specifically, patients who failed to respond to either treatment showed hyperactivation at rest in BA25 during baseline scans relative to patients who had subsequently responded to these treatments. BA25 is interconnected with medial prefrontal cortex, the anterior cingulate gyrus, but also the hippocampus and the amygdala. It hence represents a hub that interlinks areas which are involved in cognitive control, affect expression and regulation, as well as memory formation. These processes play an important role in cognitive models of depression, for instance with regards to biased processing of neural and negative information (Disner et al., 2011). BA25 itself represents a promising target area for experimental deep brain stimulation (DBS; see Chapter 1.2.1.2.4) treatment (Delaloye and Holtzheimer, 2014). Moreover, the areas it is connected with have been reported to change activity after successful psychological treatment using cognitive behavioural therapy (CBT), rendering them potential biomarkers and also target areas for stimulation-based treatment approaches such as neurofeedback (Franklin et al., 2016). Besides functional correlates, structural correlates have been suggested as prognostic markers in depression and reported for some of these areas. For instance, volume reductions of the in the body/tail of the hippocampus predicted poorer outcome for pharmacological treatment, a finding that has been repeatedly replicated (see for review; Dunlop and Mayberg, 2014). The insular cortex is another brain area that has been proposed as a biomarker for treatment success and to indicate trajectories of depressive episodes. For instance, the metabolism at rest in the right anterior insular cortex pre-treatment has been found to be higher in non-responders relative to responders of CBT (McGrath et al., 2013). For pharmacological treatment (citalopram), responders showed a higher pre-treatment activity relative to non-responders. Further, it was reported that the metabolism levels in the right anterior insula correlated with clinical improvements for CBT, and anti-correlated with clinical improvements for pharmacological treatment (McGrath et al., 2013). A more recently published study that followed up patients over 2 years identified grey

matter volume of the insular cortex as a predictor for relapse in patients who had remitted from a previous depressive episode (Zaremba et al., 2018). While patients who had relapsed showed such reduction, patients that did not show such relationship did not. However, these findings should be interpreted with caution since no direct statistical comparison between groups was reported (see also Table 1.1). Besides changes in brain function/structure in response to treatment, functional neuroimaging data acquired during exposure tasks or paradigms that involve regulation of emotion may reveal further correlates of impaired mental function in depressed patients. Notably, one recent meta-analysis that included neuroimaging data (fMRI and positron emission tomography [PET]) from over 50 studies with over 1,000 did not a find a stable correlate (Müller et al., 2017). The authors highlighted that heterogeneity across studies and paradigms may partly explain the lack of convergence, underscoring the importance of replication studies in this field. Hence, while the surveyed studies suggest that neural correlates following treatment in depression may provide stable markers and promising targets for interventions, the informative value of neuroimaging studies that employ affective cognitive tasks remains to be demonstrated.

Taken together, depression is a mood disorder with putative contribution of both neurochemical/ -endocrinological factors and life events. Affected patients show substantial heterogeneity with regards to symptoms, and biological as well as clinical subtypes that allow a better understanding for personalised treatment remain to be identified. From a neurobiological perspective, depression is a circuit disorder for which multiple neural correlates have been reported, offering new ways to interrogate with invasive and non-invasive translational medicine. Further, entraining protective cognitive factors may be worthwhile to consider when developing new approaches to treat depression. Self-regulation training of relevant brain areas through fMRI-NF may for instance allow fostering perceived self-efficacy in patients.

1.2.1.2 Treatment

1.2.1.2.1 Pharmacological treatment

The monoamine deficiency hypothesis remains the predominant pathological model in the treatment of individuals suffering from depressive disorder. Pharmacotherapy in depression acts on receptors of different monoamine transmitter systems, including Serotonin (e.g. specific serotonin reuptake inhibitors [SSRIs] such as Citalopram and Sertraline), Noradrenaline (e.g. selective serotonin and noradrenaline reuptake inhibitors such as Venlafaxine and Duloxetine).

Other pharmacotherapy is acting less selectively modulating activity of multiple receptor types. Tetracyclic antidepressant drugs for instance act on noradrenergic, serotonergic, but also dopaminergic and cholinergic receptors. The exact pharmacodynamics and thus also potential side effects of a given medication hence depend on the class that it belongs to, but also the specific compound (Geddes et al., 2012). For instance, where tricyclic medication often involves anticholinergic side effects (e.g. dry mouth and sweating), SSRIs may cause nausea and affect sexual functions (Anderson et al., 2000). More severe side effects and complications can include suicidal ideation, which has been associated in particular with SSRIs, and symptoms of Parkinsonism, which has been mainly reported for antipsychotic drugs that block dopamine receptors (Lenze et al., 2015). The efficacy of antidepressant medication compared to placebo has been repeatedly demonstrated. For instance, earlier meta-analyses suggested an on average superiority with standardized mean difference of reduction in depressive scores by about 0.3 (Kirsch et al., 2008). However, the margin of gain is partly dependent on factors such as depression severity at baseline and duration of intake (Fournier et al., 2010; Kirsch et al., 2008; Sugarman et al., 2014). For instance, standardized mean differences between treatment and placebo may range from about 0.1 to about 0.5 for a range of baseline HDRS-17 scores from 18 to 28 (Sugarman et al., 2014). Another more recent meta-analysis, which represents the largest comparison between antidepressant and placebo medication published to-date including over 500 RCT with over 116,000 patients, corroborated and extended these findings to standardized mean differences in remission rates of about 0.3 (Cipriani et al., 2018). Further, the authors analysed the efficacy and acceptability of 21 different medications compared to placebo individually after correction for factors such as depression severity at baseline. All compounds were found more effective, however, odds ratio (OR) varied between 1.4 (Reboxetine, a norepinephrine reuptake inhibitor) to 2.1 (amitriptyline, a tricyclic antidepressant) with regards to response rates. Efficacy was found in favour of the active drug for all included compounds. Acceptance was found largest for Agomelatine (a derivative of melatonin, OR 0.9) and lowest for Clomipramine (a tricyclic antidepressant, OR 1.3 in favour of the placebo). Noteworthy, quality control suggested that the majority (73%) of included studies were affected by publication bias to a moderate degree and hence reported efficacies likely were likely overestimated. First line treatment for moderate depressed patients usually includes the prescription of SSRIs. Epidemiological data obtained from the 'Sequenced Treatment Alternatives to Relieve Depression' (STAR*D) trial, which compared acute and longer-term treatment outcomes associated with up to four successive steps of treatment adaptation in over 4,000 moderately depressed patients, suggests that initial treatment with

Citalopram (an SSRI) is associated with a remission rate of over 35% (Rush et al., 2006). If patients do not respond adequately to a monotherapy, drug regimens can be combined, e.g. with another medication or psychotherapy (Chapter 1.1.13). Common medication combinations involve lithium and one of the above listed antidepressant, or a combination of a tricyclic compound and an SSRI (Anderson et al., 2000; Geddes et al., 2012). Subsequent strategies tested in STAR*D included switching to another treatment, or augmenting existing treatment, which yielded an average remission rate of about 30%. Of interest, STAR*D data also suggested that on average about 40% of patients who had completed treatment with Citalopram also relapsed after about 4 months. Taken together, pharmacological treatment yields remarkable clinical improvement, but individual treatment recommendations need to factor in potential side effects to optimise treatment compliance. Moreover, data suggested that while combining and changing medication can improve response and remission likelihood, about one third of patients remain depressed (Rush et al., 2006).

1.2.1.2.2 Psychotherapy

Besides psychopharmacology, mainstream treatments options currently include various forms of psychological treatment. The main psychological treatment that is being prescribed is cognitive behavioural therapy (CBT), which aims to modify destructive thinking patterns and behaviour in individuals suffering from mood disorder. One main goal of CBT is to equip individuals with more effective strategies to cope with challenges, it thus constitutes a form of therapy and prevention and is in particular recommended for individuals suffering from moderate to severe depression (Geddes et al., 2012). CBT employs mental imagery techniques, for instance to help patients mentalizing themselves in a different context ("reframing") or help individuals in finding relaxation (Holmes et al., 2016; Holmes and Mathews, 2010; Pearson et al., 2015). With regards to the clinical efficacy of CBT, which is often prescribed alongside pharmacotherapy, meta-analyses suggest that it can yield moderate (Cuijpers et al., 2013a; Hofmann et al., 2012) to large (Johnsen and Friborg, 2015) effects.

One challenge in demonstrating evidence for clinical efficacy of CBT is the choice of a suitable control group that can effectively control for unspecific effects including placebo. One prominent type of control involves randomizing patients to a waiting list, but this control group likely only captures relatively slowly occurring spontaneous remission and regression to the mean effects (Keller et al., 1992). Other studies refer to "treatment as usual" control groups

(TAU), against which CBT has shown superiority (Watts et al., 2015). However, the data also suggest that the exact definition of TAU form a heterogenous set of different interventions and treatment providers (Watts et al., 2015). Other meta-analyses have compared the clinical efficacy of psychotherapy in general (including CBT) to pharmacotherapy alone, and CBT to other forms of psychotherapy. Other forms of psychotherapy include psychodynamic therapy and interpersonal therapy. While the psychodynamic therapy mainly focuses on revealing and addressing unresolved unconscious intra and interpersonal conflicts, interpersonal therapy addresses interpersonal conflicts and social adjustment. One meta-analysis found that psychotherapy showed larger effects compared to tricyclic antidepressants (Cuijpers et al., 2013b). It should be noted that this type of antidepressant medication does not represent first line treatment in depression and thus factors such as duration and chronicity that the meta-analysis did not control for may have partly mediated these effects. Another meta-analysis found CBT to be more effective than psychodynamic therapy, but not more effective than interpersonal therapy (Tolin, 2010). Likewise with regards to prevention of relapse effects, both CBT and interpersonal are effective compared to groups that receive no treatment with a number needed to treat of about 22 (Cuijpers et al., 2008). Lastly, also the efficacy of combination therapy (psycho-and pharmacotherapy) compared to monotherapies (pharmacotherapy alone) has been investigated in meta-analyses. Data suggest that adding psychotherapy to medication yields larger clinical effects with a number needed to treat between 4 and 6 (Cuijpers et al., 2014). However, overall the literature suggests that therapeutic effects of CBT depend on factors such as the individual predisposing factors to develop depression, including the presence of other co-morbidities (Cuijpers et al., 2009), but also personality trait factors (Mulder, 2002). Overall, findings reported for CBT treatments require cautious interpretation because the literature is likely strongly affected by publication bias (Cuijpers et al., 2013a, 2010), and hence many reported effect sizes are likely overestimated (Figure 1.1). The effect of publication bias on reported effects in the psychotherapy literature (based on both self-reported and clinician rated measures) is also reflected in a negative trend over time with regards to the size of reported effect sizes, starting around the year 2000 when trial registration was introduced (Johnsen and Friborg, 2015). Similar patterns have also been identified for trials in other fields of medicine (Kaplan and Irvin, 2015) and can be explained by more rigorous and correct reporting that is reinforced by trial preregistration (e.g. avoiding the switch of outcome measures). Insightful epidemiological data of clinical effects have been reported by the STAR*D trial, which found CBT to yield (descriptively) superior remission rates compared to medication when patients switched from an SSRI to CBT. However, CBT

shows only comparable, or slightly inferior effects compared to other pharmacotherapy when used as an augmentation therapy (Rush et al., 2006). Taken together, meta-analytical evidence suggests that CBT yields moderate to large effects in the treatment of depression, although the choice of suitable control conditions remains a challenge. Further, when used in combination with pharmacotherapy or as a switch option, it likely yields larger effects compared to pharmacotherapy alone. However, treatment effects likely depend on the individual etiology and individual factors such as personality traits, which may mediate the acceptance of treatment.

Mindfulness training combines mental exercises such as mental imagery techniques, with relaxation training and attention regulation. Mindfulness has been recommended by the National Institute for Health and Care Excellence (NICE) to prevent relapse into and treat acute depressive episodes and is being supported by the UK's National Health Service (Marx et al., 2015). Meta-analyses suggest that the average effect size of mindfulness training is moderate (Khoury et al., 2013; Strauss et al., 2014) compared to no treatment. However, the technique has shown no superiority compared to active controls that either received CBT or psychoeducation (Strauss et al., 2014). Further, the credibility of mindfulness trials published until 2013 has been called into question by a meta-analysis that found that only one in five published trials were registered, while the remaining trials did not fulfill this basic requirement set for evidence-based medicine. Further, no registered trial specified the primary outcome measure adequately, and over 60% of registered trials remained unpublished up to 2,5 years post-trial completion (Coronado-Montoya et al., 2016), indicating strong publication bias and file drawer effects (Chapter 1.2.1). Taken together, these data challenge the evidence based on which NICE has declared mindfulness training an effective treatment for depressed patients.

1.2.1.2.3 Electroconvulsive therapy (ECT)

For severely depressed patients or patients who are at acute risk to commit suicide, sessions of electroconvulsive therapy (ECT) may be prescribed (Geddes et al., 2012), which involve eliciting electric seizures that are applied to the scalp at sites close to the frontal or temporal lobes. Introduced in 1934, it is the oldest currently performed biological therapy (Payne and Prudic, 2009). ECT treatment is usually performed in an inpatient setting or ambulatory hospital setting because it involves anesthesia in form of a muscle relaxant and sometimes mild sedation (Geddes et al., 2012) and it thus requires medical supervision. ECT usually leads to

side effects such as headaches (due to muscle spasms) and it can also impair cognitive functions. Clinical experience (Geddes et al., 2012) and meta-analytical evidence suggest, however, that the majority of patients show improved cognitive performance (e.g. with regards to processing speed in reading, working memory and aspects of executive function) compared to baseline (Semkovska and McLoughlin, 2010). This may result from general mood improvement and therapeutic effects associated with the treatment. Long-term side effects (e.g. of re-occurring ECT treatment regimens) remain to be systematically investigated (Oremus et al., 2015). With regards to treatment response, ECT shows excellent effects, given the circumstance that it is predominantly used in patients where pharmaco-and psychotherapy has failed. The remissions rates following ECT are high with about 50% (Dierckx et al., 2012), and it may even be higher in patients who show larger treatment response to initial pharmacotherapy (Heijnen et al., 2010). Predictors for ECT response include old age and the presence of psychotic symptoms, a symptom that mostly occurs in severely depressed patients (van Diermen et al., 2018). Also, depression severity itself predicted response rates to a moderate degree, but it did not predict remission rates. Comparisons between the two dominant forms of ECT, bitemporal and high-dose right unilateral ECT, suggest that they are effective to a similar degree with regards to remission rates and change in depression scores. Moreover, both montages show a similar relapse rate. However, immediate side effects such as retrograde autobiographical memory loss are more severe for bitemporal ECT (Kolshus et al., 2017). One challenge in the evaluation of *verum* ECT treatment responses lies in the absence of placebo/sham control groups for the majority of conduced trials (Rasmussen, 2009; Read and Arnold, 2017). Further, remarkable remission rates reported for ECT are accompanied with similarly large relapse rates: according to one meta-analysis, every second patient relapses within 12 months, and every third patient relapses within 6 months despite continuation pharmacotherapy or continuation ECT (Kolshus et al., 2017). Yet, compared to other accepted treatments, these relapse rates remain largely comparable, if not superior to those reported for augmentation pharmacotherapy (Rush et al., 2006). Taken together, ECT remains a viable option for severely depressed patients where it shows remarkable effects, but the field still lacks convincing evidence of superiority over placebo treatment, which remains an admittedly challenging, but nevertheless required test given the ethical implications of ECT (Read and Arnold, 2017).

1.2.1.2.4 Experimental treatment options

The treatment options reviewed in this paragraph are partly approved by the US based Food and Drug Administration and may thus not be universally considered as *experimental*. However, these treatment options are still largely investigated in early phase trials (Chapter 3.4.3) and not widely prescribed. Hence, they are labelled as experimental for this review. Despite various treatment forms that are available in healthcare systems of developed countries, many patients do not respond to treatments prescribed through respective health care professionals. Overall, about one third of patients do not respond to standard treatments (pharmacological and /or psychological), often resulting in repeated hospitalizations and chronic disability (Rush et al., 2006), necessitating further exploration of alternative treatments. Several invasive and non-invasive treatments are currently being explored, including transcranial magnetic stimulation (TMS), transcranial direct current stimulation (tDCS), and deep brain stimulation (DBS). The last decade has seen considerable investment with regards to the non-invasive brain stimulation technique TMS. One meta-analysis published in 2017 included 81 RCTs that had compared TMS treatment to sham stimulation (Brunoni et al., 2017). Different TMS protocols exist that vary in stimulation frequency (e.g. low versus high frequency) and placement of stimulation coils (uni- vs. bilateral). With regards to response rates, more established TMS interventions such as low- and high frequency stimulation as well as bilateral stimulation of the dorsolateral prefrontal cortex yielded substantial effects compared to sham stimulation with OR ranging from 2.3 to 4.6. However, the study did not provide any formal test for publication bias and thus reported findings require cautious interpretation. Of interest, a small number of RCTs have also compared clinical effects of (high-frequency 10 Hz) TMS to ECT and found no group differences, although sample sizes were relatively small (Grunhaus et al., 2003, 2000; Rosa et al., 2006). Given that no equivalence tests or non-inferiority tests were provided, these results hence remain inconclusive and differences in efficacy cannot be ruled out (Chapter 3.4.1.5). For tDCS, fewer RCTs have been published and thus meta-analytical evidence is still scarce. One meta-analysis that included 6 RCTs concluded that patients showed larger reduction in symptoms compared to sham stimulation (Kalu et al., 2012). The study further provided data that indicated successful blinding of patients. Overall response rates were still relatively small with on average about 20% in main treatment groups. It should be noted that the level of evidence for tDCS RCTs in depression is compromised by several factors: relatively few studies with relatively small sample sizes have been published to date. Data suggests significant heterogeneity with regards to reported effects and used montage, rendering comparability

difficult. Further, many studies lack of clinical information regarding severity and chronicity of depression, which limits the generalisability and interpretability of reported findings (Kalu et al., 2012).

DBS represents a set of invasive brain stimulation techniques, which has been mainly investigated for patients suffering from severe treatment resistant depression. Originally developed for the treatment of PD (Chapter 1.3.3.2.3), DBS was subsequently investigated for its therapeutic effects in obsessive compulsive disorder as the first neuropsychiatric application (Delaloye and Holtzheimer, 2014). DBS involves stereotactic neurosurgery and the implantation of stimulation electrodes. Reported perioperative side effects and complications are likely comparable to those reported for PD, where they have been reviewed more systematically (Benabid et al., 2009), including headache, seizure, infection and hemorrhage. Stimulation specific side effects depend on the location of electrodes. Previous trials have investigated the therapeutic potential of various stimulation sites including the nucleus accumbens, Brodmann area 25 (BA25) as part of subcallosal cingulate, ventral capsule/striatum and median forebrain bundle (Delaloye and Holtzheimer, 2014). For instance, one early open label study that targeted BA25 reported around 60% response and 50% remission rates, effects that lasted for several years. These results were largely replicated in larger open label studies that followed. Subsequently, the therapeutic potential of this technology was tested in a recent relatively large multi-center RCT, the Broaden trial (Holtzheimer et al., 2017). In total, 90 patients were randomized to either a main treatment group (N = 60) that received active stimulation from trial beginning, or a placebo control group (N = 30) that received delayed active stimulation after 6 months. A remission rate of 40%, (based on a score $\leq$ 10 on the physician rated Montgomery-Åsberg Depression Rating Scale (Montgomery and Asberg, 1979)) in the primary treatment group after 6 months of subcallosal DBS stimulation was defined as the trial's target. In contrast, data suggested that only about 5% of patients had reached remission after 6 months. Moreover, the placebo group showed a very comparable reduction in their depressive scores, suggesting that observed clinical improvements were largely unspecific responses (Holtzheimer et al., 2017). Of interest, the trial sponsor (and manufacturer of the stimulation device) subsequently decided to pre-terminate the trial and the originally planned sample size of 200 patients was thus not reached. Data from long term follow-up to 30 months indicated, however, that the stimulation group had reached remission rates of 25-30% after 25-30 months. Moreover, this rate was about twice as large compared to the group that did not receive stimulation from the beginning. Hence, DBS remains a promising

brain circuit-based modulation technique that requires further exploration. Further refinement, for instance by means of neural markers, may yield larger clinical responses (Delaloye and Holtzheimer, 2014).

fMRI-NF neurofeedback training represents a non-invasive way to modulate activity of brain areas (Chapter 1.3). The technique can enable patients to learn modulating activity in a specific cortical region using mental imagery and has been increasingly explored to treat psychiatric conditions including depression (Arns et al., 2017; Sitaram et al., 2017). fMRI-NF thus uses elements of CBT within a biopsychological framework that, similar to TMS and tDCS, understands psychiatric symptoms in a causal relationship with neural dysfunction Linden, 2014). Compared to (passive) external brain stimulation, fMRI-NF entails an active component such that patients train positive mental imagery and may eventually experience the ability to self-regulate their own brain activity. It has been suggested that fMRI-NF may exploit cognitive mechanisms that play a role in the etiology and treatment of depression. For instance, by learning to self-regulate brain activity, patients may experiences increases in self-efficacy (Linden et al., 2012; Linden, 2014), which describes an individual's self-reported capacity to cope with challenges (Bandura, 1982) and can be measured with the self-efficacy scale (Sherer et al., 1982). Moreover, compared to brain stimulation, fMRI-NF is spatially not constrained to cortical regions and voxels can be selected in subcortical areas with a relatively high spatial precision (Weiskopf, 2012). Lastly, in search for neurofeedback protocols that are effective in treating depression, or even certain symptoms in depression, it may benefit from a rich body of existing literature that has described cognitive processes that are affected by the disorder and their neural correlates (Arnone et al., 2012; Bar, 2009; Linden, 2012).

1.2.2 Stroke

1.2.2.1 Main clinical characteristics and pathophysiology

Stroke is the leading cause of adult disability and third leading cause for death (Losseff et al., 2009). Although stroke mainly occurs in elderly patients, it may occur at any age and even before birth. Perinatal stroke represents the leading cause of cerebral palsy and about 25% of strokes occur before the age of 65. In an increasingly aging population, the prevalence of stroke in high-income countries among individuals below 75 years has increased from about 0.4 to 0.5% between 1990 and 2000, while the incidence has decreased from about 0.15% to 0.14%

(Feigin et al., 2014). Data from individuals above 75 years shows a similar trend although changes are larger by the order of one magnitude (incidence 1990 to 2010 decreased from 3.2% to 2.6% and the prevalence increased from 4.9% to 6.2%). The costs of patient care are high and currently require an allocation of about 5% of the total health care budget in the UK. Regarding monetary investment into stroke research, stroke research represents an underfunded field in the UK relative to its health and economic impact on society and compared to other biomedical research areas including cancer and coronary heart disease (Luengo-Fernandez et al., 2015). Specifically, for every 10£ that is being spent for health and social costs associated with a respective health care condition, stroke research received in 2012 about 0.19£, making up less than 1/5 (cancer) and less than 1/3 (coronary heart disease) of received research funding in other areas.

Cerebral stroke is defined by a neurovascular injury that leads to hypo oxygenation in neighbouring parenchyma, resulting in cell death. Most cerebral infarcts are ischemic. In about 10% of cerebrovascular stroke brain vessels cannot resist the increased pressure and hence tear, leading to cerebral bleeding and a haemorrhagic stroke (Losseff et al., 2009). The aetiology of a cerebrovascular event can be categorised intro three main types: 1) endogenous causes including infection, chemical agents, 2) exogenous causes including traumatic brain injury, and 3) a combination of both. Haematogenic thrombi are the main endogenous cause. Risk factors include endogenous predispositions, including vessel malformations and genetic variants of lipoproteins, as well as modifiable exogenous risk factors such as smoking, and diabetes type II. Irrespective of the cause, hypertension and immobility are main risk factors for thrombosis, which may result in an embolism causing an infarct (Losseff et al., 2009). The time passed after stroke determines whether cases are classified as acute (1-3 weeks), subacute (1-3 months) or chronic stroke, although different classifications have been suggested (Page et al., 2008; Tohgi et al., 1990)

The vascular system that supplies the brain with oxygenated blood can be subdivided into an anterior and posterior circulation. The internal carotid artery (ICA) gives rise to the middle cerebral artery (MCA), which itself gives rise to the anterior cerebral artery (ACA). Vertebral arteries give rise to the basilar artery, from which smaller branches supply the brain stem structures including the pons, as well as the cerebellum. The basilar artery gives rise to the posterior cerebral artery (PCA), which communicates with the ACA, and it further intersects via smaller communicating arteries with the ICA via posterior communicating arteries. Also,

the left and right ACA form connections via anterior communicating arteries. This vessel architecture forms a collateral circulation system that has been described as the "Circle of Willis" (DeFelice et al., 2007; Liebeskind, 2003; Losseff et al., 2009). Anastomoses allow compensation for changes in blood perfusion and can maintain constant levels of blood flow to a certain extent, thereby preventing ischemic damage (Liebeskind, 2003). Major anastomoses are also formed between cortical branches of the ACA and MCA. Thereby, the system can compensate for perfusion deficits in the event of an infarct. Besides these main arteries, smaller vessels supply the brain stem (basilar arteries) and the cerebellum (anterior and posterior cerebellar artery).

Clinical signs resulting from stroke depend on the location of the cerebral infarct and the affected vessel. For instance, infarcts in the PCA may result in cortical signs, including dysphasia (receptive or productive) including spatial neglects, hemianopia or other forms of cortical blindness (Losseff et al., 2009). The MCA is divided into several branches which supply frontal regions (upper branch), temporal brain areas (lower branch), and the basal ganglia (sphenoidal segment). Infarcts that involve the MCA perfusion territory, as confirmed by a CT, represent the most common types of stroke (Heinsius et al., 1998). MCA infarcts may result in hemiplegia, in particular of the upper limb and proximal lower limb, hemisensory loss, but also hemispatial neglect, and apraxia. Infarcts of distal branches that perforate parts of the frontal, parietal and temporal lobule often cause cortical signs. The ACA mainly supplies frontal regions, including the medial aspect of the motor cortex. Infarcts of more distal branches result in hemiparesis of the lower limb (Bogousslavsky and Regli, 1990). Further, patients may show impaired motor planning, also called ideomotor apraxia (Losseff et al., 2009) if aspects of the supplementary motor area (SMA), a premotor area that plays a key role in motor planning, are affected. Motor planning deficits occur mainly during acute stroke stages and may improve during subacute and chronic stroke phases (Brugger et al., 2015). In their most extreme form, ACA strokes may cause an alien hand syndrome such that the patient loses the sense of agency over the control of the affected limb. Productive aphasia (i.e. impaired speech production) and mutism are associated with left ACA stroke in particular (Kumral et al., 2002).

Stroke is mainly diagnosed clinically and radiologically and indicated by a sudden acute neurological deficit (Kim et al., 2013; Losseff et al., 2009). Besides neurological examinations, neuroimaging facilitates the diagnosis and helps to determine the location and extent of the infarct. For ischemic stroke, MRI seems the most sensitive modality to image lesions, whereas

for haemorrhagic stroke, Positron emission tomography–computed tomography (PET/CT; without any contrast enhancements) is slightly more sensitive compared to MRI (Chalela et al., 2007), however, it is rarely used. CT scans remain the method of choice to detect haemorrhages. Albeit not included in guideline recommendations yet, molecular imaging may provide more sensitive diagnostic and prognostic information markers in the future (Kim et al., 2013).

1.2.2.2 Treatment

1.2.2.2.1 Acute treatment

Stroke is an acute event that can be treated acutely to decrease neural damage and prevent functional impairment of stroke survivors. Specifically, causal treatments for acute ischemic stroke mainly aim to maintain cerebral hyper perfusion using pharmacological anticoagulation and, if available within the first four to five hours, pharmacological thrombolysis (Alper et al., 2015). Given the limited time window and availability of specialised stroke units though, only an estimated small fraction (5-15%) of stroke survivors can receive acute treatment options (Cramer, 2018; Grefkes and Ward, 2014). Furthermore, neurosurgical clot removal remains an option if performed within up to seven hours of symptom onset (Saver et al., 2016), although radiological thrombectomy seems to yield preferable outcomes and is increasingly being used in large stroke centres (Jansen et al., 2018). Long term treatment of stroke aims to prevent future thrombosis in the form of pharmacological anticoagulation and life style modification, and it targets cognitive and motor deficits using physiotherapy.

1.2.2.2.2 Physiotherapy and constraint induced movement therapy

Rehabilitation treatment for stroke survivors mainly consists of physiotherapy, which combines various techniques including strength, endurance and coordination training. Meta-analytical evidence suggests that 30-60min physiotherapy training up to five days per week can yield small to medium improvement of motor function (Pollock et al., 2014). However, about one-third of patients with stroke suffer from persistent disability (Dimyan and Cohen, 2011). Several new therapies that target motor recovery of the hand and arm are currently evaluated with regards to their clinical efficacy. These include constraint induced movement therapy (CIMT), TMS (Schaechter, 2004), and motor imagery-based techniques (Sharma et al., 2006; Zimmermann-Schlatter et al., 2008). In CIMT, patients' non-paretic limb is constrained in order to promote the use and recovery of the paretic limb: stroke causes mostly unilateral brain

lesions with contralateral hemiparesis and in consequence, patients often learn to compensate with their initially not impaired "good" limb, which results in "learned non-use" of the paretic limb (Taub et al., 2002). Such behavioural compensation is postulated to counteract motor skill rehabilitation via transcallosal projections and interhemispheric competition (Allred et al., 2010). The framework of interhemispheric competition suggests that hyperexcitability of the unaffected (contralesional) hemisphere enhances inhibition of the affected (ipsilesional) hemisphere and hampers sufficient motor rehabilitation (Nowak et al., 2009; Hummel and Cohen, 2006). One meta-analysis across 15 RCTs found that CIMT treatment was associated with small additional improvement of arm motor function (~ Cohen's d = 0.3)compared to treatment as usual (Corbetta et al., 2015).

1.2.2.2.3 Non-invasive experimental therapies

Some TMS protocols have also been designed to modulate pathological interhemispheric inhibition (Ward and Cohen, 2004), either by repetitive TMS stimulation (e.g. at 3Hz) of the affected hemisphere to increase its activity, or by applying inhibiting TMS pulses over the intact hemisphere, thereby decreasing its suppressive projections to the affected hemisphere ("re-facilitation"). However, the concept of disbalanced interhemispheric inhibition has been recently challenged by findings from a meta-analysis across 112 studies (McDonnell and Stinear, 2017). The authors reported no evidence for hyper-excitability of the unaffected hemisphere or imbalances in interhemispheric inhibition, which forms a main assumption of the interhemispheric inhibition disbalance model. The authors hence concluded that active stimulation of the affected hemisphere may be more effective in promoting post stroke recovery. Indeed, while this work suggests that the mechanism of actions remains unclear, it should be noted that moderate effects compared to sham stimulation have been reported for (ipsilesional) M1 TMS stimulation, in particular subcortical stroke patients (Hsu et al., 2012).

Another form of adjunctive therapy that is currently being investigated for its potential in restoring motor function is motor imagery training, which constitutes a form of mental practice (Sharma et al., 2009b, 2006). The neural circuitry involved in motor imagery largely overlaps with the circuitry involved in motor execution, as shown by functional neuroimaging studies (see for review; Hétu et al. 2013). The available literature of mental practice-based trials in stroke motor rehabilitation is limited to a small number of studies. One meta-analysis reviewed effects reported in 5 placebo controlled studies and found that mental practice combined with

physiotherapy outperformed physiotherapy training alone (Barclay-Goddard et al., 2011). The associated effect size was found to be large, but it requires cautious interpretation and further (independent) replication given the small number of studies that mainly stem from one research group (Figure 1.1).

Robot-assisted motor training is another interactive, high-technology approach that is being investigated in clinical trials for its potential to promote upper limb recovery after stroke (Turner et al., 2013). In robot-assisted motor training patients learn to control a robotic manipulandum with their affected limb. Various paradigms have been developed that can be adapted for training needs of the individual patient, including training of coordination and force. Most robot manipulanda setups consist of a robotic arm that allows patients to control a cursor on the screen. Motors of the manipulandum can be programmed to interact physically with patients' limb kinematics, for instance to facilitate movement initiation, guide whole arm movements along a desired trajectory, or perturb movements by creating resistance during goal-oriented reaching movements. Other more recently developed paradigms have introduced implicit reinforcement learning that allow participants to learn new joint configurations without conscious awareness of underlying task dimensions (Mehler et al., 2017b). Compared to other experimental treatment forms, the literature of robot-assisted motor training features more published RCTs, providing a more robust estimate of evidence. Based on 31 trials published until February 2015, a meta-analysis estimated a small to medium effect sizes for motor function improvement of the upper limb as well as increase in arm muscle strength (Mehrholz et al., 2015).

Brain Computer Interfaces (BCI), including neurofeedback training represent a more recently developed set of techniques that combine motor imagery training with a reinforcement signal. BCIs are brain machine interfaces (BMIs) that use real-time analysis of brain signals that are transformed into either a form of feedback or a control command for an external device (Birbaumer et al., 2009). Neurofeedback experiments are one specialised form of BCIs, but not all BCIs provide neurofeedback. During BCI experiments, recorded brain signals are usually amplified and filtered and then submitted to an algorithm that computes a brain activity measure of interest. Target signatures may include power changes of a specific frequency band, lateralisation differences in activity between hemispheres, or more generally multivariate classification of different brain states (Chaudhary et al., 2016). The output of the computation can then be transformed to provide feedback to the participant, for instance in form of a visual

thermometer that represents the activity strength during motor imagery, auditory feedback whose amplitude varies with activity strength, or a visual arrow that indicates the decoded direction of imagined (left or right) hand movements. Alternatively, the output can be used to control more complex external technology such as a spelling device to study basic communication capacities in patients suffering from complete locked-in syndrome (Chaudhary et al., 2017), a robotic arm (Chapin et al., 1999) as a mind controlled protheses (Pruszynski and Diedrichsen, 2015), or non-invasive brain stimulation (e.g. TMS) of an area of interest (see overview of ongoing trials in Chaudhary et al. 2017).

Despite a growing interest in BCIs for motor rehabilitation (Chaudhary et al., 2016), currently available evidence from controlled clinical studies conducted in stroke patients remains limited. A recent meta-analysis identified 9 controlled BCI studies that targeted upper limb functional recovery (Cervera et al., 2017). The authors found moderate evidence compared to control groups, but they also noted that studies varied considerably with respect to reported effect sizes, as well as with respect to the type of control that was used. Of interest, while the authors found that clinical effects exceeded the MCID defined for stroke (a reduction of 5.25 points measured on the Fugl-Mayer scale) in 8 out of 9 trials, however, the superiority of the main over the control group exceeded the MCID in only 3 studies (Cervera et al., 2017).

Lastly, one meta-analysis compared the effectiveness of different adjunctive motor rehabilitation approaches for upper limb treatment (Langhorne et al., 2009). The review identified 19 different categories of interventions, including mental practice (i.e. motor imagery) based techniques, reinforcement feedback based robotic therapy and CIMT. Direct comparisons revealed that these treatments seemed most effective in promoting upper arm recovery. Of interest, neurofeedback training combines some of these elements and may further specifically target neural correlates that have been identified after conventional treatment and during natural recovery.

1.2.3 Parkinson’s disease

1.2.3.1 Main clinical characteristics and pathophysiology

Idiopathic Parkinson’s disease (PD) is the second most common neurodegenerative disorder, affecting 1% of people over the age 60 and 3% of people over 80 years (de Lau & Breteler,

2006). Cardinal symptoms in early PD stages include tremor (shaking), rigor (stiffness), bradykinesia (slowing of movements) and later akinesia (absence of movement), as well as postural instability. These motor symptoms form the basis of the clinical diagnosis Parkinson's disease. Besides motor symptoms, it should be noted that PD patients also suffer from non-motor symptoms including sleep disturbances, mood disorders and cognitive impairment (Marsili et al., 2018).

The underlying molecular neuropathology is complex and affects multiple neurotransmitter systems including the serotonergic raphe nuclei, the cholinergic nucleus basalis, as well as the dopaminergic midbrain (Losseff et al., 2009). Central for an understanding of key motor symptoms and successful pharmacological therapy are changes of the dopaminergic midbrain due to cell death of dopaminergic cells in the substantia nigra pars compacta (SNc). The SNc forms an integral part of the basal ganglia (Fig. 5.1). Two pathways between the SNc and the putamen have been identified that are dopamine mediated and act via excitatory (D1 receptor mediated) and inhibitory (D2 receptor mediated) pathways on the putamen. In healthy individuals, excitatory dopaminergic simulation from the SNc to the putamen disinhibit the globus pallidus internus (GPi) and thereby the thalamus. In consequence, the thalamus sends more glutamergic excitatory projections to the cortex (historically known as the direct pathway). With increasing loss of dopaminergic cells in PD, this pathway fails to disinhibit the thalamus and reduces excitatory stimulation of the cortex.

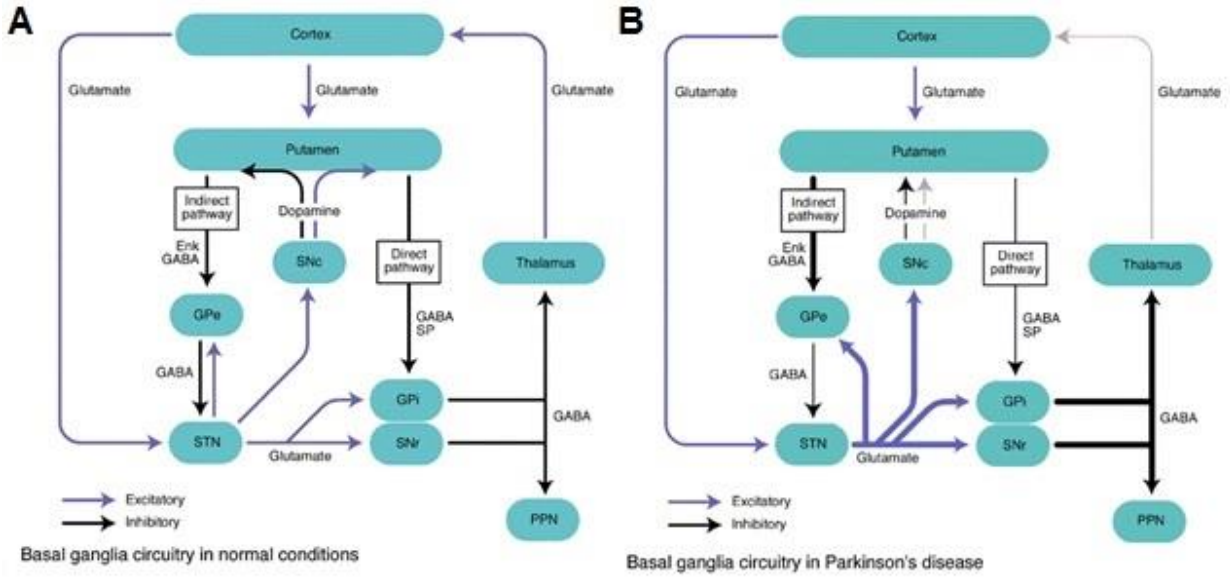

Figure 1.3 Schematic diagram of basal ganglia circuitry in healthy controls (A) and PD patients (B), modified from (Lewis et al., 2003).

A second pathway (historically known as the indirect pathway) comprises inhibitory projections from the SNc to the putamen, which disinhibit globus pallidus externus (GPe). As a result, the GPe disinhibits the thalamus, which itself is suppressed in its activity via the GPi (as described above), as well as the substantia nigra pars reticularis (SNr). Loss of dopaminergic projections from the SNc result in an imbalance such that the (movement suppressing) indirect pathway becomes hyperactive and the (movement disinhibiting) direct pathway becomes hypoactive (Figure 1.3 A, B). Moreover, projections between the STN and motor cortices (hyperdirect pathway) mediate the activation and suppression of motor programmes by interacting with the GPi and SNr (Nambu et al., 2002). Cortical structures such as the supplementary motor area (SMA) and the primary motor cortex (M1) send and receive direct projections to and from the STN. Although the role of the hyperdirect pathway in the development of motor deficits is currently less well understood (Nambu, 2008), it likely promotes voluntary movements in healthy individuals and that this function is dependent on dopamine (Chu et al., 2017). Disturbances of the pathway due to dopaminergic loss contribute to poverty, slowness and eventually absence of movement in PD patients (Nambu et al., 2015; Rodriguez-Oroz et al., 2009). Taken together, alterations in three neural pathways central to motor control have been described in PD and related to slowed movements and rigid joint control. With disease progression, patients also increasingly suffer from cognitive deficits, including learning difficulties, memory loss, attention deficits and dementia that manifest in 80-90% of late stage PD patients (Biundo et al., 2016; Gratwicke et al., 2015).

Disease progression can be classified into different stages based on the severity of motor symptoms. Two widely used staging systems for PD are the Hoehn-Yahr (HY) score, and the modified Unified Parkinson's Disease Rating Scale (MDS-UPDRS). The HY classification grossly defines PD by five stages based on the extent of motor impairment (Table 1.3). The MDS-UPDRS scale, in contrast, is a continuous scale that is partly based on patients self-reports, partly on ratings by a (blinded) assessor (Goetz et al., 2008). The MDS-UPDRS scale comprises four subparts that focus on the impact of non-motor symptoms such as mood, perception and vegetative functions (Part I), a self-evaluation of the impact of PD on activities of daily life including handwriting, tremor and hygiene (Part II), a clinician-scored monitored evaluation of motor symptoms including rigidity, bradykinesia and various evaluations of

tremor (Part III), and lastly evaluations of therapy complications including dyskinesia (Part IV).

Stage	Hoehn and Yahr Scale	Median total months to transit
1	Unilateral involvement only usually with minimal or no functional disability	-
2	Bilateral or midline involvement without impairment of balance	20
3	Bilateral disease: mild to moderate disability with impaired postural reflexes; physically independent	107
4	Severely disabling disease; still able to walk or stand unassisted	131
5	Confinement to bed or wheelchair unless aided	157

Table 1.3 Hoehn and Yahr Scale staging. Months in far right column show the median time until a respective stages was reached in an observational study with over 690 patients (Zhao et al., 2010).

1.2.3.2 Treatment

1.2.3.2.1 Pharmacological treatment

Pharmacological treatment represents the main treatment for PD and focuses predominantly on compensating for the loss of dopaminergic SNc neurons. Different pharmacological agents that act on the dopaminergic circuitry from different angles are available. These include dopamine receptor agonists that mimic dopamine effects, monoamine oxidase inhibitors (MAOI) that prolong the presence of available dopamine, and Levodopa (L-Dopa) as a brain barrier permeable precursor amino acid of dopamine that is converted into Dopamine by an enzyme (Aromatic L-amino acid decarboxylase). L-Dopa, however, may lead to involuntary excessive

movements (dyskinesia). Up to 40% of patients may experience dyskinesia after 4-6 years of treatment (Ahlskog and Muenter, 2001). To minimise the risk for such motor complications, guidelines recommend using receptor agonists or MAOI in early disease stages, thereby postponing treatment with L-Dopa to later disease stages (Connolly and Lang, 2014). On the other hand, dopamine agonists may have an adverse side effect profile including somnolence, neuropsychiatric symptoms such as hallucinations and impulse control disorder, but also (partly reversable) fibrosis of organ cavities such as the pericardium or lung pleura (Antonini et al., 2009). Overall, pharmacological treatment is an essential component in managing motor symptoms in PD patients. Yet, benefit-to-risk ratios must be considered by the clinician to minimise side effects and limit treatment discontinuation.

1.2.3.2.2 Physiotherapy

For physiotherapy, effects of exercise training have been investigated with respect to gait and postural stability. Meta-analyses found moderate effects for health care facility based compared to home based interventions (Klamroth et al., 2016; Shen et al., 2016). Likewise, another recent review concluded that physical therapy training can modify long-term motor symptoms and shows promise with regards to balance. Further, new technology such as Google Glass could be used to support patients by cuing movement (Mak et al., 2017). With regards to alternative forms of movement therapy, one meta-analysis analysis of dance therapy interventions reported moderate effects on gait stability, improvements for the UDPRS-III, and increased self-reported quality of life (Sharp and Hewitt, 2014). However, it should be noted that in particular earlier physiotherapy studies show strong indications of publication bias and lacked meaningful follow-up periods (Gage and Storey, 2004). Although it remains to be tested, study quality may also have improved for physiotherapy trials in PD with the introduction of trials registration and CONSORT, resulting in more published null findings. Noteworthy, two independent recent well controlled RCTs question the futility of physiotherapy as an add-on treatment in PD: the studies compared changes in motor symptoms and quality of life for different forms of physiotherapy (patient-centred and high intensity exercise) that were prescribed in addition to pharmacotherapy versus TAU (i.e. pharmacotherapy alone) and found no clinically meaningful (MCID) superiority for additional physiotherapy (Clarke et al., 2016; Schenkman et al., 2017). Given these findings, larger futility trials are warranted. Noteworthy, although group means did not show clinically meaningful changes (>3.5 points) compared to TAU, individual data points in either group indicated that some patients improved remarkably

(by about 20 points on the UDPRS-III), whereas others deteriorated to a similar degree. Taken together, albeit widely prescribed in clinical practice, the evidence for the therapeutic value of physiotherapy remains unclear. Moving the focus beyond group means, however, data indicates that some patients deviate largely in their response. Employing cross-over trial designs in combination with detailed phenotyping of patients (including a personal sports and exercise histories) could allow identifying predictors for whether patients would benefit or deteriorate from a specific treatment form.

1.2.3.2.3 Invasive deep brain stimulation (DBS)

Deep brain stimulation (DBS) remains a treatment option for individuals where pharmacological treatment and adjunctive physiotherapy do not provide sufficient improvement of motor symptoms, or where medication associated dyskinesia becomes unbearable (Benabid et al., 2009). DBS involves the implantation of electrodes whose activity is controlled by a battery-operated implantable pulse generator. The procedure involves stereotactic neurosurgery, requiring that patients are in a general good health condition. DBS stimulation protocols mostly operate with high-frequency stimulation (~140Hz) of the STN, but also the GPi and thalamus may be targeted and provide good clinical outcome in reducing resting tremors. Optimal stimulation frequencies are determined in patients individually during follow-up phase (Benabid et al., 2009). Identifying the underlying neurophysiological mechanism(s) of action remains an active field of research and several models have been suggested based on electrophysiological recordings in PD patients, animal models, and computer simulation work (Chiken and Nambu, 2015; Santaniello et al., 2015; van Wijk, 2017). However, one main observation that has replicated in multiple studies is an amplitude reduction in beta frequency band (13-30Hz) after STN (Hammond et al., 2007; Santaniello et al., 2015; van Wijk, 2017) and pallidal DBS (Wang et al., 2018), which correlates with improvements of motor symptoms. Within the STN, beta oscillations are mainly detected within the dorsolateral portion, which is organised somatotopically (Miyachi et al., 2006). The potential significance of this finding in the search for potential electrophysiological markers for EEG-NF training will be discussed in Chapter 8. DBS has been found highly effective in reducing motor symptoms. One comprehensive meta-analysis based on 15 years clinical trial experience with DBS treatment in advanced PD patients reported 30-70% improvement for rigidity, 60-90% improvement of resting tremor, and 30-70% on the UPDRS-III. As a result, patients' medication dose (measured in the equivalence of L-Dopa) could be decreased by 20-

65% (Benabid et al., 2009). In consequence, also medication related dyskinesia decreased in a substantial proportion of patients (40-90%). Recent work suggests that the hyperdirect pathway plays a pivotal role in the clinical effects that can be achieved with DBS (Akram et al., 2017; Chou et al., 2015; Muthuraman et al., 2017). This suggests that direct communication between these two areas may relate to some key clinical motor features observed in PD.

Regarding the side effects that can occur during DBS treatment, one can distinguish between complications arising from the implantation procedure (peri-operatively), from physical irritation of surrounding tissue by the hardware, and from the high frequency stimulation itself. Implantation related complications mainly include haemorrhages, which occur on average in about 8% of patients, and postoperative confusion (about 10%). Peri-operative mortality can occur, but currently available data limits estimating its prevalence. One study found that in 109 DBS PD patients that received some of the first available STN implantations between 1998 and 2002, two patients died due to complications related to surgical procedures (embolisms and pneumonia) that can occur particularly in elderly patients (Umemura et al., 2003). The second type of complications are related to the implanted hardware and are more prevalent. These complications mainly consist of infections with reported incidences varying largely between studies (from 2.5% up to 50%), as well as malfunctioning of hardware parts. Although hardware failure represents a relatively common complication among all DBS treatment associated complications, it is well manageable in most cases and requires treatment discontinuation in a fraction (~6%) of patients (Benabid et al., 2009). Lastly, simulation related neurological side-effects can occur, but these are rarely permanent and can be reversed, e.g. by optimising the stimulation protocol. Stimulation related side effects include dysarthria and hypophonia (4-17%), as well as voltage-dependent dyskinesia, which can be reduced or stopped by reducing the stimulation strength (Benabid et al., 2009). To summarise, DBS treatment provides effective reduction of key PD motor symptoms and can reduce motor complications that result from dopamine replacement therapy. Moreover, DBS treatment in combination with simultaneous recordings provide a better understanding of electrophysiological correlates of movement deficits observed in untreated PD patients. However, the risk of side effects and their potential impact on the patient need to be considered on a case by case basis.

1.2.3.2.4 Non-invasive experimental forms of treatment

Experimental nonpharmacological and non-invasive interventions include brain stimulation (Chou et al., 2015), motor imagery training (Harris and Hebert, 2015), and neurofeedback training (Esmail and Linden, 2014). These techniques are being investigated for their clinical potential as add-on therapies, and if they are found being effective, they may help improving motor symptoms, delay disease progression, or possibly allow reducing the dose of on-going pharmacological treatment. However, current evidence for the effectiveness of all these techniques is scarce. For non-invasive TMS brain stimulation, one recent meta-analysis suggested moderate superiority compared to placebo control sham TMS (Chou et al., 2015). Data further indicates that effects for certain brain sides and stimulation parameters may lead larger clinical benefits. For instance, high-frequency (>5Hz) TMS over M1 seems to show higher clinical efficacy compared to high-frequency TMS over frontal regions. With regards to motor imagery training, preliminary work found that motor imagery training in combination with physical exercise can reduce bradykinesia (Tamir et al., 2007). These findings have more recently been corroborated in a randomised, single-blinded controlled trial with a relatively large sample size (N = 66) that included 40 training sessions of mental imagery training per patients over 8 weeks (Ajimsha et al., 2014). The main treatment group received in addition to physical exercise training instructed mental imagery training that consisted of motor imagery exercises, but also relaxation and body scanning mentalisation techniques. The control group received physical exercise training only. Both groups showed significant improvements in motor symptoms on the UPDRS-III, but the combination group showed a larger improvement of 6.8 points at the primary end-point, including a stronger reduction of bradykinesia (2 points) and resting tremor (1.5 points). Superiority was also maintained four weeks later at follow-up with 4.3 points, a group effect that lies within the range of MCID defined for PD motor symptoms (Horvath et al., 2014). Lastly, neurofeedback training, which combines motor imagery training with positive reinforcement learning (Chapter 2.1.1), may provide another way to target brain activity non-invasively. Neurofeedback in PD has been investigated using fMRI and EEG in a small number of studies (see for review Chapter 2.1.5 and Chapter 8).

2 REVIEW OF CLINICAL REAL-TIME FMRI NEUROFEEDBACK APPLICATIONS

2.1.1 Principles of learning and transfer

Neurofeedback training constitutes a special form of Brain Computer Interfaces (BCI) and hence a special form of biofeedback training (Birbaumer et al., 2013). In biofeedback training, biological signals, for instance the pulse or skin conductivity, are measured in real time. Signal intensities are transformed to a measure of activity that reflects the training goal, for instance a percent signal change compared to a rest period. The obtained value is then transformed into a feedback cue, for instance a thermometer with filling bars, that reflects the activity level and that is fed back to the participant. The goal of biofeedback techniques is to train participants in attaining volitional control over the underlying signal that informs the feedback they are presented with. At first, participants usually have no concept of the cause of signal, although initial strategies such as mental imagery techniques may be provided to accelerate the early learning phase. Through trial and error, participants can learn to manipulate the signal by the means of instrumental operant (reinforcement) learning (Birbaumer et al., 2013). Classical operant learning requires a contingent association between a stimulus that is being conditioned (CS) and a reinforcing signal that is an unconditioned signal (US). In the context of neurofeedback training, the CS represents the brain response that is being trained, and the US constitutes the feedback signal which acts as a reward. The specific task that participants perform during training (e.g. mental imagery) constitutes the operant behaviour that is being rewarded. Of interest, it has been shown that learned up-regulation of target motor areas through motor imagery training can be modulated by the presentation of cues that have been rewarded in a different context (Mendelsohn et al., 2014). Specifically, activation in motor imagery related areas as well as areas associated with motivation in general (e.g. the ventral striatum) could be enhanced during concurrent presentation of reward-related cues. This finding indicates that the association between CS and US can be modulated by Pavlonian-instrumental-transfer (PIT) learning, rendering mental imagery-based reinforcement biofeedback training a viable option to modify cognition and behaviour. PIT represents a key concept in the study of cue related behaviour and cognition (Cartoni et al., 2016) and may thus provide a therapeutic mechanism in the treatment of psychiatric conditions, including depressive disorder (Huys et al., 2016), but also in neurorehabilitation by modifying preference of limb use after neural injury (Hosp and Luft, 2013; Mosberger et al., 2016).

The first reports of neurofeedback training used electroencephalography (EEG) in the late 1950’s (Thibault et al., 2016). During EEG neurofeedback (EEG-NF) training,

electrophysiological activity recorded from the scalp comprises the biological signal that participants are trained to control (see Chapter 3.1 for a more detailed description). Real-time functional magnetic resonance imaging neurofeedback (fMRI-NF) is a rather recent development (Weiskopf et al., 2003). During fMRI-NF training, the blood oxygenation level dependent (BOLD) contrast represents the biological signal that participants are trained to control. The BOLD contrast is an indirect measure of neural activity because its biological basis is the local level of deoxyhaemoglobin (see Chapter 3.2 for a more detailed description). BOLD is regulated by various molecular pathways that are linked neural computations, the vascular architecture and metabolic demands (Shmuel, 2010). Hence, during real-time fMRI-NF training, the biological signal that is being fed back to the participant is partly modulated by neural activity, but also other non-neural factors that can modulate the BOLD signal including changes in breathing, heart rate and carbon dioxide (Murphy et al., 2013). Moreover, it depends on the ability of brain vessels to adapt local arterial blood supply by the means of constriction and dilatation in response to changing metabolic demands (that result from variations in neural activity during a given task). Birbaumer and colleagues have thus hypothesised that successful NF training constitutes a form of skill learning and highlighted the central role of the basal ganglia, which form a key structure in reward processing and reinforcement operant learning. This conjecture was partly supported by a recent meta-analysis (Emmert et al., 2016), which reported that parts of the basal ganglia, as well as the anterior insula were activated across various fMRI-NF paradigms. However, it should be noted that this analysis did not account for mere self-regulation attempts in the absence of feedback. Indeed, earlier work suggests that these areas, together with structures such as the supplementary motor area (SMA), form part of a frontoparietal network that is involved in cognitive control (Ninaus et al., 2013). Hence, the specificity of neural correlates identified during fMRI-NF tasks remains to be shown to allow dissociating in how far these reflect true learning of self-regulation, or merely exerting cognitive control. Additional models that have been proposed to describe neurofeedback learning mechanisms are summarised in (Sitaram et al., 2017).

Besides neural and behavioural effects co-occurring during neurofeedback training, training effects that last after training periods are of interest. Briefly, lasting training effects can be achieved at different levels of organisation, representing different degrees of generalizability. For instance, individuals may learn to self-regulate the activity of an area or a network when no feedback is being provided. Depending on the way the feedback is being calibrated, individual successful strategies may vary with regards to content and modality between

participants. The ability to self-regulate activity is predominantly tested with *transfer runs* during which participants apply learned mental strategies but do not receive feedback. Another training goal may aim to enhance mentalising specific content (e.g. motor imagery, or positive autobiographical mental imagery), for which fMRI-NF training mainly represents a training vector that enables participants to develop successful mentalising strategies. Such training effects can be tested using mental imagery questionnaires or mental imagery-based tasks where accuracy and reaction times provide additional measures of mental imagery performance. Furthermore, successful training strategies may be associated with additional training effects, and for instance generalize as a behavioural effect that can be detected with a psychometric measure, a clinical scale or an independent behavioural task. Generalisation beyond the training task constitutes the ultimate goal behind fMRI-NF training interventions because they can modify behaviour in a way that yields favourable clinical outcomes, for instance by reducing pathological, and/or increasing physiological or compensatory behavioural strategies. (Arns et al., 2017). Studies that aim to demonstrate generalisability of training effects ideally include adequate control groups, randomised treatment allocation and blinded assessment (RCTs), allowing to demonstrate specificity of training effects and minimise potential biases (see Chapter 3.4.4). Besides behavioural changes, physiological changes may result from training effects. These include functional measurements such as peripheral skin conductance, or the covariance structure of fMRI time series between brain regions, as well as changes related to the integrity of white matter pathways (Sitaram et al., 2017).

2.1.2 Clinical applications of fMRI-NF

Main training goals of fMRI-NF paradigms include reinforcing specific mental strategies of participants and their associated neural states. Clinical fMRI-NF training based on motor regions may for instance aim to improve stroke survivors' ability to imagine movements performed with a lesioned arm (Sharma et al., 2009b; Sitaram et al., 2012). Another application for fMRI-NF may be to support patients in achieving a mental state with the aim to reduce specific symptoms. Patients suffering from depression for instance often report lacking the ability to mentalise positive thoughts. fMRI-NF training based on regions that are associated with positive affect may for example aim to facilitate this process. These examples constitute a form of supervised mental imagery training, in which the teaching signal is provided by the feedback that indicates activation of a region of network that is of interest for the individual condition. Most neurofeedback experiments that have been conducted to date represent a form of supervised mental imagery training (Sulzer et al., 2013). Compared to traditional EEG-NF

or EEG based BCI paradigms, fMRI-NF provides a relatively high spatial resolution and thus allows targeting motor regions with higher specificity. In particular, the ability to achieve whole brain coverage renders fMRI-NF a promising technique for non-invasive neurological and psychiatric rehabilitation, allowing to provide feedback from subcortical brain areas (Weiskopf, 2012). Most current fMRI-NF protocols instruct participants to engage in a form of mental imagery (Sulzer et al., 2013), but neurofeedback has also been provided on other tasks such as motor execution (Neyedli et al., 2017). Another set of techniques that employs multivariate statistical analyses (decoded neurofeedback training; DecNef) does not require conscious mental strategies. The technique entrains activity patterns associated with desired perceptive or attentive states, opposed up- or down-regulation training of regions in conventional univariate fMRI-NF (Shibata et al., 2011; Watanabe et al., 2017). A detailed review of DecNef would exceed the scope of this thesis, but discussions in Chapters 5 and 9 will touch on potential advantages compared to conventional fMRI-NF training.

Various fMRI-NF paradigms have been piloted for application in psychiatric (Arns et al., 2017; Sulzer et al., 2013) and neurological patients over the last seven years (Wang et al., 2017). With regards to psychiatric applications, proof-of-concept (PoC) studies have been conducted in patients suffering from substance use disorder involving alcohol or cocaine, as well as in nicotine dependence (Hartwell et al., 2016; Karch et al., 2015; Kirschner et al., 2018; Li et al., 2012). For anxiety disorders, fMRI-NF training has been explored in patients suffering from general anxiety disorder, phobia and post-traumatic stress disorder, demonstrating general feasibility (Cordes et al., 2015; Gerin et al., 2016; Hampson et al., 2012; Keynan et al., 2016; Scheinost et al., 2013; Zilverstand et al., 2015). Other studies have demonstrated general feasibility in patients suffering from unipolar depressive disorder (Linden et al., 2012; Young et al., 2014). However, evidence for clinical efficacy from randomised controlled studies conducted in patients remains scarce. To-date (January 2018), the published literature of fMRI-NF for psychiatric conditions consists of only one registered RCT that was conducted in depressed patients (Young et al., 2017). With regards to other psychiatric conditions, several RCTs are currently registered on clinicaltrials.gov (as of January 2018) that investigate therapeutic effects of fMRI-NF training in alcohol addiction (Linden et al., NCT02486900), anxiety in adolescents (Cohen-Kadosh et al., NCT02440451), autism spectrum disorder (Castelo-Branco et al., NCT02440451), eating disorders (Hallschmid et al., NCT02148770) and various registered trials in depression (Chapter 2.1.3).

In comparison to fMRI-NF studies with psychiatric applications, less work with neurological patients has been published. Two PoC studies focused on the rehabilitation of cognitive functions, including one study that aimed to improve memory in Alzheimer's disease (Hohenfeld, Christian Nellessen et al., 2017), and one study that trained higher visual areas with the aim to reduce visuospatial neglect in stroke patients (Robineau et al., 2017). With regards to motor rehabilitation, more PoC studies have been published, including one PoC study with Huntington's disease patients (Papoutsi et al., 2018), two small PoC studies with stroke patients (Liew et al., 2015; Sitaram et al., 2012) and two PoC studies with Parkinson's disease patients (Buyukturkoglu et al., 2013; Subramanian et al., 2011). The PoC studies in Parkinson's disease patients has informed the design of a more recently published registered RCT (Subramanian et al., 2016). Lastly, one fMRI-NF RCT in stroke is currently (as of January 2018) registered on clinicaltrials.gov (Cohen et al., NCT02089776).

2.1.3 fMRI-NF in Major Depressive Disorder

Major depressive disorder (MDD) is currently the most intensively studied psychiatric condition in the context of fMRI neurofeedback. Out of 14 psychiatric RCTs that registered on ClinicalTrials.gov as either completed or still active, five trials (~36%) focused on depression (the search term "fMRI neurofeedback" was used; studies with status terminated, withdrawn or unknown as of January 2018 were excluded). The higher spatial precision of fMRI-NF and better access to deep brain areas compared to EEG-NF (Thibault et al., 2016) comes in particular to fruition in the treatment of depression, where the technique can target neural mechanisms underlying widely accepted cognitive models of depression (Disner et al., 2011). Specifically, neuroimaging studies indicate that negative cognitive biases observed in patients can be explained in terms of an increased drive of subcortical emotion processing and attenuated top-down cognitive control. Published and ongoing PoC studies and trials employ self-regulation of emotions using mental imagery, recall of autobiographical memories, or cognitive control strategies during exposure to affective external stimuli. fMRI-NF research published to-date has focused on one of either two aspects of the cognitive model: One set of studies (Johnston et al., 2011; Linden et al., 2012; Young et al., 2017, 2014; Zotev et al., 2011) has focused on increasing positive thoughts by training participants in up-regulating brain areas using positive mental imagery and memories. For instance, pilot studies with healthy participants (Johnston et al., 2011; Zotev et al., 2011) showed that self-regulation of brain areas associated with positive images was feasible. Similar paradigms have been translated to PoC

studies with depressed patients (Linden et al., 2012; Young et al., 2017, 2014). Another set of studies (Brühl et al., 2014; Hamilton et al., 2016; Paret et al., 2014) has focused on modifying patients' attentional bias towards negative stimuli by training individuals in exerting cognitive control to overcome such bias when being exposed to triggering stimuli. Specifically, PoC studies in healthy participants successfully trained individuals in down-regulating the response in limbic areas such as the amygdala when participants were presented with aversive stimuli (Brühl et al., 2014; Paret et al., 2014). However, both studies lacked any additional outcome measure (e.g. psychometric measures) to test whether the down-regulation was related to changes in affective states. A comparable fMRI-NF protocol was used more recently in depressed patients who trained down-regulating the anterior cingulate cortex (Hamilton et al., 2016). The study reported a decrease of about 20% in ratings of negative affective stimuli for the group that received fMRI-NF, but not for the control group that received yoked feedback. Yet, a direct comparison between groups was not provided and thus no superiority shown. A more detailed review of to-date published fMRI-NF trials will be provided in the following paragraphs.

2.1.3.1 Comparison of trial designs

In the following paragraph I will shortly review the designs of fMRI-NF studies published until January 2018 that were conducted with depressed and stroke patients. Figure 2.1 describes the main design features of 4 published studies, including one RCT (Young et al., 2017; NCT02078610). For comparison, it also shows the key design features of an unpublished RCT that will be presented in detail in this thesis (NCT01544205), as well as details of an ongoing RCT (Young, NCT02709161). Besides these three RCTs, other previous studies in healthy participants and patients were not registered and did not randomize patients, and two studies (Hamilton et al., 2016; Young et al., 2014) blinded patients and assessors. Overall, data indicated that published studies were rather heterogeneous with regards to the design, patient population and outcome measures. For instance, the number of training sessions was overall relatively low and ranged between 1 and 4 training sessions. Trials were also rather heterogenous regarding their inclusion and exclusion criteria for patients, for example with regards to other ongoing treatment and co-morbidities. While the study by Linden and colleagues (Linden et al., 2012) only included patients that had received psychopharmacological treatment for at least six weeks, both studies by Young and colleagues (Young et al., 2017, 2014) reported only data from unmedicated patients. The choice of the

control group represents another major difference between trial designs: whereas (Linden et al., 2012) included a control group that performed similar mental imagery but did not receive brain scans, Young and colleagues (Young et al., 2017, 2014) trained a proportion of patients on a control target area (the intraparietal sulcus) with an instruction (positive autobiographical memories) that did not allow participants to up-regulate the target area. Another form of control was delivered by (Hamilton et al., 2016), who used *yoked-feedback*, which constitutes a replay of feedback that is being matched randomly from another participant. These different control conditions address different aspects of neurofeedback training. The approach used by Young and colleagues aimed to control for motivation levels and the belief that one is receiving veridical feedback. In contrast, the approach used by Hamilton and colleagues aimed to match groups for the amount of (rewarding) feedback that groups are provided with (Thibault et al., 2016). Patients in both trials did not report awareness of group allocation. However, neurofeedback data reported by Young et al. indicated that patients in the control group did not attain self-regulation to a similar extent as the main experimental group. Hence, groups were not matched for the rewarding experience (Thibault et al., 2016). Moreover, it is possible that patients in this control group experienced different levels of frustration because they did no achieve self-regulation during most training runs. This circumstance may have partly confounded results: depressed patients are highly susceptible to negative effects on mood caused by frustration because of they attend more to negative thoughts and experience (Disner et al., 2011; Gotlib et al., 2004). Taken together, these examples illustrate the challenge in choosing an adequate control condition for complex clinical interventions such as fMRI-NF, which will be discussed in more detail in Chapter 9.

Study / Registration	Paradigm	Control	Tx	Randomisation/ Blinding	Registration	#sessions
Linden et al., 2012	↑ of areas involved in positive emotions during MI of positive emotions (N=8)	MI group (N=8) without fMRI-NF	Meds (> 6 weeks), CBT not specified	No / No	No	4
Young et al., 2014	↑ of amygdala during positive AM (N=14)	Sham feedback from control region (N=7)	No Meds, other not specified	No / Yes (double-blind)	No	1 + TR
Hamilton et al., 2016	↓ of salience-network; mainly anterior insula (N=10)	Yoked feedback from fMRI-NF group (N=10)	Meds (?) and no Meds, CBT not specified	No / Yes (double-blind)	No	1
Young et al., 2017	↑ of amygdala during positive AM (N=19)	Sham feedback from control region (N=17)	No Meds, CBT not specified	Yes / Yes	NTC02079610	2 + FU
Mehler et al., in submission	↑ of areas involved in positive emotions during MI of positive emotions (N=16)	↑ of areas involved in scene processing during MI of scenes (N=16)	Medicated Patients (> 3 months), no CBT	Yes / Yes (single-blinded)	NCT01544205	5 + TR+ FU
PI: Young 2016-2019	↑ of amygdala during positive AM (N=19)	Sham feedback from control region (N=17)	In combination with CBT	Yes / Yes (double-blind)	NCT02709161	2

Legend: ↑ = Up-regulation of target area; ↓ = Down-regulation of target area; AB = Autobiographical Memories; MI = Mental Imagery; Tx = Treatment; Meds = Psychopharmacological treatment; CBT = Cognitive Behavioral Therapy; TR = transfer run; FU = Follow up clinical assessment.

Figure 2.1 Overview of fMRI-NF studies conducted in depressed patients as of January 2018.

2.1.3.2 Comparison of main outcomes

The choice of an adequate outcome measure that shows sufficient specificity and sensitivity represents another key aspect in designing a successful intervention. Previous studies of fMRI-NF in depression varied largely with regards to their main outcome measures (Table 2.4) and included the Profile of Mood States (POMS) (McNair, 1971), Hamilton Depression Rating Scale (HDRS) (Hamilton, 1960) and the Montgomery-Åsberg Depression Rating Scale (MADRS) (Montgomery and Asberg, 1979). Whereas the HDRS and MADRS are physician rated scales accepted to diagnose and stage clinical depression, the POMS merely is a psychometric, multiscale measure, whose subscales and total score show satisfactory discriminate validity compared to other validated scales (Nyenhuis et al., 1999). Hence, the level of generalisability of different scales varies between published studies. Further, the comparability of reported clinical outcomes between trials is likely limited.

Study	Main Outcome
Linden et al., 2012	HDRS-17: significant 40% decrease in main group; POMS: no superiority of main group (no effect size reported).
Young et al., 2014	POMS: Significant decrease in depression subscale for main group with very large Cohen's d = 3.74[$], no superiority compared to control group with a medium effect size of r = 0.3[$].
Hamilton et al., 2016	Ratings of affective stimuli: significant decrease of about 20% only in experimental group, no superiority compared to control group.
Young et al., 2017	MADRS after 2 sessions (primary*): significant 49% decrease in main experimental group, 8 % reduction in sham control group. Group difference significant with Cohen's d = 1.22[$]. HDRS-21 after 2 sessions: significant 46% decrease in main experimental group, 10 % reduction in sham control group. Group difference significant, no effect size reported.
Mehler et al., 2018	HDRS-17 after 5 sessions (primary*) – reported in Chapter 4.
Young (ongoing) 2016-2019 NCT02709161	MADRS after 2 sessions (primary*) – to be determined.

Table 2.4 Reported main outcome measures for published fMRI-NF trials conducted in depressed patients. HDRS-17/HDRS-21 = 17-item and 21-item version of the Hamilton Depression Rating Scale; POMS = Profile of Mood States; MADRS = Montgomery-Åsberg Depression Rating Scale. *) Studies that preregistered primary clinical outcome. [$]) Conversions to standardised effect sizes were performed based on reported t-values using formulas provided by (Lakens, 2013).

Linden and colleagues demonstrated in a first PoC study that patients suffering from mild to moderate depression could activate brain areas that were localised with positive affective visual

stimulation using positive mental imagery (Linden et al., 2012). However, the study did not include pure mental imagery or transfer runs in the fMRI protocol and thus could not test if providing of feedback increased activation levels. Moreover, as pointed out by the authors, voxel selected was adapted from run to run for some patients (Linden et al., 2012) and hence analyses were partly circular such that the same data was used for selection and selective analyses (Kriegeskorte et al., 2009). With regards to the clinical results, the authors found that only the group that completed fMRI-NF training (but not the control group that completed exclusively mental imagery) showed significant improvement in mood (measured on the HDRS-17 score) within four training sessions (Linden et al., 2012). Further, an interaction analysis demonstrated superiority of the fMRI-NF group, which remained after correcting for gender as a covariate. However, the analysis did not correct for baseline HDRS-17 scores, which may have biased results. The POMS data revealed an improvement in both groups, but this effect did not remain significant after correcting for baseline scores. Hence, these findings remained limited given the relatively small sample, lack of rater blinding, and the circumstance that the control group did not interact with a high-tech environment that may have induced unspecific clinical effects (Thibault et al., 2018b, 2016), requiring replication in larger RCTs (Linden et al., 2012; Linden, 2014).

The PoC study published by Young and colleagues included a double-blind design and also employed a positive mental imagery task (Young et al., 2014). Their main experimental group trained exclusively up-regulating of the left amygdala, while a control group trained up-regulating of the intraparietal sulcus as a control region. The authors reported that the amygdala was only up-regulated in the main group. Moreover, a transfer run after the training indicated that patients retained the capacity to up-regulate the left amygdala. With regards to clinical training effects, the study only provided very limited insight because patients' symptom severity was not assessed with a clinical scale (e.g. the HDRS or the MADRS). Instead, the authors only reported pre-post changes in mood using the POMS, which they only found for the experimental group for the depression subscale. Superiority over the control group, however, was not reported (Young et al., 2014). Linking the neuroimaging with clinical changes, the study reported only for the experimental group a negative relationship between up-regulation success and patient's scores on a subscale of the Toronto Alexithymia Scale (Bagby et al., 1994), which measures the inability to identify and describe one's own emotions verbally. A direct comparison to the control, however, was not performed and hence the specificity of the finding is questionable. Moreover, it should be noted that this correlation

analysis was not declared as exploratory. The authors reported that 4 subscales were administered before training together with two other scales related to the experience of emotions, totaling in at least 10 possible correlation analyses that could have been explored to test a similar hypothesis. Yet, no other correlations besides the significant correlation were reported and no correction for multiple comparison was carried out for any statistic. The reported results should thus be interpreted with caution. To conclude, (Young et al., 2014) showed that an fMRI-NF paradigm that was earlier successfully piloted in healthy participants (Zotev et al., 2011) could be translated to patients suffering from depression in a double-blind design, but it did not provide further insight into potential clinical effects.

A different approach of fMRI-NF was studied by Hamilton and colleagues, who trained patients to down-regulate the activity in areas of the salience-network while being exposed to images that evoke a negative affect (Hamilton et al., 2016). The authors compared a main experimental group to a control group that received yoked feedback and report that the PoC study was conducted double-blinded. However, the study only consisted of a single training session and did not assess clinical outcomes. The authors reported that only the main group showed reduced activity in training areas, and that only the main group also showed a reduction in affective ratings of images, but no formal demonstration of superiority was reported, and non-significant results were reported as being significant. Hence, also this study was affected by some questionable research practices. Taken together, the evidence from previous PoC with regards to the clinical efficacy of fMRI-NF was very limited, and reported evidence partly compromised due to poor design quality and research practices, highlighting the necessity for clinical trials (Linden, 2014; Sacchet and Gotlib, 2016).

Indeed, standards were remarkably improved in a recently reported first double-blind, placebo-controlled randomized clinical trial (RCT) that was conducted in unmedicated patients (Young et al., 2017). The training protocol as largely similar to earlier work by the same authors (Young et al., 2014; Zotev et al., 2011) and compared amygdala fMRI-NF to fMRI-NF of a control area (the intraparietal sulcus) using positive mental imagery instruction. Participants attended one weekly training session for two subsequent weeks, which included a baseline and a transfer scan (both without feedback) at the beginning and end of each session to assess pre-post differences in target region up-regulation and whether participants retained the capacity to up-regulate in the absence of feedback. The authors found only participants in the experimental group showed activation of the amygdala. This activation was retained during the transfer run

and exceeded activation during the pre-clinical baseline run. Data further indicated that patients in the control group did largely not activate the control region over the trial. Only for the last training and transfer run the authors reported significant differences between groups for both target regions. The authors found over 40% reduction in depressive symptoms in the intervention, and only 8-10% improvement in the placebo neurofeedback control group on two established scales to rate depression severity (Young et al., 2017). However, since patients in the control group did not attain self-regulation success for most training runs of the trial, groups were not matched for the reward experience. Therefore, it remains unclear whether this efficacy is specific to feedback from emotion-regulating regions, or whether the general experience of self-regulation during fMRI-NF may be therapeutic. To test for potential psychophysiological mechanisms underlying the remarkable clinical effects, the authors included an in-house developed autobiographical memory test that allowed assessing whether training success was related to performance changes in cued recalling of positive autobiographical memory. The authors modeled the relationship with a mediation analysis and found that neurofeedback success during the last transfer run large explained the relationship between gains in recalling positive autobiographical memories at follow-up. On a critical note, the model seemed flawed such that the measurement of the mediator variable (transfer run) preceded the measurement of the predictor (change memory performance) and response variable (change). Conceptually, a mediation analysis requires that the mediator variable should be measured at a time point that follows the predictor and precedes the response variable. Hence, the authors' approach reduced the analysis to a hierarchical (manually selected) stepwise regression model in these sense that adding neurofeedback success to the model improved the model fit. Hence, presumably exploited degrees of freedom and the lack of preregistration for this analysis (for a critique of these biasing factors see Chapter 1.2.1) reduce confidence in this analysis.

In conclusion, the published literature of fMRI-NF studies conducted in depressed patients provides evidence for technical feasibility, tolerable attrition rates and safety of the intervention. Moreover, first evidence for clinical efficacy has been provided, but more RCTs are required that allow replicating reported findings, explore alternative designs, and identify potential neural mechanisms.

2.1.4 fMRI-NF for motor rehabilitation in stroke

2.1.4.1 Available evidence from previous fMRI-NF studies in stroke

With regards to stroke, the available data is very limited with only two small published PoC studies (Liew et al., 2015; Sitaram et al., 2012). Sitaram and colleagues tested whether stroke patients can activate the ipsilesional ventral premotor cortex (vPMC). The authors used an action observation task, which consisted of a video sequence showing a person holding and moving a coffee mug. This approach allowed to target voxels in the vPMC in stroke patients who were compromised in executing movements. The study included two stroke patients with right-sided hemiparesis (i.e. left hemispheric stroke) with mild impairment, as indicated by 55 and 60 out of 66 points on the Fugl-Mayer (FM) scale (Sitaram et al., 2012), and four healthy participants for comparison. Participants used motor imagery to learn to increase activation in the vPMC. Moreover, the authors compared two forms of feedback presentation: a video replay similar to the one used during the localiser and where the picture frames grew proportionally to the activity, and a traditional thermometer display which indicated both activation and deactivation. Results suggested that the percent signal change was positive to a similar extent for both feedback forms and for healthy participants and patients alike. Moreover, the authors reported that PSC values increased over time, suggesting that stroke patients were able to learn up-regulating the region within a 3-day training period with daily fMRI-NF sessions. Noteworthy, the study also included a reference region in the white matter to control for nonspecific global effects, including head motion.

The fMRI-NF study by Liew and colleagues included four stroke patients who were severely impaired with an average FM score of about 20 and a range of 13 to 28 points (Liew et al., 2015). Patients were trained on two days for two hours each, the feedback signal was based on a functional connectivity measure between the perilesional M1 and the ipsilesional thalamus (measured by Pearson correlation). The authors reported that connectivity estimates increased mainly for more severely impaired stroke patients, arguing that these could benefit in particular. However, an alternative explanation for this seemingly surprising result is head motion: the study did not correct for spurious correlations induced by head motion and may thus have been confounded in their results, as documented for healthy participants (Power et al., 2014, 2012) and stroke patients (Siegel et al., 2017). Severe stroke patients likely show more head motion (Wylie et al., 2014), and are thus more likely to induce larger spurious correlations in connectivity estimates. Head motion may also explain why connectivity estimates towards the

end of a two-hour long training session were larger than at training beginning for most patients, except the most recovered patient (Liew et al., 2015). Taken together, the available data suggest that stroke survivors tolerate fMRI-NF scan sessions. Further, preliminary results suggest that stroke survivors can likely activate target regions and may also learn to increase activation over repeated sessions. However, adequate control for head motion remains a challenge. Hence, replication and extension of findings in larger samples is required to demonstrate that the technique is technically feasible and to warrant testing for clinical effects in future RCTs. Given the limited evidence from published fMRI-NF conducted with stroke patients, the following paragraph will provide a review of potential targets.

2.1.4.2 fMRI correlates of recovery and potential fMRI-NF targets

Neuroimaging studies provide insight into functional and structural changes occurring during different phases after stroke. Changes in activity of motor regions and motor networks have been related to motor outcomes, allowing to conceptualise preliminary network models of motor recovery after stroke. For instance, fMRI experiments of hand movements conducted in acute and chronic stroke patients with upper limb impairment identified bilateral activation patterns over motor and premotor cortices compared to rather lateralised activation patterns in healthy controls (Grefkes and Ward, 2014). Early work in chronic stroke patients that investigated the extent of contralesional recruitment of sensorimotor areas reported an inverse relation between residual motor function and recruitment of ipsi- and contralesional areas including M1 and dPMC (Ward et al., 2003). The study presented a set of inverse correlations between the strength of activation in different motor areas and a motor outcome measure across stroke survivors. However, this set of bivariate correlational analyses could not indicate to which extent bilateral activation was related to poor motor scores within individual patients. Addressing this question would require using a multivariate model. Further, given the large range of time past stroke (3 months to 6 years), it is not clear when lateralisation differences had built up. One longitudinal study provided further insight into the dynamics of neural correlates of outcome after stroke in severely to mildly impaired patients (Rehme et al., 2011b). The study reported that bilateral activation (hyperactivation) built up gradually during the first two weeks following ischemic stroke, and then decreased at later (3-6 months) follow-up measurements (Rehme et al., 2011a). The authors further reported that a decrease in bilateral activation had coincided with an increase of the group's mean motor performance, but the relationship between these variables was not directly tested. One main question that arises is

whether increased bilateral activation represents merely an early compensatory mechanism, or whether it may be indicative of the motor outcome and potential to recover.

Several studies indeed suggest that an increase in contralesional (i.e. ipsilateral) motor areas likely compensate for neural loss to some extent. For instance, a combined fMRI-TMS study found that paired pulses applied over the contralesional dorsal premotor cortex (dPMC) at rest correlated with a faciliatory state (measured by motor evoked potentials; Bestmann et al., 2010). This effect correlated with the level of motor impairment (measured by a composite clinical score). Further, applied TMS pulses over contralesional dPMC during a hand grip task resulted in greater modulation of ipsi- and contralesional motor areas (dPMC and SMA), and this modulation was also correlated with the degree of motor impairment. These results indicate that the neurophysiological and behavioural role of the contralesional dPMC varies with the state of the motor system and therefore depends on the level of motor impairment. The experiment corroborated and extended earlier reports which found that interfering TMS pulses over the contralesional/ipsilateral dPMC increased reaction times, and more so in more severely impaired patients, who also showed more bilateral cortical activations during a finger movement task (Johansen-Berg et al., 2002). Hence, these findings suggest that contralesional dPMC may compensate for neural loss in the contralateral hemisphere. Another approach to study the role of contralesional/ipsilateral dPMC in stroke recovery was taken by Lotze and colleagues. The authors applied repetitive TMS pulses over different motor areas, including the dPMC, while stroke patients who had nearly fully recovered from initially severe impairment were performing sequential finger movements (Lotze et al., 2006). Compared to an age-matched healthy control group, they found reduced motor performance after stimulation of contralesional dPMC, as well as M1 and the superior parietal lobule, indicating that these contralesional regions may play a role in the recovery process. Taken together, several cross-sectional perturbation studies using TMS suggested that contralesional cortices including dPMC and M1 facilitate residual motor function, in particular, but not exclusively for more severely impaired patients.

In an elegant series of TMS induced virtual lesion experiments, O'Shea and colleagues explored functional reorganisation of dPMC following disruption of neural communication (O'Shea et al., 2007). The authors applied TMS pulses over the dominant dPMC of healthy participants while these performed an action selection task and found immediate compensatory increases in the contralateral dPMC. This activation was only observed during the action

preparation, but not action execution phase in the absence of a significant change in task performance. This result indicated that the contralateral dPMC compensated for dominant dPMC that had been perturbed with TMS. The authors replicated and extended these findings in a second experiment, in which a second TMS pulse was applied to the contralateral dPMC during action selection. As expected, this second perturbation interfered with task performance, supporting the interpretation that the "contralesional" dPMC can flexibly compensate when the other hemisphere is being perturbed. More generally, these data suggest that the brain can counteract lesions by recruiting homologue or other interconnected areas in a way that can maintain task performance.

Apart from perturbation studies, meta-analyses reviewing fMRI correlates of motor impairment provide further insights. For instance, one meta-analysis that computed activation likelihood estimation maps across 36 studies (including 472 patients with upper limb impairment) that had studied various motor tasks in stroke survivors reported that the activation of contralesional dPMC (and the bilateral SMA) was anticorrelated with time post-stroke when controlling for impairment. Further, activation in the contralesional dPMC (and bilateral pre-SMA) was anticorrelated with motor impairment when controlling for time post-stroke (Rehme et al., 2012). Thus, both time post stroke and motor impairment correlate independently with less activation in contralateral dPMC and pre-SMA/SMA. In contrast, no cluster in contralateral dPMC was reported for motor impairment when controlling for task complexity. Although interpretations based on null findings are limited if no evidence for the absence of an effect is being provided (Chapter 3.4.1.5), these contrasts together indicate that task complexity may mediate the involvement of the contralesional (ipsilateral) dPMC during motor execution. Preliminary data from Riecker and colleagues illustrate this aspect (Riecker et al., 2010). The authors studied fMRI correlates of right-hand finger tapping at different speeds in right handed left-side chronic MCA stroke patients and age-matched healthy controls. As expected, increased tapping speeds resulted in a linear increase in activity of the left (contralateral and ipsilesional) SMA in both groups. However, only the patient group also showed a linear increase in the right (ipsilateral and contralesional) dPMC in response to increasing tapping speeds. In other words, the involvement of ipsilateral/contralesional motor cortices varies as a function of the challenges that the impaired motor system is confronted with (Riecker et al., 2010; see also Schaechter et al., 2008). Studies in healthy participants support this view: recruitment of the ipsilateral motor cortex has been reported during challenges resulting from neural perturbation (O'Shea et al., 2007) and behavioural complexity (Verstynen, 2004), as

well as skill acquisition learning (Waters et al., 2017). Moreover, recent fMRI work suggests that movement kinematics (e.g. endpoint variability) are likely processed in parietal areas across hemispheres (Haar et al., 2017b, 2017a). In the context of stroke and motor impairment, in particular neural correlates of movement variability are of interest because excessive (task relevant) movement variability represents one main symptom in stroke (Cirstea et al., 2003; Cirstea and Levin, 2000). One potential explanation why the motor system integrates information about movement variability across arms and hemispheres may be that such interhemispheric cooperation facilitates motor learning and allows compensating more flexibly for potential unilateral perturbations. In other words, bihemispheric monitoring of movement variability (and hence potential movement errors) may come at a higher computational cost for the motor system, it allows both hemispheres to learn and act based on experience sampled by either limb (Mehler and Reschechtko, 2018).

Longitudinal studies during stroke recovery that explored correlations in activity between intra- and interhemispheric regions (*functional connectivity* analyses) can provide additional insights (Rehme et al., 2011a). For instance, correlations in activity between the ipsilesional SMA and M1 change in the months following stroke, predicting motor function 3-6 months post stroke. Similar relations were reported for correlations between activity of ipsilesional SMA and M1, as well as dPMC and M1 within days after stroke and two weeks post stroke These data suggest that early synergistic reorganisation between premotor areas and M1 may facilitate restoring some motor function. Moreover, a positive relationship between structural integrity of ipsilesional intrahemispheric connections between motor cortices and motor function were reported from a cross-sectional study (Schulz et al., 2015). After correcting for potential confounding factors such as the lesion site and patients' age, the authors found positive relationships between motor function and a measure of tract integrity (fractional anisotropy) between ipsilesional M1 and the (ventral and dorsal) premotor cortices. The corticospinal tract (CST) originates from M1, as well as from premotor areas including the SMA, and the structural integrity of these tracks is a reliable predictor for residual motor function and treatment effects of upper limb therapy (Riley et al., 2011; Schulz et al., 2012). In fact, it has been suggested that the integrity of descending CST may be a better predictor for motor impairment in stroke compared to other measures such as the size of lesions (Park et al., 2016). For instance, residual structural integrity of CST was found to correlate with grip strength across motor and premotor cortices (Schulz et al., 2012). Further, damage to the CST after stroke correlates with contralesional compensatory cortical activation (Lotze et al., 2012;

Schaechter et al., 2008). Of interest, it has been suggested that the integrity of the CST mediates the relationship between tract integrity between premotor-motor (vPMC and M1) connections and residual motor function. The study used a relatively large sample (N = 53), and was thus was presumably sufficiently powered to detect such interaction (Schulz et al., 2017). Specifically, the authors reported the outcome of a multiple regression model that found a positive relationship between M1-vPMC integrity and residual upper limb motor score (measured as a combination of force grip and FM score). Importantly, the slope of this relationship gradually declined as a function of CST integrity, suggesting that M1-vPMC projections mainly mediate residual motor output in patients with more severe CST disruption (Schulz et al., 2017). Altogether, these studies suggest that dynamic inter- and intrahemispheric changes in function of motor cortices allow compensating for loss in motor function early after stroke, and that these processes largely depend on the structural integrity of CST descending motor fibres.

Another body of literature focuses on non-crossing corticofugal motor fibres that may facilitate residual motor function and can partly compensate for damage to the CST (Grefkes and Ward, 2014; Peters et al., 2018; Plow et al., 2015; Rüber et al., 2012). For instance, in preliminary data in chronic stroke patients the microstructure of bilateral tracts to the red nuclei were found to be correlated with motor function (Rüber et al., 2012). These findings were more recently corroborated in a lager study (N = 43) that investigated the relationship between microstructural integrity of tracts between M1/SMA and subcortical areas, including the red nucleus in chronic stroke patients (Peters et al., 2018). Moreover, the compensatory role of the red nucleus seems supported by a longitudinal study that measured tract integrity over 3 months following stroke (Takenobu et al., 2014). The authors reported an increased in structural density of the red nucleus, which correlated with a change in motor function on the FM. Albeit being limited by a relatively small sample (N = 10) and likely inflated estimates (by pooling within and between subject variance within the same correlation analysis), these preliminary data are in line with previous reports. Noteworthy, the SMA contributes substantially to corticofugal motor fibres (Peters et al., 2018), as also supported by work in primates (Palmer et al., 1981). Other work in primates suggests that the SMA likely sends direct projections to spinal motor neurons that innervate the hand (Dum and Strick, 2002).

Based on the reviewed literature, the following motor imagery-based fMRI-NF paradigms could promote motor recovery after stroke:

1) The Ipsilateral/contralesional dPMC is likely involved in compensatory motor function of stroke patients. This part of motor cortex could be trained in an up-regulation fMRI-NF paradigm.

2) fMRI-NF could be used to retrain lateralisation of cortical activation, for instance during motor imagery tasks (Chiew et al., 2012). However, such training bears the risk of interfering with compensatory mechanism that are sustained by the contralesional hemisphere, and that may facilitate natural recovery.

3) Bilateral SMA activation has been associated with better motor outcome (Favre et al., 2014; Rehme et al., 2012). Further, the SMA contributes fibres to the CST and corticofugal tracts, which can partly compensate for lost motor function. Patients could train up-regulating the SMA (Subramanian et al., 2016).

This review has informed the design for a preregistered PoC study in MCA stroke survivors, which is summarised in Chapter 7.

2.1.5 fMRI-NF training in PD

Most neurofeedback training paradigms that aim to improve motor symptoms in PD combine motor imagery training with a visual or acoustic feedback signal that is based on a neural target correlate. As such, motor imagery based neurofeedback training combines two therapeutic approaches that may yield improvements of motor symptoms for PD patients in their own right (Mirelman et al., 2013).

The first fMRI-NF study conducted in patients was a PoC study with early-stage PD patients (N = 5) who engaged in motor imagery based fMRI-NF training of the SMA for two scan sessions (Subramanian et al., 2011). A control group of PD patients (N = 5) also trained motor imagery whilst lying in the scanner and receiving yoked feedback from a matched patient in the experimental group. In contrast to other fMRI-NF studies that have employed yoked feedback (Hamilton et al., 2016; Neyedli et al., 2017; Ninaus et al., 2013; Rance et al., 2018; Thibault et al., 2016), patients in the control group were instructed to ignore the presented feedback. Hence, patient groups were mainly matched for a similar visual input, rather than sham controlled for similar levels of motivation or self-regulation experience. Both groups

were further matched for their PD severity and medication status. Apart from motor imagery training during scanning sessions, both groups practiced motor imagery strategies at home for a period of 2-6 months before they attended a second training session. Neuroimaging data indicated that the experimental group, but not the control group, successfully activated the SMA during motor imagery practice. However, no group comparison was performed to demonstrate that fMRI-NF was associated with larger signal change in the SMA compared to motor imagery only. The authors interpreted this result as an indication that the feedback presentation enabled patients in the experimental group to increase the activation level of the SMA. An alternative explanation why the experimental group but not the control group showed significant SMA activation compared to rest could be that only patients in the experimental group likely intended to control the feedback. Previous work suggests that motor imagery combined with yoked feedback and the belief to be under control of the feedback can yield SMA activation (Ninaus et al., 2013). From a clinical perspective, the main finding of this PoC study was that only participants in the experimental group showed significant improvements (37%) in motor symptoms relative to baseline on the MDS-UPDRS-Part III. Specifically, four out of five patients showed improvement in limb bradykinesia. This clinical finding was also supported by an independent motor task (finger tapping) that showed speed gains only for participants in the experimental group. Although these findings are promising, their generalisability may be limited given the small sample size and the lack of blinding of the assessor, which could have biased results, as well as of patients, which could have resulted in a placebo effect in the experimental group.

Based on this PoC study, a larger registered, single-blinded randomised controlled trial (RCT) has been conducted more recently (Subramanian et al., 2016). Using a two-armed design, the RCT compared a group that completed three sessions of fMRI-NF of the SMA (N = 13) to an active control group that completed three sessions of supervised physical exercise (N = 13). The physical exercise training consisted of a gaming console that provides visual and auditory feedback based on motor performance and has been shown to improve motor symptoms in PD patients (Barry et al., 2014). Hence, both groups were controlled for several factors that may modulate positive expectations and thus induce placebo responses, including the regular interaction with research personnel, regular visits to a research facility, and training in feedback-based training that involves high-technology components. Additionally, both groups were instructed to continue with either with unsupervised motor imagery training at home (fMRI-NF group), or with supervised physical exercise training (exercise group). This design

hence increased the potential efficacy of neurofeedback sessions by combining it with additional treatment that was matched between groups. Blinded clinical assessment indicated that the fMRI-NF improved on the MDS-UDPRS-MS by about 4.5 points. Although the experimental group showed a statistically significant improvement, a formal test for superiority over the active control group remained insignificant. Importantly, a decrease of 4.5 points on the MDS-UDPRS-MS lies within range of minimal clinically important change score (MCID) defined for PD (Horváth et al., 2015), and thus represents a clinically significant improvement. Moreover, this improvement is descriptively also larger than placebo effects of 1 to 3 points that have been observed for sham TMS stimulation in PD patients (Chou et al., 2015). The neuroimaging data further suggested that PD patients could activate the SMA when feedback was provided, as well as during a final transfer run when no feedback was provided. Since no baseline motor imagery data was acquired, the trial could not demonstrate the feedback from the SMA increased its activation during motor imagery. Further, the trial reported a correlation (at trend level) between neurofeedback success, quantified as the average activation of the SMA, and clinical improvement. This finding may indicate that the average up-regulation success of patients related to clinical improvements. Nevertheless, as noted earlier, the intention to control feedback may be sufficient to yield SMA activation (Ninaus et al., 2013), and thus this finding may also be explained by motivational factors. One way to control for potential confounding factors would be a design in which PD receive well designed yoked feedback from the SMA. However, previous work indicates that such design likely requires relatively large sample sizes to demonstrate statistically significant differences in SMA activation levels between groups. An alternative approach may be to employ a different measure of self-regulation success that better reflects self-regulation and partly controls for potential confounding cognitive factors such as a general intention to self-regulate. Taken together, first studies have provided preliminary evidence that training of the SMA is associated with clinically meaningful improvements of motor symptoms in early-stage PD patients. Yet, the relationships between measures of feedback success and clinical improvement and neural mechanism underlying clinical effects remain unclear. Future trials will likely benefit from designs that allow controlling for general cognitive effects associated with self-regulation training, thereby allowing to separate mere intention to control feedback from true feedback self-regulation. EEG signatures for motor imagery-based EEG-NF paradigms, will be reviewed in Chapter 8.

2.2 Research questions and aims

Based on the above reviewed literature, I define seven main research questions and aims that will be addressed in this thesis.

1) For fMRI-NF in the treatment of depression, clinical efficacy remains to be tested under controlled conditions that include blinding and randomisation of participants. Previous studies have trained patients in up-regulating the activity in areas that are involved in processing positive affect. The groups in these controlled paradigms, however, were not matched for the rewarding experience associated with successful fMRI-NF self-regulation. It thus remains to be tested whether self-regulation of areas involved in processing positive affect yields superior clinical effects compared to a control condition in which patients receive veridical feedback from areas that are not mainly involved in processing positive affect. Chapter 4 and 5 will address these questions, presenting clinical and neuroimaging data acquired in a registered single-blind RCT (NCT01544205) that tested if fMRI-NF up-regulation training of areas processing positive emotions (NFE) yields superior efficacy compared to up-regulation of a control region activated by scenes (NFS). Moderately to severely depressed medicated patients were randomly assigned to five sessions augmentation treatment of either NFE or NFS training, Clinical effects were evaluated at trial completion and a follow-up assessment several weeks after the last training session.

2) First open label and controlled studies of fMRI-NF training have reported clinical improvements for depression (Linden et al., 2012; Young et al., 2017), but underlying mechanisms of action remain to be identified. Cognitive models in depression postulate that the self-concept self-efficacy can mediate therapeutic effects. It has further been postulated that self-efficacy could be enhanced by the experience of successful regulation of one's own brain activity. Chapter 4 will thus explore whether successful self-regulation improved self-efficacy ratings in depressed patients, and whether changes in self-efficacy relate to clinical changes. Chapter 5 will explore for relationships between measures of self-regulation success changes in self-efficacy and depression severity, respectively.

3) One important design aspect in fMRI-NF training is the choice of a suitable target region. Besides potential therapeutic effects that may unfold from activating or deactivating a specific target region, its trainability should be established first. Chapter 6 summarises a comparative study of motor imagery-based fMRI-NF training of two target regions (SMA and M1) that are of particular interest for motor rehabilitation in stroke and Parkinson's disease. Healthy participants will train with graded neurofeedback, which allows to separate generic activation of training regions (e.g. resulting from the intention to control the feedback) from true self-regulation effects. Chapter 5 will thus introduce a framework that yields a new measure to test for self-regulation success. The findings of this study will inform the translation of a similar design in a PoC study with stroke patients (Chapter 7).

4) Chapter 7 will present a preregistered PoC study of motor imagery-based fMRI-NF with stroke patients. Specifically, middle cerebral artery (MCA) stroke survivors completed sessions of motor imagery based graded fMRI-NF training of the SMA. Based on the experience gained in the pilot study with participants in a similar paradigm (Chapter 6), *a priori* hypotheses were formulated, and a complete analysis plan anticipated. Together with a detailed documentation of the used methodology, this information has been preregistered and made publicly available on the Open Science Framework before data collection started: https://osf.io/qnsv7/

5) Besides fMRI-NF, EEG-NF remains a viable option for motor rehabilitation. For instance, EEG is more affordable, portable and available for health care settings compared to fMRI. For Parkinson's disease (PD), electrophysiological correlates of motor symptoms and treatment success have been intensively studied, providing a rich body of literature that can inform translational EEG-NF paradigms. Surprisingly, only a small number of motor rehabilitation EEG-NF studies have been conducted in PD patients (Esmail and Linden, 2014). Chapter 8 will review potential EEG signatures that could be translated as targets for EEG-NF training in PD patients. This review has partly informed the design of a registered EEG-NF PoC study that is currently being conducted in PD patients (ISRCTN16783092).

6) Replicability and reproducibility of scientific work are tightly linked to the way statistical tests are used and results are being reported. This thesis addresses various

issues that have suggested to enhance these pillars of robust science. Specifically, planned and exploratory hypotheses are explicitly declared as such and kept separate, following one key principle of preregistration. Further, recent developments of recommended statistical and reporting practices are employed throughout the thesis, including reporting of effect sizes and conducting statistical tests that allow providing testing for the absence of an effect (e.g. frequentist equivalence tests and Bayesian hypothesis testing). To comply with best practice in evidence based clinical research, the RCT conducted with depressed patients (Chapters 4 and 5) and the PoC study (Chapter 7) conducted with stroke survivors have been preregistered before data collected started.

7) The thesis will close with a general discussion (Chapter 9) that embeds the main findings of the experimental chapters 4 to 7 in a broader context of evidence-based translation of neuroimaging. The role of reliable control conditions, data quality, and sufficient statistical power will be reviewed. Providing convincing evidence for clinical usefulness of neurofeedback training may represent "a long winding road" (Arns et al., 2014). However, building an intersection between neuroimaging, computational neuroscience and translational clinical medicine, the field benefits from rich interdisciplinary collaboration. Neurofeedback researchers can thus make use of various tools that help overcoming these challenges and possibly establish the technique as a valid treatment alternative for certain neurological and psychiatric conditions. Moreover, neurofeedback research is a useful test dummy for translating neuroimaging from the bench to the bedside, possibly creating "spill over effects". The multidisciplinary field may drive innovations that can be adapted and adopted by researchers in other areas of translational neuroscience.

3 METHODS

3.1 Basic principles of Electroencephalography (EEG)

Electroencephalography (EEG) was invented by German psychiatrist Hans Berger in the 1920s as the first device that allowed recording brain activity non-invasively. EEG uses scalp electrodes which measure local field potentials (LFPs), which represent the summed activity of local neural populations. LFPs reflect the electric potential in the extracellular space, which is largely determined by post-synaptic activity. When neurons generate electrical currents, negatively and positively charged ions are exchanged with the extracellular space along the neuron. The direction of ion flow is determined by the electric potential that is generated at the post-synapse: Excitatory post synaptic potentials (EPSP) entail an inflow of positively charged ions, resulting in a depolarisation of the post-synaptic neuron. In contrast, inhibitory post synaptic potentials (IPSP) entail an inflow of negatively charged ions, resulting in a hyperpolarisation of the post-synaptic neuron. These local ion currents change concentrations of ions in the extracellular space, creating an electric potential along the neuron. For instance, in the case of an EPSP, negatively charged ions (e.g. chloride) dominate and create negative polarity (a sink). In contrast, further along the neuron passive ion movements follow this gradient, resulting in positive polarity (a source). Together, sink and source form a bipolar potential. The sum of these bipolar potentials results in local field potentials (LFPs), which form the basis for the neural signal that can be detected with electrodes placed on the scalp (Lopes da Silva, 2010). In contrast to LFP recorded directly in the extracellular space of brain tissue, the signal detected by EEG has first propagated through various media that lie between the electrode and the extracellular space, including the cerebrospinal fluid, the cranium and muscle.

Electrical activity measured with EEG can be characterised by their frequency (measured in hertz, Hz) and amplitude (measured in microvolt, µV). Raw EEG data consists of a mixture of various frequency bands, however, through digital filtering (e.g. a fast Fourier transformation) data is transformed from the time domain into the frequency domain. Additional band pass filtering allows separating a frequency spectrogram into discrete frequency bands. Some main types of frequency bands include the alpha-band (8-12Hz), beta-band (13-30Hz), and low gamma-band (30-70Hz) and high gamma-band (70-300Hz). Alpha, beta and lower-gamma oscillations play a critical role in motor control (Lopes da Silva, 2010). For instance, one predominating view in the motor control literature is that gamma oscillations encode feedforward information from the sensory signals (predictions) whilst beta oscillations encode

feedback information (prediction errors) modulating feedforward signals based on the current context (Bastos et al., 2015; Tan et al., 2016). Changes in these frequency bands have been identified in response to treatment for Parkinson's disease (PD) and will be discussed in Chapter 8 with regards to their potential as target frequencies for an EEG-NF intervention in early-stage PD patients. Other slower frequency bands include the delta (0.5-4Hz) and theta rhythms (4-8Hz).

The strength of oscillations within a certain frequency band is measured in amplitude (μV) and reflects the relative level of neural synchrony at a given frequency range. When measured over time, this allows to quantify EEG changes relative to baseline periods (e.g. as percent signal change) in response to specific events. Relative to baseline, a given task may either increase (event-related synchronisation; ERS) or decrease the synchrony (event-related desynchronisation; ERD) of cortical neural populations. ERDs and ERSs are of interest for our understanding of motor control, pathological processes, and therapeutic responses (Chapter 8). Further, they provide potential EEG markers that can be targeted with neurofeedback training. For instance, for motor imagery event-related, lateralized changes can be expected in the alpha and beta band over central electrode (Pfurtscheller et al., 2005): for left hand motor imagery, typically ERD occurs in the alpha and beta band over the contralateral (i.e. right) hemisphere, while ERS occurs for both frequency bands over the ipsilateral (i.e. left) hemisphere. Besides motor imagery, also motor execution shows similar patterns of ERD and ERS, indicating that they are likely involved in the preparation, execution, but also in the evaluation of a movement outcome (Tan et al., 2016). ERDs in the alpha band have been successfully used for motor imagery based EEG-NG training (Ring et al., 2015), such that healthy participants learned to voluntary control the ERD. Moreover, in particular lateralization patterns are used in BCI experiments that use classification algorithm to decode the side and limb with which motor imagery is being performed (Morash et al., 2008).

3.2 Basic principles of functional MRI

The results and discussions presented in this thesis were mainly informed by data stemming from functional magnetic resonance imaging (fMRI) measurements. Specifically, the image contrast used for the outlined real-time fMRI neurofeedback studies is based on the blood oxygenation level dependent signal (BOLD), which is inversely proportional to the local level of deoxyhaemoglobin (desoxyHB; Shmuel, 2010). Hence, BOLD only represents an indirect

measure of neural activity. The neural basis of the BOLD signal lies in LFPs, which constitute the activity in local neural populations (Logothetis, 2002). However, LFP amplitudes depend on various factors, including the strength of synaptic input, the transmembrane potential and the synchrony of current. Further, the functional anatomy of the cortex including the cytoarchitecture modulates local LFPs and thus the BOLD signal. Whereas large Purkinje cells in the cerebellum induce a large BOLD response, small inhibitory neurons do not contribute to the LFP (Lauritzen, 2005). Hence, the neurophysiological correlate of the BOLD signal mainly represents the local input and processing of information, but it does not inform directly about the output signal that is sent to other regions. Of interest, the relationship between BOLD and LFPs is particularly pronounced for certain frequencies, suggesting that these mainly contribute to the measured signal. These include alpha (8–12 Hz), beta (18–30 Hz), and gamma (40–100 Hz) LFP bands (Magri et al., 2012).

Neurophysiological processes are not the only correlate of the BOLD. The BOLD contrast is based on the idea to exploit an endogenous contrast that has originally evolved to regulate blood flow. One mechanism that contributes to blood flow regulation is autoregulation, which counteracts fluctuations of arterial blood pressure. Autoregulation is limited to only a few organs, including the brain and the kidneys. Yet, changes in neuronal activity require changes in blood flow, a process that is regulated via a second mechanism: neurovascular coupling.

Neural processes and vascular responses that regulate the supply of oxygenated blood are linked through neurovascular coupling (Iadecola, 2004). When local metabolic demands increase due to increased firing rates, different vasoactive mediators including ions (e.g. potassium), neurotransmitters (e.g. gaba-aminobutyric acid (GABA)) and neuromodulators (e.g. dopamine) are released. These vasomodulators may merely constitute by-products of metabolic processes (e.g. carbon dioxide), or they may play a key role in the neural processing itself (e.g. dopamine and GABA). Also, interneurons that project neurites directly to blood vessels release vasoactive factors when neural activity increases. Vasoactive modulators regulate local arterial cerebral blood flow (CBF), and thereby the concentration of desoxyHB decreases in the local vessels, resulting in a positive BOLD signal compared to rest (Shmuel, 2010).

Noteworthy, temporal delays in neurovascular coupling render the BOLD response a temporally low-pass-filtered version of the underlying neurophysiological response that is

delayed by 1.5 – 2s (Ashby, 2011; Shmuel, 2010). Moreover, the peak blood flow and hence the BOLD plateau are reached in healthy individuals after 4-6s. These delays, however, may vary between brain regions and also change with aging and pathological processes including stroke (Lin et al., 2011; Schroeter et al., 2007).

The brief outline of the underlying neurophysiology also provides a better understanding of some limitations of BOLD contrast, which shall be shortly summarised. The signal arises from a multitude of vasoactive factors that may relate to different underlying physiological processes including changes in the metabolic rate due to changes in firing rates, the computational architecture including inhibitory and excitatory processes realised via neuromodulators such as GABA, and other context dependent aspects including memory processes and motivational aspects. The BOLD signal thus contains information of brain function dependent processing, but also neuromodulators and physiological processes. Therefore, comparisons between magnitudes of the BOLD contrast, e.g. between two brain regions or between two tasks within the same brain region, require cautious interpretation (Logothetis, 2008).

3.3 fMRI Data pre-processing and analysis

3.3.1 fMRI processing

The main goal of fMRI pre-processing is to minimise the impact of physical and physiological artefacts and to validate model assumptions of parametric statistics used for fMRI analysis (Lindquist, 2008), which will be discussed in the next paragraph (3.3.2). This paragraph serves as a brief overview of some predominant pre-processing steps that are relevant for the univariate analyses of fMRI data that will be presented in this thesis.

Brain volumes are acquired in slices, often with an inter slice time of 70 milliseconds or more. In contrast, statistical analyses assume that data in all voxels was acquired simultaneously. With a repetition time (TR) of 2 seconds for instance, the true temporal difference between slices may be, however, up to two seconds. Moreover, within this time the participant may have moved their head. fMRI analyses assume that each data point in a time series from a given voxel only consists of data from a specific voxel. Hence, two important early pre-processing steps in fMRI data analysis are slice timing correction and head motion correction (Ashby, 2011; Lindquist, 2008). During slice timing correction, time courses are shifted with respect to an indicator. Head motion represents one main source for variation in fMRI time series

(Lindquist, 2008), which varies between contextual factors such as the task and certain clinical populations such as stroke (Seto et al., 2001), or age groups (Yuan et al., 2009). When uncorrected, head motion violates model assumptions because signal from neighbouring voxels "contaminates" the signal from a specific voxel of interest (Lindquist, 2008). At the beginning of motion correction procedures, a refence image is determined to which other volumes are registered and corrected, e.g. based on contrast differences at edges of the image. This reference image may be the first volume of a scan, such that all following volumes and scans within a scan session are aligned to this image ("intrasession alignment"). Different approaches exist to solve this optimisation problem, which vary between fMRI analysis software (Oakes et al., 2005) and can be characterised by the interpolation method (e.g. Fourier or trilinear), the optimisation technique (e.g. iterative gradient descent), and the cost function used to avoid overfitting and control optimisation speed (e.g. weighted lest squares). One predominant technique of motion correction is a rigid body transformation, which involves 3 translation variables and 3 rotation variables (along the x, y, and z axis, respectively). During the process, a cost function is minimised to achieve optimal parameter estimates for displacement along these 6 dimensions between a given image and a reference image. Transformation parameters can further be used as nuisance regressors in subsequent statistical analyses to account for spurious variations in the BOLD signal that have occurred due to head motion. Once motion correction is completed, estimated displacement parameters are used to interpolate a new image which is assumed/treated as "motion free" for the parametric analysis (Lindquist, 2008).

The spatial resolution of BOLD fMRI is limited by the architecture of the brain's vascular system. Hence, compared to a structural anatomical acquired at the same field strength, the BOLD image will feature less spatial acuity. However, spatial acuity in the evaluation of functional images can be increased by co-registering functional to structural data. Specifically, the co-registration of functional and anatomical data allows to map functional activity onto a high-resolution anatomical image. Moreover, a transformation to a standard space such as Talairach (Talairach and Tournoux, 1988) or MNI (Evans et al., 1992) allows comparing coordinates across participants and studies. Transformation to a standard space hence makes it possible to pool data across participants and perform group analyses, which greatly improves statistical power (Desmond and Glover, 2002). However, in contrast to image alignment within one acquisition mode (e.g. functional BOLD imaging), co-registration across acquisition modes (e.g. functional BOLD imaging and structural T1 weighted imaging) entails a higher degree of interindividual anatomical variations, requiring non-linear transformations (for

instance trilinear or sinc interpolation) as opposed to mere rigid-body transformations (Ashby, 2011; Lindquist, 2008).

When working with group data and *mass univariate statistics* (Chapter 3.3.2), spatial smoothing is another commonly practiced pre-processing step. Spatial smoothing is a form of data filtering such that the intensity values of voxels become more similar to neighbouring voxels. Practically, smoothing kernel is a function that is convolved with image, yielding voxel intensities that represent a weighted average of neighbouring voxel intensities (Lindquist, 2008). Smoothing reduces between subject variations in anatomical and functional variations, which may interfere with the co-registration process. Spatial smoothing can increase the signal-to-noise ratio because it reduces the effect of random voxel fluctuations due to noise. Moreover, smoothing reduces the effect of outliers on the data and thereby contributes to a normal distribution of residuals (Ashby, 2011). This latter aspect increases the validity of inferential parametric statistical tests (such as a t-test) that are commonly applied in imaging data. The size of smoothing kernels (measured in mm) is dependent on factors such as regions of interest as well as the task at hand (Ashby, 2011). One disadvantage of spatial smoothing, however, is that the spatial resolution is compromised. It may hence reduce the power for other sets of statistical techniques such as multivariate pattern analysis (Hendriks et al., 2017). Taken together, the application of smoothing kernels should be informed by the research question.

3.3.2 fMRI analysis

General Liner Models (GLMs) represent the predominant statistical approach in analysing univariate neuroimaging data (Ashby, 2011). GLMs are statistical models that can be solved with ordinary least square methods. In principle, GLMs as used in most block design fMRI experiments attempt to model a measured time series (Y) of every voxel using a given set of predictors and an assumed shape of the hemodynamic response function (HRF) that is convolved with (and hence scaled by) predictor variables. Predictors may represent events, for instance the onset and offset of a task ("box-car" function"), or confounding variables such as head motion parameters as well as other factors that may impact the BOLD signal. If available, physiological signals such as respiration or end-tidal carbon dioxide ($P_{ET}CO_2$) can for instance be measured and included as predictor variables once it has been pre-processed accordingly (Murphy et al., 2013). All predictors are entered into a matrix (X), for which the number of

predictors determines the number of columns, and the number of time points determines the number of rows:

Eq. (3.1): $Y = XB + e,$

It is hence another way to represent a multiple linear regression model in the form of

Eq. (3.2): $Y = \beta_0 + X_1\beta_1 \ \ldots \ X_n\beta_n + \varepsilon$

, β_0 is the estimated constant (baseline), the numerated betas correspond to the estimated beta regression weights of the entered predictors, and ε represents the residual error. Predictors are entered into the GLM and regression weights (i.e. β_n) are estimated using a closed form ordinary least square approximation. Through rearrangement of the equation, it can be solved with ordinary least square methods (Friston et al., 1994). One main assumption that parametric methods make is that residuals (errors) are normally distributed. However, due to the nature of the BOLD response, which is related to cerebral blood flow and other physiological noise such as heart rate and respiration but also other factors, periodic fluctuations in the BOLD signal that have not been modelled in the GLM result in autocorrelated residuals. One way to tackle temporal autocorrelation is by estimating the temporal dependencies, readjust the time course and re-iterate the model estimation (Bullmore et al., 1996). In approaches that study the representational geometry of functional imaging data (e.g. representational similarly analysis), one can also normalise for spatial dependencies using spatial "pre-whitening" (Ejaz et al., 2015). Lastly, multivariate techniques, so called model-free techniques such as independent component analysis (ICA), represent an alternative to model-based correction procedures. ICA allows dissociating the BOLD signal into additive subcomponents (Beckmann and Smith, 2004) such that signal components related to noise can be identified and discarded in a semi-automatised way.

With regards to statistical analysis of model-based *mass univariate statistics* fMRI data, appropriate procedures for multiple comparisons need to be considered to allow valid inferences and minimise false positives (see Chapter 3.4.1.3 for more details). Specifically, contrasting of beta weights between conditions (e.g. in form of a t-test) is performed on a voxel-by-voxel basis and hence requires appropriate correction methods. Frequentist statistics assume that no effect is present, i.e. that the null hypothesis is true (Chapter 3.4.1.1). The null

hypothesis can only be rejected if the data is more surprising than it could be expected with a certain probability if there was no effect, which is commonly called a significant effect. This probability is represented as the alpha level, an error rate that is considered tolerable and that is usually defined as 5% (i.e. $p < 0.05$) in the life sciences. However, this error rate is set per test and as the number of tests increase, the overall error rate increases proportionally (see Chapter 3.4.1.3). To maintain the validity of statistical inferences, multiple comparison correction is required, in particular for fMRI data which comprises many thousand voxels and hence often many thousand (univariate) tests. Four main approaches of multiple testing correction shall be shortly reviewed here (Ashby, 2011): 1) family-wise-error-rate (FWER) correction aims to retain the α-level at 5% across experiments and represents one of the most conservative approaches. 2) False discovery rate (FDR) based approaches aim to limit the number of false positives per experiment. It thus aims to control the α-level per experiment. 3) Another set of techniques are cluster wise inference approaches, which instead of correcting at voxel level, estimate the minimum number of adjacent voxels one would expect to find under the null hypothesis. Cluster extent-based correction techniques use information about the spatial smoothness of the data and require users to set two thresholds: a cluster defined threshold (CDT) that retains only voxels whose p-value is below it (e.g. CDT $p < 0.001$), and a cluster-level extent threshold (e.g. $p < 0.05$), which represents the error rate of falsely retained clusters (assuming that an appropriate CDT was applied, and all other model assumptions are met). Parametric cluster wise inference methods are widely used to correct for multiple testing and reliable in controlling for the type-I error rate when appropriate thresholds are chosen (Eklund et al., 2016). 4) Another set of techniques are non-parametric permutation tests. In contrast to parametric approaches that *assume* the distribution of the statistic of interest under the null hypothesis, non-parametric permutation based procedures can *find* the statistic under the null hypothesis by permuting the labels and reiterating analyses (Smith and Nichols, 2009; Winkler et al., 2014). Both techniques have been shown to reliably control FWER around the nominal 5% level, although permutation tests show slightly better performance (Eklund et al., 2016), and cluster inference based methods can control reliably for FDR nominal 5% level (Kessler et al., 2017). One advantage of cluster wise inference over permutation based methods is their higher sensitivity in detecting effects (Woo et al., 2014). This sensitivity may come at the cost of inflated false positives if too liberal CDTs are chosen (Eklund et al., 2016). However, if used with appropriate primary thresholds (CDT $p < 0.001$), they represent a reliable and economic way to control for multiple testing and hence a justifiable trade-off

between statistical rigorousness and sensitivity. For this reason, they have been the primary method that was applied to inferential whole brain analyses reported in this thesis.

3.3.3 Real-time fMRI setup and processing

Most real-time fMRI setups consist of three main elements that require computing infrastructure: an acquisition element, a real-time analysis and processing element, and a feedback presentation element (Figure 3.1). In many current setups, these three elements are processed by at least two dedicated computing units. The real time acquisition assures timely export of data and performs real-time image reconstruction. Reconstructed images are transmitted (often via a Transmission Control Protocol/Internet Protocol [TCP/IP] network connection) to the analysis element, a computer system that is equipped with fMRI analysis software capable of performing real-time analyses. The experiments presented in this thesis were exclusively analysed with the commercial software package Turbo-BrainVoyager (Brain innovation, Maastricht, The Netherlands). However, it should be noted that freely available alternatives have been more recently developed as well (Koush et al., 2017; Sato et al., 2013). Similar to conventional (offline) fMRI analyses, real-time fMRI analyses include pre-processing steps. In particular realignment to reference volumes to correct for head motion is critical to ensure that training can be performed on selected voxels. Also, spatial smoothing is usually performed to increase the signal to noise ratio. When feedback is provided based on model estimates obtained from an incremental GLM, head motion regressors can be included to account for spurious correlations caused by the head motion. Also linear drift term can be included to account for scanner drifts, e.g. due to thermic changes and instabilities (Lindquist, 2008). These result in relatively slow confounding changes of voxels intensities which can be modelled within the incremental GLM. Given that real-time analyses are usually constrained to regions of interest (but also limited with regards to the temporal signal to noise ratio), real-time fMRI experiments usually do not control for multiple comparison correction during online analyses.

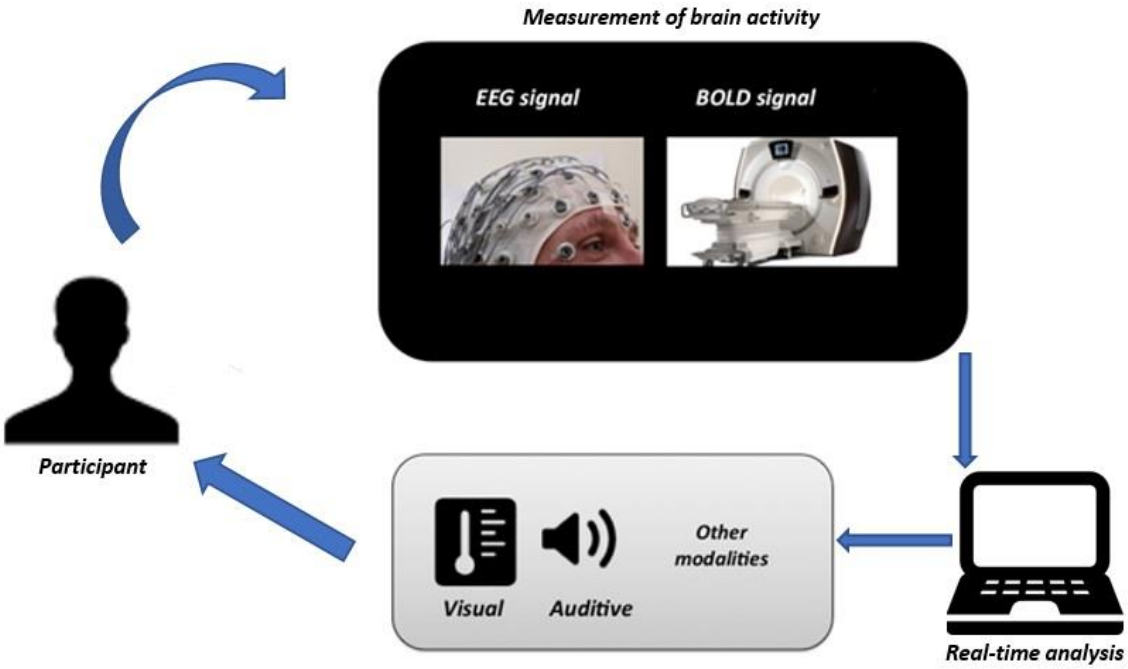

Figure 3.1 Schematic real-time neurofeedback setup. Modified from (Arns et al., 2017).

3.4 Statistical approaches to hypothesis testing

Statistics play a central role in the interpretation and presentation of data (Chapter 1.1). In this thesis statistical approaches more recently introduced to neuroscience will be applied, including frequentist equivalence tests and Bayesian statistics. Hence the rationale of classical frequentist hypothesis testing, potential pitfalls, and basics of Bayesian hypothesis testing will be shortly reviewed. The goal of this section is to provide the reader with an intuition of these techniques and refer to relevant literature for a more detailed discussion.

3.4.1 Frequentist hypothesis testing

3.4.1.1 Principles of frequentist hypothesis testing

Two general forms of statistics can be distinguished: descriptive statistics are concerned with quantifying the shape of data points and provide measures of centrality (e.g. mean), variability (e.g. variance) and skewness. Frequentist inferential statistics test in how far a given distribution of data deviates from a distribution when no effect is present (i.e. the null distribution). Frequentist statistics provide test statistics that allow to test how probable the empirical data is given that the null hypothesis is true. Specifically, they test for the probability of obtaining a result as extreme (or more extreme) as the empirical result, given that the null hypothesis is true. This probability is expressed as a probability value, or *p-value*. It thus

implies the notion of uncertainty. In frequentist statistics, it is assumed that measurements underlie various forms of error, sampling error: data collection constitutes sampling data points from a population. Even if there was on average no effect on a measure of interest in the population that one is sampling from, it is possible that one is mainly sampling data points from extreme data points of the null distribution, resulting in a false positive effect. However, in theory repeated sampling would converge eventually over time on the true population average. This thought experiment describes the rationale behind *frequentist* statistics, which treat probabilities as long run relative *frequencies*. On the level of the individual experiment or test, potential errors are accounted for by a pre-defined error rate that is deemed tolerable, the α-level (or *type-I error*). If the empirical p-value is smaller than the prespecified α-level, the test statistic deviates *significantly* from test statistics that can be expected under the null hypothesis, and the null hypothesis is rejected. This principle shall be shortly illustrated with an example in which one wants to test whether two group means differ. To test for a group difference on some measure x, using a two-sample t-test and assuming similar variances between groups, a t-statistic can be obtained,

Eq. (3.3): $$T = \frac{\bar{X}_1 - \bar{X}_2}{s_p \cdot \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

, where $\bar{X}$ represents the means of two independent samples, S_p their pooled standard deviation and n their respective sample size. This t-statistic hence represents a ratio between measures of centrality and variability that are weighted by the respective sample sizes. Given the degrees of freedom (df; in this example $df = n_1 + n_2 - 2$), the obtained t-statistic can be expected with a certain probability under the null hypothesis, yielding a p-value. As noted earlier, the p-value indicates the probability of obtaining a result as extreme (or more extreme) as the observed one, given that the null hypothesis is true. It does not, however, indicate the probability that the null hypothesis was true, which represents one common misconception that occur in the interpretation of frequentist statistical results (Greenland et al., 2016). To express such probability, it is required to know the probability of the alternative hypothesis, which is impossible to derive using a frequentist approach because empirical distributions are always tested for their comparability with the null distribution. This circumstance indicates one main limitation behind frequentist null hypothesis testing, such that the p-value is biased against the null hypothesis because it does not allow to quantify evidence for the absence of an effect (see Chapters 3.4.1.4., 3.4.1.5, and 3.4.2). Another limitation of p-values is that they are not

indicative of the size of an effect (Greenland et al., 2016). As shown in Eq. 3.3, for a given sample size, a large t-value can be obtained when the difference between group means is large, or when the standard deviation is relatively small. Moreover, with increasing sample sizes p-values will decrease. Effect sizes, on the other hand represent this information, and it is thus recommended to report these together with their confidence interval alongside inferential statistical results (Ioannidis, 2018). Confidence intervals estimate the range of values for a point estimate (e.g. a mean or effect size) that can be expected for a given percentage (e.g. 95%) of experiments if it was repeated many times, representing its long run frequency distribution.

3.4.1.2 Statistical Power

Besides the type-I error rate (α), another concept central to frequentist statistics is statistical power. Statistical power describes the probability of a statistical test to detect an effect of interest, representing the inverse of the type-II error rate (β) of missing to detect an existing effect. Statistical power depends on the effect size one is interested in (e.g. a Cohen's d of 0.5), the specific statistical test (e.g. a two-sample t-test), and the error rate (e.g. an α -level $p = 0.05$). Hence, when researchers define an effect that they hope to detect with an experiment, the probabilities α and β allow to calculate the required sample size (Figure 3.2).

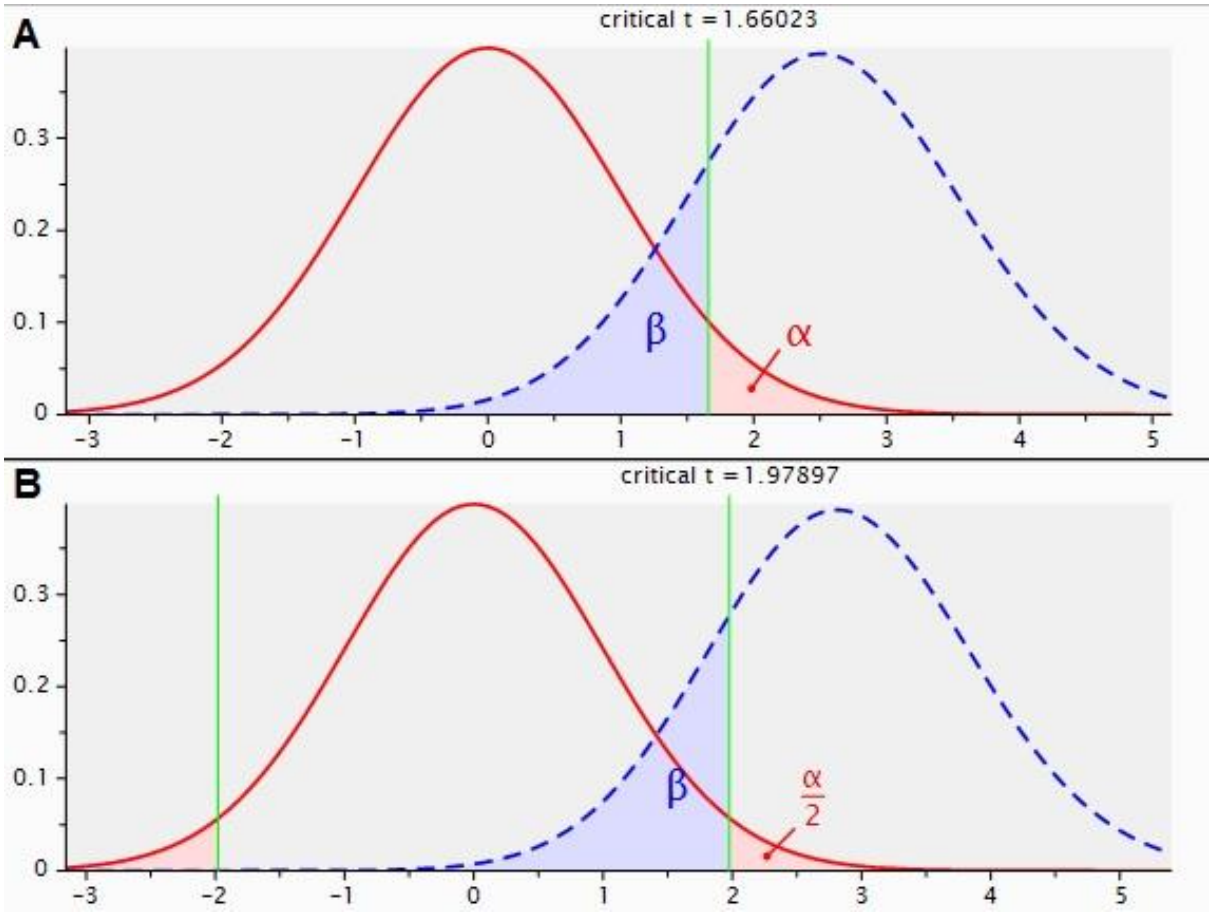

Figure 3.2 Example power calculation for a one-sided (A) and two-sided (B) independent t-test conducted in G*power software (Faul et al., 2007). Calculation was based the following input parameters: an effect size of Cohen's d = 0.5, a desired power of 0.8 (β = 0.2), an α-level of 0.05, and an allocation ratio N2/N1 = 1. The null distribution is shown in red, the simulated empirical distribution corresponding to the pre-specified input parameters is shown in blue, the critical t-values to reject the null hypothesis are indicated by green vertical lines.

A priori calculated statistical power gives researchers an indication of the required sample size to detect an effect of interest with a certain probability. Moreover, if studies are designed in way that they can detect smallest effect sizes of interest (SESOI), researchers increase the probability to reject the null hypothesis, as well as a range of SESOIs if the null hypothesis test remains inconclusive (see Chapter 3.4.1.5). In contrast, *a posteori* (i.e. post-hoc, empirical) power calculations result from a mere transformation of the null-hypothesis p-value. Post-hoc power estimates thus represent a frequency probability that is contingent upon the individual design, including the effect sizes used for the original power calculation that may have determined the acquired sample size, rather than the effect one is interested in (Hoenig and

Heisey, 2001). For this reason, but also factors such as researcher bias and publication bias (Algermissen and Mehler, 2018), post-hoc power calculations do not provide a reliable estimate to plan prospective *a priori* power calculations (see also Chapter 1.1).

3.4.1.3 Multiple testing correction

To maintain valid inferences, frequentist statistics rely on maintaining the α-level at an acceptable rate (e.g. 5%) in order to control type-I errors (i.e. falsely rejected null hypotheses). When multiple tests are carried out, the chance to commit a type-I error increases with the number of tests. Hence, techniques that correct for multiple testing either adjust the obtained test statistics (i.e. p-values), or the α-level and thereby control for the rate of false positives. One prominent way to correct for multiple testing is the Bonferroni method, for which the critical α-level is adjusted by the number tests that are being conducted. However, this approach assumes that the hypotheses and their associated tests are independent. In the case of (partly) correlated hypotheses, for instance when two similar hypotheses are tested using two correlated items from a questionnaire, this assumption does not hold, and Bonferroni correction is too conservative. One alternative way is to correct for the "False Discovery Rate" (FDR), which adjusts obtained p-values to control for the proportion of potential false positives among all tests that reject the null hypothesis (Benjamini and Yekutieli, 2001). FDR thus represents a less conservative approach that is particularly suitable if data is partly correlated. Multiple testing correction also plays an important role in neuroimaging, where usually many univariate tests (e.g. t-tests) are performed on a given set of voxels. Therefore, the neuroimaging field has developed other techniques that also consider spatial correlations between neighbouring voxels (i.e. smoothness and spatial autocorrelation) to generate more appropriate null distributions from the data (see Chapter 3.3.2).

3.4.1.4 Some pitfalls of frequentist statistics

The need to adjust for multiple testing reduces power of frequentist tests. Specifically, because frequentist statistics operate with long run frequencies of probabilities (the p-value), statistical test results require adjustment when being obtained during an ongoing sampling plan. Hence, the ability to implement flexible data collection stopping rules is limited with frequentist statistics because power is reduced (Schönbrodt et al., 2015; Schönbrodt and Wagenmakers, 2016). Further, flexible stopping is performed without multiple testing correction, such

undeclared and uncorrected flexible stopping rules constitute a (relatively common) questionable research practice (Chapter 1.1.1). Another pitfall of frequentist hypothesis testing represents its earlier mentioned (Chapter 3.4.1.1) bias against the null hypothesis (Berger and Sellke, 1987; Rouder et al., 2009). However, frequentist equivalence tests (Chapter 3.4.1.5) allow testing for other null hypotheses such as a range of SESOIs. Another often discussed disadvantage of frequentist statistics is that they remain agnostic to any prior knowledge. For instance, it may be possible that certain effect sizes are implausible for an experiment. Integrating prior knowledge and observed data underlies the principle of Bayesian statistics, which will be briefly introduced in Chapter 3.4.2.

3.4.1.5 Equivalence Tests

When a statistical test does not reach a declared significance level and the null hypothesis can thus not be rejected, it may be that there is truly no effect present, or that the study was not sufficiently powered to detect the effect of interest. Null hypothesis testing implies a point hypothesis that assumes that the effect is zero. This scenario, however, is rarely observed, partly owed to the fact that it is likely impossible to completely isolate phenomena and their effects from all other (correlated) influences that remain measurable. Moreover, if independent measurements are compared, independent (and identically distributed; *iid*) noise is assumed independent and hence even in the presence of no real effect, subtle differences in noise will remain for small samples and in particular for noisy measurements such as fMRI (Cremers et al., 2017; Eklund et al., 2016). Hence, most empirical non-significant results rather represent small effects that were not detectable with the given sample size, and that may be negligible in the first place. This raises the questions what constitutes a too small, and thus negligible effect, and how can negligible effects be rejected in a given data set? Frequentist equivalence tests allow addressing this question by testing whether a range of smallest effect sizes of interest (SESOIs) can be rejected in a given sample (Lakens et al., 2018; Walker and Nowacki, 2011). SESOIs are ideally specified *a priori* and chosen based on theoretically plausible effect sizes. When following up a null hypothesis test with an equivalence test based on a given SESOI, both tests may be not significant, indicating that the data remains inconclusive to reject null hypothesis as well as a range of smallest effect sizes of interest (Figure 3.3 A, B). In this situation, more data collection would be required to yield a conclusive test statistic. Alternatively, it is possible that the null hypothesis test remains inconclusive, but the equivalence test can reject the SESOI (Figure 3.3 C). In this situation, it is possible to show

that even though no overall significant effect can be established, smallest effect sizes that are worthwhile detecting can be rejected.

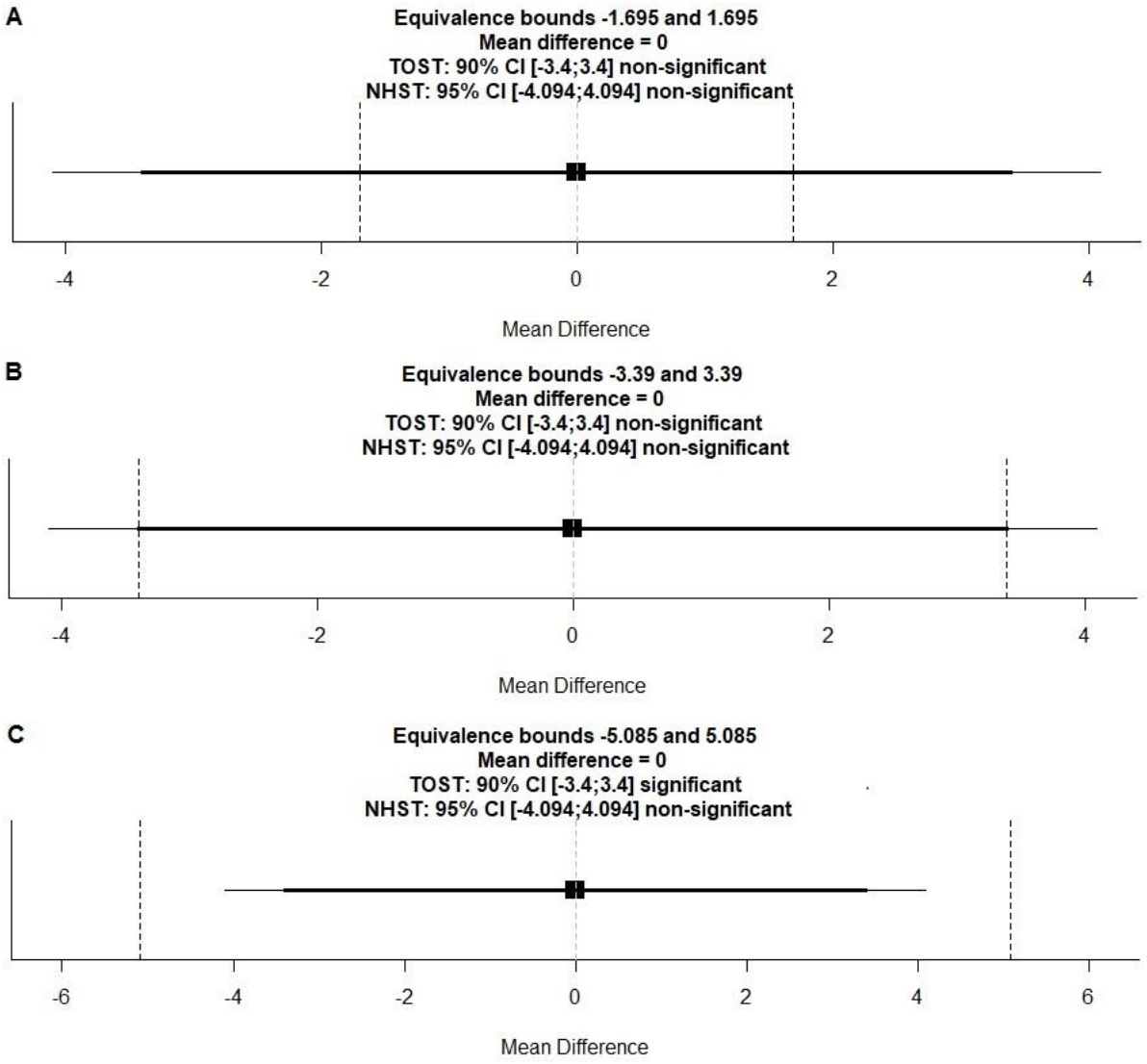

Figure 3.3 Example equivalence tests for a two-sided independent t-test and a group mean difference of zero and different SESOIs. Equivalence bounds set to Cohen's d = 0.3 (A), 0.6 (B) and 0.9 (C), corresponding bounds in native units shown with dotted vertical lines. Group mean difference centred at zero with its 90% confidence interval (CI; thick line) and for comparison also the 95% CI (thin line). Only if the 90% CI does not include the equivalence bounds (C), one can declare equivalence. TOST: Two one-sided test. NHST: Null hypothesis significance test.

3.4.2 Bayesian hypothesis testing

Bayesian statistics form a branch of statistics named after the 18th century British philosopher and statistician Thomas Bayes. They can address some of the limitations that are associated

with frequentist statistics and have hence been used alongside frequentist tests in this thesis. The following paragraphs intend to provide a short overview of the principles underlying Bayesian hypothesis testing and highlight features that can complement frequentist hypothesis testing.

3.4.2.1 Principles of Bayesian hypothesis testing

When researchers conduct an experiment, they are rarely completely agnostic about possible outcomes. In fact, they may have prior knowledge about plausible effects that could be used to inform their analysis and hypothesis tests. Further, this prior knowledge is likely associated with some uncertainty (if it was not, the experiment would likely be trivial and not funded/conducted in the first place). However, frequentist statistics do not allow to incorporate *a priori* knowledge and its associated uncertainties into statistical analyses and hypothesis tests. Bayesian statistics, in contrast, operate with conditional probabilities and thus allow incorporating prior knowledge. Frequentist statistical tests calculate the probability of the observed data given that the null hypothesis is true. In contrast, Bayesian statistical tests calculate the posterior probability that a specific model (and a hypothesis derived from it) is true given the evidence (i.e. data). In other words, Bayesian statistics equip researchers who intend to test the plausibility of a theory (and the models and hypotheses they have derived from this theory) with a technique that allows testing how well a model is supported by observations. Notably, the influence of new observations is weighted by prior beliefs about an event. The calculation of Bayesian updating is based on Bayes' theorem, which can be derived by applying the rules of conditional probabilities:

Eq. (3.4) $$P(A|B) = \frac{\mathrm{P(A\ AND\ B)}}{\mathrm{P(B)}}$$

, where A and B can be thought of as separate events. Applying the rules of conditional probabilities, Eq. (3.4) yields

Eq. (3.5) $$P(B|A) = \frac{\mathrm{P(B\ AND\ A)}}{\mathrm{P(A)}} = \frac{\mathrm{P(A\ AND\ B)}}{\mathrm{P(A)}}$$

Reordering Eq. (3.5) for $P(A\ AND\ B)$ yields

Eq. (3.6) $$P(A\ AND\ B) = \mathrm{P(B|A)} * P(A)$$

Substituting the numerator in Eq. (3.4) by the derived expression of the right-hand side of Eq. (3.6), Bayes' theorem is derived,

Eq. (3.7) $$P(A|B) = \frac{\mathrm{P(B|A)} * P(A)}{P(B)}$$

, where $P(A|B)$ represents the posterior (e.g. the probability for model A given the observed data B), $\mathrm{P(B|A)}$ the likelihood (i.e. the non-normalised probability function for a range of parameters/values under the model), $P(A)$ the prior (i.e. *a priori* beliefs in the model), and $\mathrm{P(B)}$ the marginal likelihood (i.e. the normalised likelihood function of the data integrated over the complete parameter/model space):

Eq. (3.8) $$posterior = \frac{likelihood * prior}{marginal\ likelihood}$$

The derived Bayes theorem allows parameter estimation by integrating respective probability distributions (e.g. a prior distribution, likelihood distribution of data) and thereby can ignore the marginal likelihood, which indicates the model evidence. In contrast, for Bayesian hypothesis testing the marginal likelihood is central because it indicates the probability of the data under a given model. The ratio of marginal likelihoods between competing models (for instance a null and an alternative models) is the Bayes Factor (BF; (Kass and Raftery, 1995)):

Eq. (3.9) $$BF10 = \frac{p(y|M1)}{p(y|M0)}$$

Eq. (3.9) indicates that Bayesian hypothesis testing allows model comparison, which represents a desirable property (Chapter 3.4.2.2). The Bayesian approach thus does not only allow to calculate the evidence for the null model relative to the alternative, but also any other model of interest. While the BF calculation is relatively simple for simple models, the underlying calculations become exponentially more computing intensive for complex models (Wagenmakers et al., 2010). However, it is possible to calculate the BF directly from the ratio between the density prior and posterior functions at a point of interest ("Savage-Dickey trick"):

Eq. (3.10) $$BF10 = \frac{p(\delta=0|M1,y)}{p(\delta=0|M1)}$$

Figure 3.4 provides an example for a two-sided independent t-test two groups where no group mean difference was found (Figure 3.3).

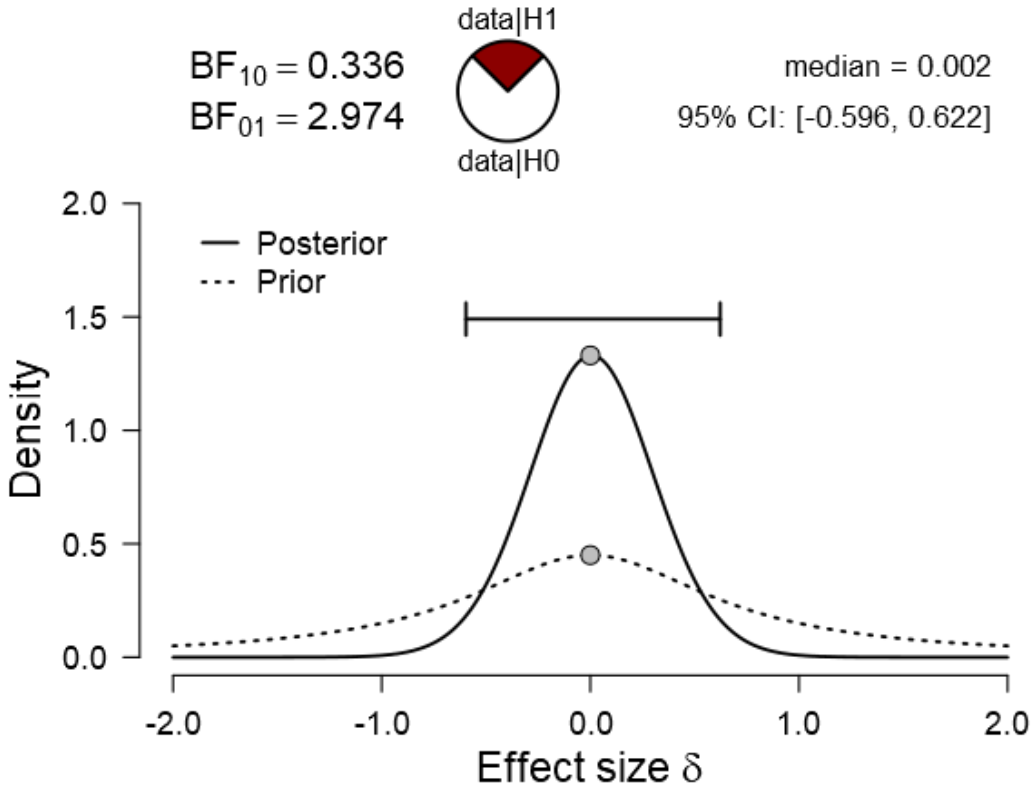

Figure 3.4 Example calculation of the Savage-Dickey ratio for a two-sided Bayesian t-test. A wide tailed Cauchy prior distribution scaled by an effect size r = 0.707 was used, representing a prior belief that 50% of expected effect sizes will fall within the range of -0.707 and 0.707. The point hypothesis was set to zero. The BF in favour of the alternative hypothesis (BF_{10}) is below 1, indicating that the evidence for the null hypothesis is larger (BF_{01} = 2.97). The relationship of evidence between the alternative and the null hypothesis is illustrated in the cake diagram. Calculations performed with the JASP software package (Team, 2017).

As mentioned earlier, prior distributions allow incorporating knowledge about plausible sizes of effects, as well as the uncertainty around these estimates, indicated by the density. For analysis based on continuous data, this uncertainty is expressed in the form of the prior distribution tails. For the example illustrated in Figure 3.4 a Cauchy prior distribution was used, which features wide tails and thus assigns some probability to extreme values. Besides Cauchy

priors, other prior distributions such as the normal or t-distribution can also be used and scaled by values that are deemed plausible effect sizes for the research question at hand (Rouder et al., 2009).

Bayes Factors allow expressing the plausibility of competing models (and hypotheses) in terms of likelihood ratios and can thus inform decision making, for instance with regards to best practice recommendations in medicine (Ashby, 2009). Hence, categories to aid the interpretation of BF calculations have been developed (Table 3.1) that provide a general indication for what is considered for example strong versus anecdotal evidence.

Statistic (BF_{10})	**Interpretation**
> 100	Decisive/extreme evidence for H_1
30 – 100	Very strong evidence for H_1
10 – 30	Strong evidence for H_1
3 – 10	Moderate evidence for H_1
1 – 3	Anecdotal evidence for H_1
1	No evidence in favour of H_1 or H_0
0.333 – 1	Anecdotal evidence for H_0
0.1 – 0.333	Moderate evidence for H_0
0.033 – 0.1	Strong evidence for H_0
0.001 – 0.033	Very strong evidence for H_0
< 0.001	Decisive/extreme evidence for H_0

Table 3.1 Evidence categories for Bayes Factor (BF_{10}), modified from (Wetzels et al., 2011).

This terminology will be used in this thesis where BF results are interpreted. Taken together, we have derived Bayes theorem that allows us to express beliefs about an event in terms of conditional probability to yield a posterior belief. In this probability calculation, uncertainty about prior beliefs is incorporated. The next paragraph will highlight a few aspects of Bayesian hypothesis testing that overcome some limitations of frequentist hypothesis testing.

3.4.2.2 Advantages of Bayesian statistics and hypothesis testing

Bayes factors allow quantifying the evidence for a hypothesis over another, given the observed data and prior beliefs. Important, BFs thus allow specifying evidence for the null hypothesis

relative to the alternative hypothesis. This property seems in particular relevant with regards to questionable research practices that have contributed to the replication crisis (Chapter 1.1) because it can render results more informative and conclusive (Dienes, 2014; Morey and Rouder, 2011; Rouder et al., 2009). Moreover, the Bayesian approach allows comparing an arbitrary number of models because the Bayesian approach is comparative (Eq. 3.9). Hypotheses are tested against each other in a way that yields three desirable properties:

1) Bayesian inference-based model comparison avoids overfitting of complex parameters because marginal likelihood integrates over the complete parameters space. Specifically, complex models assign lower likelihoods to a larger range of parameters (e.g. possible effect sizes), whereas simpler models assign larger likelihoods to a smaller set of parameters. This built-in "Occam's razor" (Smith and Spiegelhalter, 1980) protects Bayesian model estimation against overfitting and yields a model selection mechanism that prefers parsimonious models that are more likely generalisable.

2) One convenient mathematical property of BFs is their *transitivity*, such that they can be re-combined as a ratio that expresses the belief that one model is more likely than another. Given that BFs represent the ratio between the posterior to a prior distribution (Salvage-Dickey trick, Chapter 3.4.2.1), transitive combination of BFs requires that these were computed using the same data, that both original tests were carried out against one common hypothesis (e.g. the same null hypothesis), and that identical prior distributions were used. Transitive combinations of BFs allows asking new forms of research questions, for instance for the relative evidence of a positive opposed to a negative correlation or group mean difference between two conditions (see for example Chapter 7.5.4.2).

3) Bayesian inference is not affected by multiple testing and thus does not require multiple comparison adjustment. This is because Bayesian statistics do not treat probabilities as long run relative frequencies (as opposed to frequentist null hypothesis testing) and hence results are not subject to type-I error rates. Hence, repeated testing does not alter results and does not require adjustments of the test statistic. This allows investigators to describe how evidence is accumulated in an experiment as a function of new data being acquired (Figure 3.5). This property implies that Bayesian Inference allows flexible stopping rules for data collection (Schönbrodt et al., 2015). Employing a Bayesian sequential sampling (BSS) approach, one can design experiments in which data collection is continued until a predefined threshold of

sufficient evidence is reached (see sampling plan used in Chapter 7.2.3.2). Simulation results indicate that such an approach allows more economic, still robust designs (Schönbrodt et al., 2015; Schönbrodt and Wagenmakers, 2016). However, although the concepts of power and type-I error do not exist, Bayesian inference can still yield false positive and false negative results. The probability of false positive and false negative errors is determined by the choice of parameters (the scaling factors and the BF threshold), the sample size and the true effect determine the false negative rate false positive (Schönbrodt et al., 2015). These can be balanced against each other when designing a BSS plan.

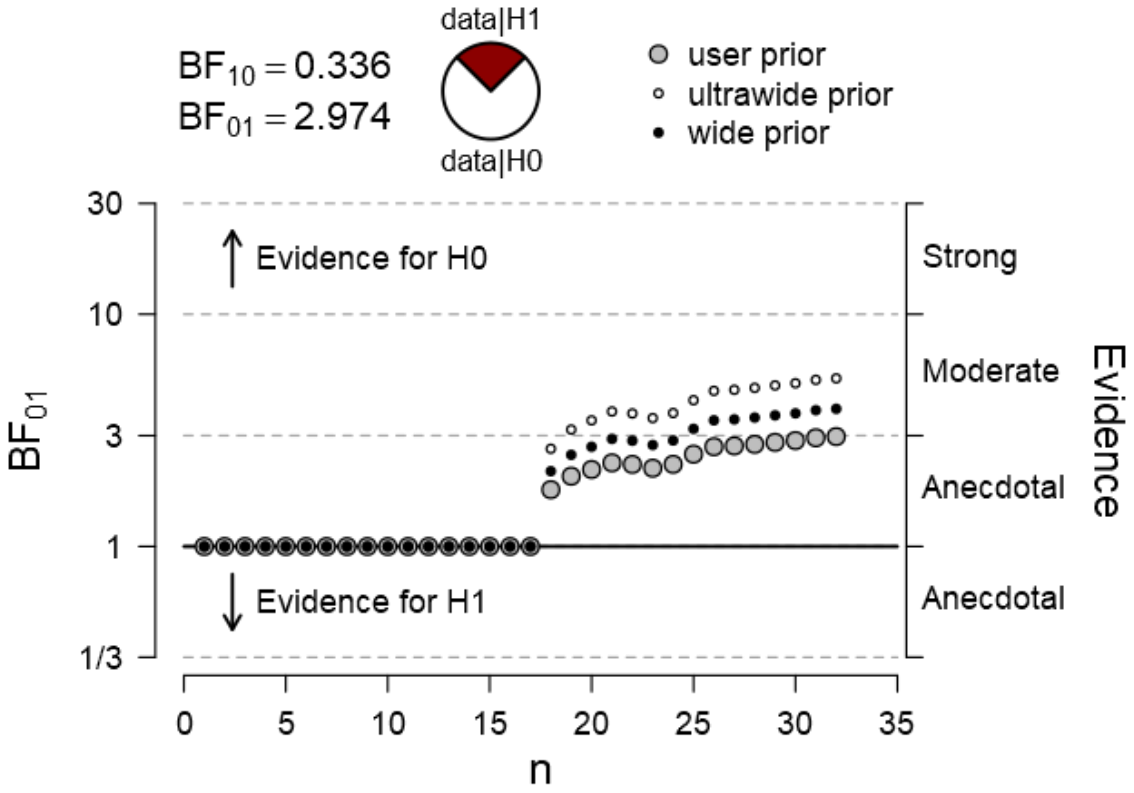

Figure 3.5 Bayesian Sequential analysis for a two-sided independent t-test. Data and prior identical to those used for Figure 3.4. The plot shows the accumulation of evidence in favour of the null hypothesis as a function of n (the number of participants). Dots of differences sizes and colours indicate different priors, indicating the sensitivity of the result to the prior distribution. Given that the data yielded a zero group mean difference, wider priors that cover a larger range of effect sizes yield relatively more evidence for a null effect.

Lastly, one potential criticism against Bayes may be that priors can be selected in arbitrary way (possibly after data has been analysed) and data is thereby overfitted. This criticism can be addressed in a number of ways, e.g. either by choosing very generalizable, so called default prior distributions (Morey and Rouder, 2011), or by providing convincing justifications for a given prior distribution choice (e.g. priors that are informed by SESOIs). Additionally, analyses can be amended by a sensitivity analysis that informs about the influence of different prior distribution on the Bayes Factor (see legend Figure 3.5).

3.4.3 Randomised Clinical Trials

Randomised Clinical Trials (RCTs) are considered the gold standard to demonstrate empirically the superiority of an intervention over a control condition (Bothwell et al., 2016). Tested interventions may include pharmacological substances but also training programmes that aim to modify behaviour or cognitive processes. Some main trial types, control procedures and analysis principles will be briefly outlined here.

One main rationale that guides trial design is minimising potential biases during data acquisition, analysis and interpretation. These aims are achieved by three main principles: 1) randomization of participants (Chapter 3.4.3.1) to minimise risk of confounding and bias, 2) control conditions (Chapter 3.4.3.2) to control for unspecific (e.g. placebo) effects, and 3) blinding of assessment as well as of treatment where applicable (Chapter 3.4.3.3).

For the development of new medical interventions and medical substances, RCTs are traditionally categorized in different phases (Table 3.3) based on their development stage and focus. Early phase trials focus on assessing mechanisms (in Phase 0, e.g. pharmacodynamics), where later trials in humans focus on establishing safety (Phase 1), efficacy (Phase 2). Final confirmation of safety and efficacy is provided in large sample RCTs (Phase 3) and follow-up safety studies (Phase 4) that are conducted after the intervention has received approval from medical administration councils.

Phase	Objectives	Dose	Approximate Size/Population
Phase 1	Pharmacodynamics, bioavailability	Often subtherapeutic, but with ascending doses	First in vitro and/or animal models;

	Testing of drug on healthy volunteers to confirm safety and likely therapeutic dose		Later healthy participants
Phase 2	Testing of drug on patients to assess efficacy and safety	Therapeutic dose	Several dozen up to a few hundred patients
Phase 3	Testing of drug on patients to assess efficacy, effectiveness and safety	Therapeutic dose	Over a few hundred patients
Phase 4	Post-marketing surveillance – monitoring the use of the drug after approval	Therapeutic dose	Anyone seeking treatment from their doctor

Table 3.2 Clinical trial phases for drug development, modified from (Friedman et al., 2010).

While these traditional categories may largely apply for RCTs involved in drug development and other medical procedures, they may not fully apply to non-invasive translational work where animal models are not available. Further, an understanding of a mechanism of action and efficient therapeutic doses may still be refined in late stages. For translational technologies, including brain circuit-based interventions such as DBS and fMRI-NF, it is particularly challenging to identify causal interactions in a complex system such as the human brain that can largely be only assessed non-invasively (e.g. using neuroimaging techniques). In the case of clinical fMRI-NF development, working hypotheses based on published literature about circuit dynamics usually replaces empirical Phase 0 studies. First empirical knowledge about technical feasibility is gained from pilot and feasibility studies in healthy humans (Phase 1), followed by small, often uncontrolled and not randomised (Thibault et al., 2018b) feasibility studies with patient groups (Phase 2). Albeit to a limited extent, first translations into RCTs (Phase 3) have been achieved (Chapter 2.1.2). However, in contrast to the phases of traditional drug development trials, uncovering mechanisms of actions remain an important component of Phase 3 trials in the development of clinical fMRI-NF. Taken together, while traditional phases of trial development provide a guidance for the pathway from a preclinical model to a

successfully developed health product, the phases of clinical fMRI-NF development partly deviate and may necessitate a theoretical framework.

3.4.3.1 Randomisation

Randomisation procedures assign participants to treatment groups. The goal of randomisation procedures is to eliminate potential pre-treatment differences between groups to allow a pre-post treatment comparison where group results mainly stem from differences in the intervention. Thereby, randomisation procedures aim to avoid pre-existing group differences regarding the primary (or secondary) outcome, as well as for factors that may interact with (i.e. confound) treatment effects. Different approaches exist: for instance, during *simple randomisation*, participants are randomly assigned to intervention groups, but the randomisation procedure remains agnostic to individual factors. In contrast, during *adaptive randomisation* a minimisation procedure can be applied such that new participants are assigned to a group in a way that minimises between group difference of possible confounding factors (see Chapter 4.2.2). These factors typically include age, gender, duration and severity of an illness, as well as the medication status (if applicable) in clinical trials. Hence, an adaptive randomisation sequence is informed by the data and aims to balance for potential confounding factors between groups. With regards to randomisation procedures of fMRI-NF studies, it should be mentioned that additional randomisation factors may be relevant depending on the choice of control conditions (Chapter 3.4.3.2). For instance, researchers may expect to find differences in the ability of self-regulation based on certain neuroimaging or psychometric markers.

3.4.3.2 Control Conditions

Control conditions in RCTs mainly serve to minimise the influence of potential confounders and provide evidence for the specificity of potential treatment effects. An effective control condition can be designed based on epidemiological insight of the condition and the pathophysiological model of interest. Moreover, the nature of the intervention may require different forms of control because multiple elements of the intervention may interfere with therapeutic processes and confound results. For fMRI-NF, multiple purpose of control groups have been identified (Thibault et al., 2016): Groups may for instance be matched for received feedback (*yoked feedback*) by providing participants in the control with non-contingent sham feedback from participants of the experimental group. Another purpose may be to control for motivation levels and the belief of self-control by providing contingent sham feedback from an

area that is not related to a pathophysiological model of interest. Inverse sham feedback represents an alternative control condition. In this setup, participants engage in neurofeedback training for which opposite effects are expected, for instance up-regulation and down-regulation motor imagery-based fMRI-NF training of the motor area). Other control conditions do not involve fMRI acquisition but merely exposure to similar cognitive training such as mental imagery training only control groups. Altogether, fMRI-NF experiments are complex experiments and thus consist of many elements that can be controlled for. Moreover, since fMRI-NF training involves an interaction of the individual with the setup, the implementation of control conditions should be informed by the level of blinding that is desired (Chapter 3.4.3.3). A more in-depth discussion of pros and cons for different control conditions will be provided in the general discussion (Chapter 9).

3.4.3.3 Blinding

Blinding in RCTs aims to minimise potential biases and confounds that may result from knowledge about treatment allocation. Studies that do not use blinding procedures are considered open label studies. Traditionally, three grades of blinding procedures have been described for RCTs: single-blind, where conventionally only the participant (or for some designs only the investigator) is blinded to the intervention allocation, double-blinding where both the participant and investigator are blinded, and triple-blinding where additionally the data analyst is blinded (Umscheid et al., 2012). A lack of sufficient information about blinding procedures of trials has been documented and criticized, for instance every second study published in PubMed journals in December 2000 did only refer to the blinding type but not provide details on the procedure (Chan and Altman, 2005). Hence, applied blinding procedures require scrutiny in the evaluation of RCTs, for interventions such as fMRI-NF where active involvement of the patient forms part of the procedure. While these grades of blinding are generally applicable for trials that test interventions which do not require an active interaction of the patient, feasible blinding strategies for fMRI-NF trial designs largely depend on the choice of control conditions (Chapter 3.4.3.2).

3.4.4 The CONSORT statement

In response to reporting deficits of clinical trials, including lacking details on blinding status, but also outcome specification and randomization procedures, the Consolidated Standards of Reporting Trials (CONSORT) statement was introduced in 1996 with the aim to improve the quality of reporting in clinical trial research. CONSORT is considered best practice in clinical

trial reporting and forms a prerequisite for article submissions of RCTs to most medical journals. CONSORT includes detailed specification items on blinding procedures since its in 2001 revised version (Moher et al., 2001). In its currently recommended version, which was revised in 2010, CONSORT comprises a 25-item checklist and a flow chart diagram about patient flow (Appendix A Figure A1). The checklist requires the provision of details about the methodology (e.g. blinding and randomization), results (e.g. attrition rates, main clinical outcome and exploratory analyses), as well as information about a public registration, trial protocol and funders. According to one large COCHRANE meta-analysis that included 50 published evaluations of reporting standards across more than 16,000 entries (Turner et al., 2012), the introduction of CONSORT was associated with an improved quality of clinical trial reporting. For instance, trials increasingly reported randomization concealment (i.e. separating randomization from recruitment) and mention the rationale behind an intervention. Yet, details of the blinding procedure remained insufficiently specified in about 80% of trials (Turner et al., 2012). Taken together, the CONSORT statement and its endorsement by major medical journals has likely improved the reporting standards in trials research. However, given that certain aspects of the CONSORT checklist remain underspecified, it likely does not (yet) constitute a decisive element in the peer-review process.

3.4.5 Trial registration

Trial registrations are a public record of clinical trials that contain key information including outcome measures, start and end dates, as well as planned sample sizes. They have been introduced in response to reporting biases in the medical literature and have been a requirement by the International Committee of Medical Journal editors since 2004 (De Angelis et al., 2004). To avoid being biased by preliminary results, trials registrations need to be completed before data collection. Registered trials receive ID numbers, which are required to be reported in submitted manuscripts. The largest registry for clinical trials is ClinicalTrials.gov, which has been founded in February 2000. As of August 2013, more than 150,000 entries were made (Jones et al., 2013). Moreover, since September 2008 a results data base is available, which contained data of more than 7,600 trials as of February 2013 (Williams et al., 2015).

3.4.6 Missing data and data analysis

Clinical trials require the documentation of missing outcomes and noncompliance to allow assessing safety and potential sources for biases. For instance, when comparing two treatment methods, it is possible that attrition rates vary between treatment A and treatment B, which may indicate worse compliance for treatment B. Furthermore, if attrition interacts with certain patient characteristics such as the severity of condition, outcome estimates may bias. For instance, it may be possible that despite similar drop-out rates between treatment conditions, predominantly severe cases interrupt treatment B. To address this issue and test for potential bias, researchers can use intention-to-treat (ITT) analyses (Gupta, 2011), which are considered best practice. Another aspect worth emphasizing is that analyses of RCT outcomes may benefit from adjustment of covariates. Four main reasons for main goals of covariate adjustment include 1) to correct for potential slight imbalances in baseline prognostic covariates (e.g. baseline severity or medication status), 2) to increase statistical power by modelling variability in the outcome variable that is explained by prognostic covariate, 3) to arrive at treatment effects that are more representative to treatment effects of individual patients (because individual variation in dependence of covariates such as age or gender is taken into account), and 4) to incorporate specific study design features in the data analysis (Yu et al., 2010). CONSORT recommends adjusting for covariates and predeclaring these in the trial registration to restrict degrees of freedom for p-hacking. However, the majority of trials often do not follow these recommendations (Yu et al., 2010).

3.5 Preregistration and Registered Reports

Trial registration and CONSORT statements inspired similar developments outside of clinical medicine. In particular, open science methods including preregistration and registered reports (RRs) have been suggested as a solution to address *researcher bias* and thereby make science more reproducible and replicable (Chambers, 2017; Ioannidis, 2012; Munafò et al., 2017; Peng, 2011). For instance, by defining and publishing hypotheses and analysis pipelines before data collection, the robustness of statistical inferences is improved as post-hoc analysis options are rendered transparent (Chambers, 2013). This public declaration is central to the new publishing formats of study preregistration and RRs. These two formats ask researchers to predefine their hypotheses, methods and analysis plans which are then published. Hence, both preregistrations and RRs make the crucial distinction between confirmatory hypothesis testing and post-hoc exploratory hypothesis generating and thus address a major contributor to the replication crisis

(Munafò et al., 2017). Preregistrations are comparable to clinical trial registrations and comprise a document that describes the rationale and hypotheses, however, they usually describe the methodology in more detail and they should provide an analysis plan that is registered and accessible at a public repository. RRs represent a peer-reviewed extension of preregistrations (Nosek and Lakens, 2014). Specifically, the initial peer review of RRs (Stage 1) should be based on criteria such the studies' scientific validity, the plausibility of statistical power calculations, and its novelty (Figure 3.6). Approved submissions receive an acceptance in principle. Originally launched at the neuroscience journal Cortex (Chambers, 2017), more than 100 journals offer a form of RR as of January 2018.

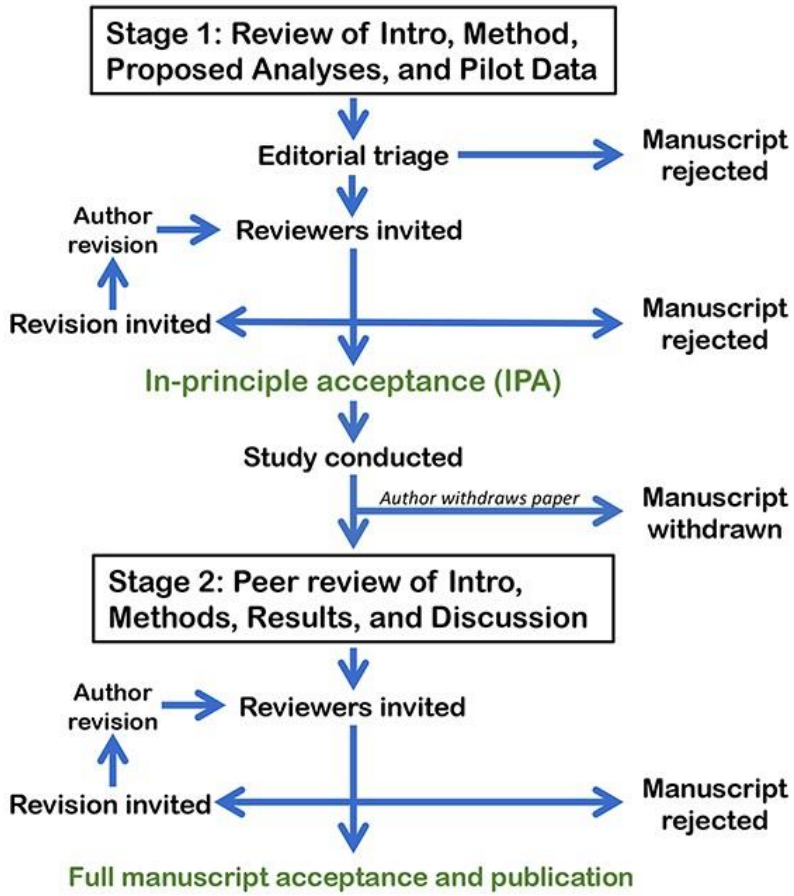

Figure 3.6 Registered Reports (RRs) Workflow Diagram adapted from (The Registered Reports Steering Committee, 2018).

After data collection and analysis is completed, a second review is performed (Stage 2) to assess that analyses were performed according to the analysis plan and that authors draw valid conclusions. Final manuscripts may also contain exploratory analyses that were not preregistered, but importantly these are indicated as such, reducing their evidential value. Crucially, final manuscript acceptance is not contingent upon results and authors can hence

publish studies even in the event of an inconclusive null result. Taken together, RRs can effectively reduce, or even eliminate *researcher bias* including *file drawer effects*, as well as *publication bias*. Transparency can be further increased by providing access to data, computer code and research materials. Overall, transparency as evidence by preregistrations and registered reports is becoming increasingly desired and required by journals and founders (Bloom et al., 2014; Eglen et al., 2017; Picciotto, 2018). Moreover, increased transparency encourages and facilitates direct replication attempts and allows immediate comparison between studies, fostering one pillar of the scientific endeavour. Necessary infrastructures and incentive structures are largely not in place yet, however, they become increasingly available. Taken together, publishing formats such as preregistration and RRs likely transform the way research is conducted and communicated, equipping scientists with new tools to control for bias and falsify previous work to advance theory and science (Popper, 2005).

4 CLINICAL EFFECTS OF FMRI-NF TRAINING IN DEPRESSION

4.1 Introduction

Neurofeedback can enable patients to develop personal strategies that are effective in self-regulating brain areas and networks associated with mental imagery through the feedback of signals that reflect their own neural activation patterns. Mental imagery can be therapeutic for depression by increasing cognitive flexibility and capacity for positive mental simulation (Holmes et al., 2016). Thus, with neurofeedback, patients' engagement in mental imagery can be enhanced by monitoring and feeding back the associated brain activation. Proof-of-concept (PoC) studies with mildly to moderately depressed patients have shown that patients can learn to up-regulate brain areas localised with positive affective stimulation (Linden et al., 2012; Young et al., 2014). Moreover, one study suggested that only the group that completed real-time fMRI-NF training, but not a mental imagery only control group, experienced significant improvement in mood within three training sessions (Linden et al., 2012). These findings have recently been corroborated in a double-blind, placebo-controlled RCT conducted in unmedicated patients (Young et al., 2017). This RCT found over 40% reduction in depressive symptoms in the intervention, but not a placebo neurofeedback group. However, patients in the placebo group of that trial did not attain self-regulation success, and thus groups were not matched for the reward experience. Hence, it remains unclear whether this efficacy is specific to feedback from emotion-regulating regions, or whether the general experience of self-regulation during fMRI-NF may be therapeutic. Further, the clinical efficacy of fMRI-NF as an add-on treatment to psychopharmacology remains to be tested. Also, no follow-up tests have yet been conducted to determine whether clinical effects are lasting. This chapter presents results from an RCT of rt-fMRI-NF training in medicated patients with moderate to severe depression. The current trial compared up-regulation of brain areas involved in emotion processing (NFE) with an active control procedure that involved up-regulation of brain areas involved in higher visual processes (NFS). Specifically, both groups were provided with information that enabled patients to up-regulate selected target areas. Whereas the NFE group was instructed to imagine positive images similar to those seen in the respective blocks of the localiser run (Linden et al., 2012), the NFS group was instructed to imagine relaxing scenes (Habes et al., 2016).This design also gave patients of the control group the opportunity to experience self-regulation success and reward experience. The main hypothesis was that neurofeedback of areas involved in processing positive emotions would produce clinical improvements exceeding those seen in the active control group. Patients' clinical status was assessed again at a follow-up approximately 6 weeks after the completion of the intervention.

4.2 Methods

This section exclusively covers methodological details about patient recruitment, randomisation, the overall trial design, clinical and psychometric measures, the power calculation for the main clinical outcome measure, and statistical analyses of planned and exploratory analyses. For methodological details about the MRI acquisition and analysis, please refer to Chapter 5.2.

4.2.1 Patient recruitment

The study was approved in January 2012 by the South East Wales Research Ethics Committee and registered in February 2012 (NCT01544205). The first patient started the study in March 2012. All patients provided written informed consent and were compensated for their time and travel costs in cash. Patients were recruited via general practitioner (GP) surgeries and Community Mental Health Teams (CMHTs) in South Wales and the National Centre for Mental Health (NCMH). Patients had to meet all the following inclusion criteria:

- diagnosis of unipolar depression,
- currently moderate or severe depression confirmed with the Mini International Neuropsychiatric Interview (MINI),
- and current antidepressant treatment (with no change of dose in the preceding three months).

Patients meeting following criteria were not included in the trial:

- psychotic symptoms,
- current or lifetime substance or alcohol dependence,
- eating disorders,
- claustrophobia and other MRI contraindications,
- and ongoing non-pharmacological treatment (e.g. psychotherapy).

4.2.2 Randomisation and masking

Patients were randomly assigned to one of two groups using an adaptive randomization protocol developed by the South East Wales Trials Unit (SEWTU). The randomization protocol allocated patients to two groups, minimising for differences in age, gender, duration of illness, medication type (with three categories: SSRI only; non-SSRI antidepressant; combination treatment) and baseline depression severity as measured with the Hamilton Depression Rating

Scale (HDRS-17). After the patient had consented and completed all baseline measures, these were entered in a computer program (scripted in Microsoft Excel) and an allocation was provided to the investigators conducting the study. Investigators running the MRI sessions needed to know group allocation to apply the appropriate imaging protocols, but those conducting the assessments were blind to group allocation.

4.2.3 Trial design

The intervention in both groups consisted of five training sessions, starting with four weekly sessions followed by a consolidation session after a break of one month to test if patients retained the ability to self-regulate target ROIs. The third session served as a transfer session during which no neurofeedback was provided to test if patients could up-regulate target ROIs based on successful strategies learned during the previous two neurofeedback sessions. The actual duration of the intervention period was on average 12 weeks because of the need to schedule sessions around patients' availability. Before the randomisation, baseline measures of all clinical outcomes were recorded. Clinical measures were recorded again after the fifth neurofeedback session (week 12), and at follow-up (FU; on average week 18 after start of the intervention). All patients who completed the trial received verbal debriefing at FU.

4.2.4 Primary (registered) clinical outcome measure

The primary outcome measure was group mean difference in the Hamilton Depression Rating Scale (HDRS-17) (Hamilton, 1960) at the end of the intervention (session 5), which was administered by a clinician who was blinded to treatment group.

4.2.5 Secondary (registered) clinical outcome measure

Secondary clinical outcome measures were the group mean differences in the HDRS at follow-up and group mean differences at both time points in the Hospital Anxiety and Depression Scale (HADS) (Zigmond and Snaith, 1983), which includes anxiety (HADS (A)) and depression subscales (HADS (D)), the Quality Of Life scale (QoL) (Flanagan, 1982), and lastly the EuroQol research foundation questionnaire (EQ-5D-5L), which assessed the subject's health utility.

4.2.6 Other (registered) outcome measure

The following self-rated psychometric measures were also acquired before and after the intervention and at follow-up: The Thought Control Ability Questionnaire (TCAQ) (Luciano et al., 2005) to measure the perceived ability to control unwanted, intrusive thoughts; the Thought Control Questionnaire (TCQ) (Wells and Davies, 1994) to measure techniques that individuals use to control unpleasant and unwanted thoughts; the Self-Efficacy Scale (SES) to measure optimistic self-beliefs to cope with difficult life demands, with subscales for General Self-Efficacy (SES (G)) and Social Self-Efficacy (SES (S)) (Sherer et al., 1982); the Behavioral Inhibition System and Behavioral Activation System (BIS/BAS) (Carver & White, 1994) to assess approach and avoidance motivation. Further, patients' employment status and healthy service utilisation was also recorded (not reported here). Lastly, patients completed the Profile of Mood States (POMS) questionnaire before and after every neurofeedback session to assess the current mood state and level of fatigue (McNair, 1971).

4.2.7 Power calculation

The data from our pilot study (Linden et al., 2012) yielded an effect size (Cohen's d) of 1.5 for the group difference in HDRS-17 improvement between active neurofeedback and a control intervention. Expecting a slightly more conservative effect size of 1.2, it was estimated that a sample size of 15 patients in each group would achieve >80% power for post-hoc t-tests (Bonferroni corrected, alpha-level 0.025, two-sided) and set a recruitment target of 40 patients to allow for 25% attrition. Although this effect size is unusually large for add-on medication treatments (Cuijpers et al., 2014), similar effect sizes have been reported for some non-pharmacological interventions (Janssen et al., 2016), as well as for rt-fMRI-NF compared to a sham feedback condition (Young et al., 2017). Furthermore, because fMRI-NF is still in the early phases of clinical evaluation and there is no agreement on the most suitable comparators for clinical trials, logistic and economic factors were also taken into consideration when setting up the trial.

4.2.8 Statistical tools and planned analyses

The factors used to randomise patients (gender, age, duration of illness, dummy-coded medication type, and baseline HDRS-17) were first tested for any systematic differences between treatment conditions using frequentist and Bayesian statistical tests (independent t-test for continuous variables, chi-square association test for categorical items). For Bayesian tests, default Cauchy priors (r=0.707) were used.

The primary and secondary outcome measures for patients with complete data were analysed using Statistical Package for Social Science (SPSS, version 23). To test for any significant differences between groups, linear regression analyses were performed on the outcome measures post-intervention (separately for the primary endpoint after session 5 and for the follow-up session) with pre-intervention scores as regressors of no interest. Further, the following regressors of no interest were entered to reflect the minimisation procedure applied during randomisation: gender, age, duration of illness, dummy-coded medication type, the score for the respective outcome variable at baseline and HDRS-17 score at baseline (Yu et al., 2010).

P-values for all secondary and other measures were adjusted for a False discovery rate at q = 0.05. For POMS scores that were assessed immediately before and after each neurofeedback session, the research question was slightly different. The focus of analyses was changes in the POMS $_{TMD}$ (total mood score) and the depression subscale (POMS $_{Depression}$). Hence, this data could indicate immediate changes in mood and depression following neurofeedback training. Second, they gave an indication of changes that occurred between the first and fifth session for which no clinical data was assessed. Data from one participant was incomplete and thus excluded from analyses. Further, pre-post scores for POMS $_{TMD}$ were not available at session 3 for one patient. Hence, analyses conducted on POMS $_{TMD}$ were carried out with N = 30, and POMS $_{Depression}$ were carried out with N = 31. To test for pre-post changes, potential effects of group, or time, within-session difference scores for each session were submitted to mixed ANOVA with session as the within and group as the between factor. To account for the severity of patients' depression, pre-score differences between the first and last session, as well as their (dummy coded) medication status were entered as covariates. To explore for potential gender differences, these were entered as a grouping factor (Linden et al., 2012). These analyses were carried out in the Statistical Package for Social Science (SPSS, version 23). To follow-up for pre-post differences across the intervention, the pre-test scores (i.e. baseline) of session 1 and post-test scores of session 5 were submitted to two-sided frequentist and Bayesian paired samples t-tests. The Bayesian t-tests were carried out using a default prior (Cauchy r = 0.707). To compare within and between-session changes, data was plotted, and descriptive statistics reported. To assess whether within session changes showed an effect of time, pre-post difference scores of each session were submitted to a repeated measures ANOVA consisting on the within-subject factor session (5 levels).

Pre-post differences within group means were also computed for all primary, secondary and other clinical outcome measures. For all pre-post comparisons, %-change and 95%-confidence intervals of the changes in scores were calculated separately for both groups. Remission was defined by an HDRS-17 score ≤ 7, and remission rates were calculated based on all patients for whom we had complete data up to the relevant time point. Effect sizes (ES) of clinical outcome on the HDRS-17 were quantified using Hedges g based on difference scores with respect to baseline for both session 5 (post-intervention) and FU. Confidence intervals were bootstrapped based on 10,000 iterations using the R package bootES (version 1.3-20) (Canty and Ripley, 2017).

4.2.9 Exploratory analyses

4.2.9.1 Intention-to-treat analyses

Intention-to-treat (ITT) analyses aim to test whether patients who have completed an intervention differ to patients who have interrupted an intervention with regards to intervention related, or relevant prognostic factors (e.g. group allocation or clinical severity; 2.3.6). The ITT was not specified and is thus listed as an exploratory analysis. For the current study, ITT was limited to follow-up tests of baseline characteristics because clinical re-assessment of the primary outcome measure (as well as all other clinical measures) was performed after trial completion (session 5). Hence, with the data available in the present trial we could address two questions:

1) Systematic differences between treatment groups regarding side effects, comfort or training feasibility would likely be reflected in differences of attrition rates. Hence, the group allocation of patients who did not complete the intervention (i.e. did not attend their 5^{th} scanning session) was tested with a frequentist and a Bayesian binominal test (using a test value of 0.5, which reflects chance level).

2) It was possible that patients with certain baseline characteristics were more likely to stop the trial. Hence it was tested whether patients who did not complete the intervention differed on average in their baseline HDRS-17 scores and the duration of illness using an analysis of covariance (ANCOVA), in which baseline HDRS-17 or duration were the dependent variable, respectively, and binary coded completion status a fixed factor, and age, gender, dummy-coded

medication type and group allocation entered as covariates. Both baseline characteristics (HDRS-17 and duration of illness) were further compared between completers and non-completers using a frequentist Welch t-test (Delacre et al., 2017) and a Bayesian independent t-test with a normal prior distribution (N(0, 0.3)), which is considered minimal clinically important difference.

4.2.9.2 Testing for null effects

Given the absence of group effects for the main clinical outcome (HDRS-17; see Results), post-hoc tests were conducted to assess if 1) evidence for a null effect based on an assumed effect size (Cohen's d = 1.2) that was used for the power calculation could be found (Quintana and Donald, 2017), and 2) if a smallest clinically meaningful effect size could be rejected (Lakens, 2017; Walker and Nowacki, 2011). To assess evidence for the null, Bayesian ANCOVA with treatment group as a fixed effect and the randomisation factors (gender, age, duration of illness, medication type and the HDRS-17 score at baseline) as nuisance covariates were carried out using JASP (JASP Team (2016), version 0.8.0.1). The prior for group as a fixed effect was scaled by r = 1.2, which reflects the effect size used for the power analysis. The prior for covariates was scaled by a default value (r = 0.354).

To test if the smallest effect size of interest (SESOI) could be rejected, two one-sided tests for equivalence (TOST) were performed (Schuirmann, 1987) using the TOSTER package (version 0.2.3) in RStudio (version 1.0.136). A SESOI = -0.3 to 0.3 [with raw score lower limit 90% CI = -1.695 and upper limit 90% CI = 1.695] was chosen, which has been suggested as a common effect size for the additional benefit of anti-depressive medication compared to placebo (Kirsch et al., 2008). Although this was lower than our originally expected effect size, this analysis allows to test whether group effects that are clinically still relevant could be rejected at the nominal 5% alpha level. Moreover, based on the confidence interval around the empirical effect size for a group difference, equivalence bounds were calculated (i.e. the lower and upper bound for the smallest SESOI that could be rejected in either direction). This allowed exploring the limit of effect sizes that one could reject based on the data from the current sample.

Given the absence of group effects, the following analyses were conducted based on the complete sample (i.e. data was pooled across both interventions NFE and NFS).

4.2.9.3 Test against placebo effect reported for high-technology intervention in depression

Given the absence of a group mean difference on the HDRS-17, but an overall strong and similar clinical response in both groups, it was tested whether unspecific effects reported for high-technology augmentation therapy could be rejected. For example, an improvement of 5.5 points on the HDRS-17 was reported for sham whole-body hyperthermia treatment (Janssen et al., 2016).

4.2.9.4 Relationship self-efficacy and clinical improvement (HDRS-17)

Self-efficacy describes an individual's self-reported capacity to cope with challenges (Bandura, 1982). To test whether pre-post changes in self-efficacy (combined scores of the general and social self-efficacy subscale) predicted changes in depression scores, a regression analysis was performed on residualised HDRS-17 scores of the primary endpoint (session 5). Specifically, to minimise potential confounding by baseline HDRS-17 and baseline self-efficacy scores, these were first regressed from primary endpoint HDRS-17 scores. Obtained residuals were submitted to a robust (iteratively weighted least squares) regression analysis, a technique that is more robust against extreme data points (Rousseeuw and Leroy, 1987).

4.2.9.5 Changes in HDRS item scores

HDRS-17 sum scores only reflect overall changes over a range of symptoms, including depressive mood, feeling of guilt, suicidality, and anxiety, as well as physiological and somatic symptoms including different aspects of insomnia, gastrointestinal symptoms (GIT), genital symptoms (GS) and weight loss (WS) (Hamilton, 1960). Further, Cohen's d effect size was computed, including its 95% CI between baseline and primary endpoint, as well as between baseline and follow-up using the R package MBESS (version 4.4.3). Given the high number of items (N = 17), the relatively small sample (N = 16 per group) and the lack of an *a priori* hypothesis, no inferential statistics were applied. Any inferential statistics would have required substantial correction for multiple comparison, after which any incidental effect would likely be insignificant.

Apart from potential group differences, changes over time for different items were also of interest. For instance, SSRI treatments are often associated with an improvement of most clinical scores, with the exception of GIT, GS, and WS (Hieronymus et al., 2016). In fact, scores on these items tend to increase in response to SSRI treatment, which likely result from

side effects on associated physiological systems (Geddes et al., 2012). To limit the number of tests, a sign rank test was only carried out for these three items (GIT, GS and WL) for the difference between the primary endpoint and baseline and FDR corrected for multiple testing.

4.2.9.6 Relationship between POMS and HDRS-17

To investigate the relationship between POMS and HDRS-17 data, their relationship between baseline measures as well as between their between-session difference scores (i.e. difference between HDRS session 1 and 5 vs difference between POMS score at baseline session 1 and post-test session 5) were explored for POMS $_{TMD}$ and the POMS $_{Depression}$ subscale using right-sided Spearman's rho correlation analyses.

4.3 Results

4.3.1 Recruitment, randomisation and termination

The patient flow is summarised in the CONSORT diagram (Figure 4.1). A total of 147 patients agreed to be screened for the present study, of which 104 patients were excluded because they did not meet inclusion criteria or because they decided not to participate in the trial. The remaining 43 patients were randomised. Recruitment ended in June 2014 when the sample size target was achieved, and the last follow-up assessment was completed in September 2014. 32 patients (16 per group) completed the intervention: their demographical data are documented in Table 4.1. Overall, 28 patients (88%) also participated in the FU assessment. With respect to the total number of randomised patients, the attrition rate across both groups was ~26% at the end of the intervention (session 5) and ~35% at FU. Three patients in the NFE group increased their antidepressant medication during the trial, and five patients (four in the NFE and one in the NFS group) decreased or stopped medication. Hence, overall attrition rates in our trial were relatively low and comparable to other non-invasive intervention methods tested for MDD (Hollon et al., 2014; Janssen et al., 2016), suggesting that the procedure was well accepted by patients.

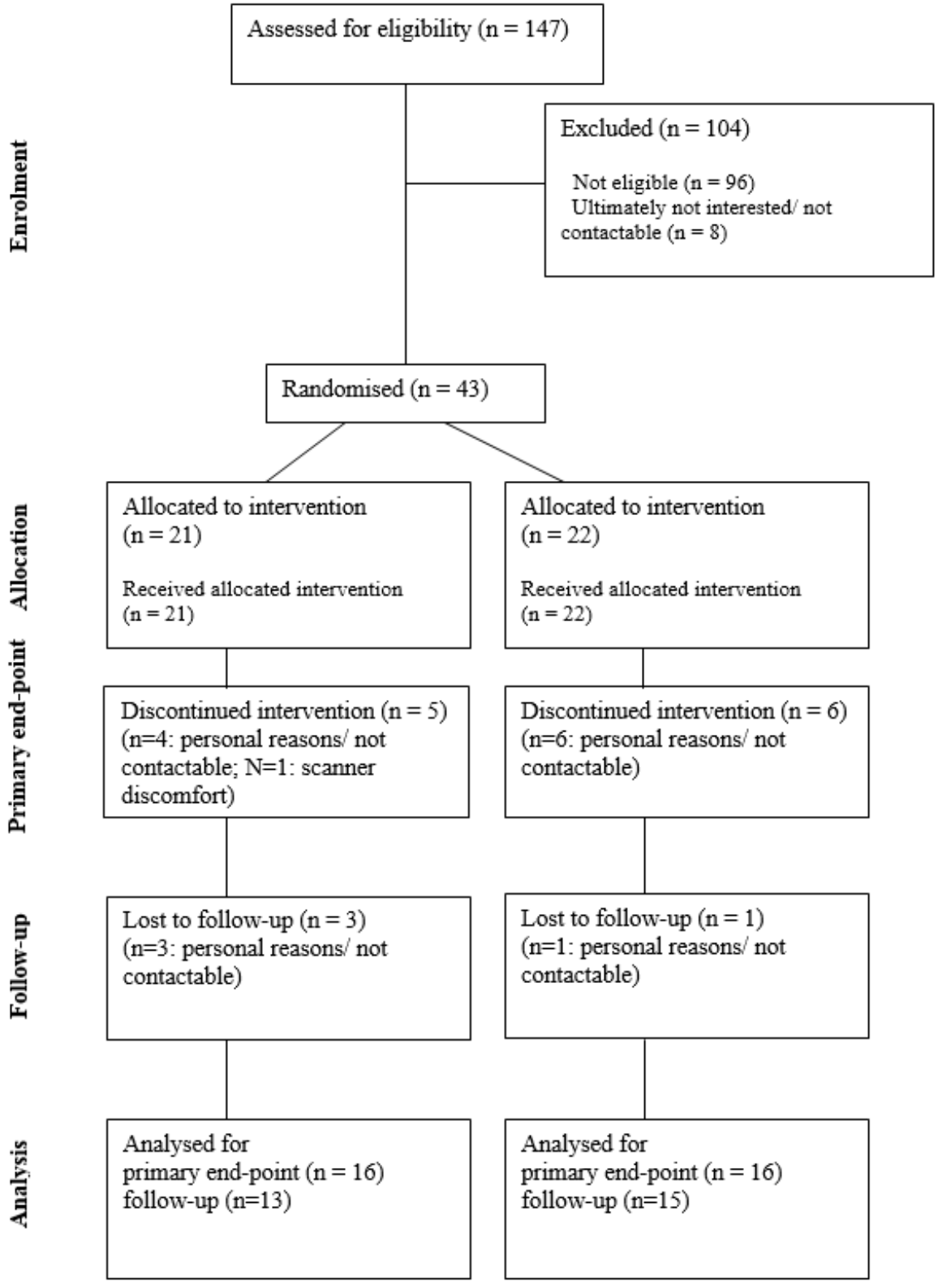

Figure 4.1 CONSORT patient flow diagram. Adapted from (Mehler et al., 2018).

4.3.2 Attrition, baseline characteristics, and completion rates

In total, 32 patients (16 in each group) completed the intervention, yielding an attrition rate from intervention start to the primary endpoint of about 26% (11/43) across both groups. All remaining 32 patients were included in the analysis of clinical, psychometric and neuroimaging data at the primary end point. Demographical and baseline data of patients who completed the intervention are documented in Table 4.1. Post-hoc analyses showed that randomisation was successful such that there no group differences with regards to the randomization factors gender, age, duration of depressive disorder, dummy coded medication type, and baseline HDRS-17 (Table 4.1). Also, with regards to handedness no group differences were detected.

Characteristics	NFE Group (n = 16)	NFS Group (n = 16)	Group differences	
			Statistical values	P values / BF_{01}
Gender	Female = 11; Male = 5	Female = 10; Male = 6	Pearson Chi-square = 0.139	0.71 / 2.374
Age (years)	47.19 ± 12.50	46.94 ± 12.74	$t_{30} = 0.6$	0.96 / 2.702
Handedness	Right = 14; Left = 1; Ambidextrous = 1	Right = 13; Left = 3	Pearson Chi-square = 2.037	0.36 / 1.133
Depression severity (HDRS-17)	19.88 ± 3.65	19.09 ± 5.09	$t_{30} = 0.5$	0.62 / 2.965
Duration of depressive	19 ± 12.39	18.56 ± 14.76	$t_{30} = 0.9$	0.93 / 2.970

disorder (years)				
Medication	SSRI only = 4; Non-SSRI=6; Combination*=6 (Combined = 3; Augmentation = 3)	SSRI only = 7; Non-SSRI=5; Combination*=4 (Augmentation = 4)	Pearson Chi-square = 1.31	0.52 / 2.557

Table 4.1 Demographical and clinical baseline data of completers. Estimates of duration reported as group mean ± standard deviation. (*): Combination included patients who received combined therapy with two compounds classified as antidepressant medication, and patients who received augmentation with an additional compound that is classified as a mood stabiliser (lithium or 2nd generation antipsychotic in addition to antidepressant). Adapted from (Mehler et al., 2018).

Of the remaining 32 patients, 28 patients (~88%) also attended follow-up assessments and were included in the analysis of clinical and psychometric data at FU, yielding an attrition rate from intervention start to FU of about 35% (15/43) across both groups. Further, three patients in the NFE group (~19%) increased their antidepressant medication during the trial. Moreover, four patients in the NFE (25%) and one patient in the NFS group (~6%), either decreased or stopped their prescribed psychopharmacological medication during the trial.

4.3.3 Primary (registered) clinical outcome measure

Second-rater scores for the HDRS-17 were available for 12 sessions, with high interrater correlation (Pearson's correlation: $r = 0.95$, $p < 0.001$). There was no significant difference between groups on HDRS-17 (B = - 0.415, [95% CI -4.847 to 4.016], p = 0.848) at the primary endpoint (12 weeks). Pre-post comparisons of HDRS-17 (and all secondary and other outcome measures) across the three sessions (baseline, end of intervention and follow-up) are documented in Table 4.3. Depressive symptoms in both groups improved similarly on the HDRS-17 (NFE: -8.34 [95% CI -4.92 to -11.77]; NFS: - 8.34 [95% CI -5.81 to -10.87]; Figure 4.2A), such that the NFE group improved by 42% on the HDRS-17 and the NFS improved by 44%. Overall, the effect size for HDRS-17 was g = 1.46 [95% CI 0.97 to 1.95] at session 5 (NFE g = 1.23; [95% CI 0.63 to 1.92] vs. NFS g = 1.67 [95% CI 0.94 to 2.48]). Figure 4.2 B

shows the remission rates, which were overall 38% (12/32) with 25% for NFE (4/16) and 50% for NFS (8/16).

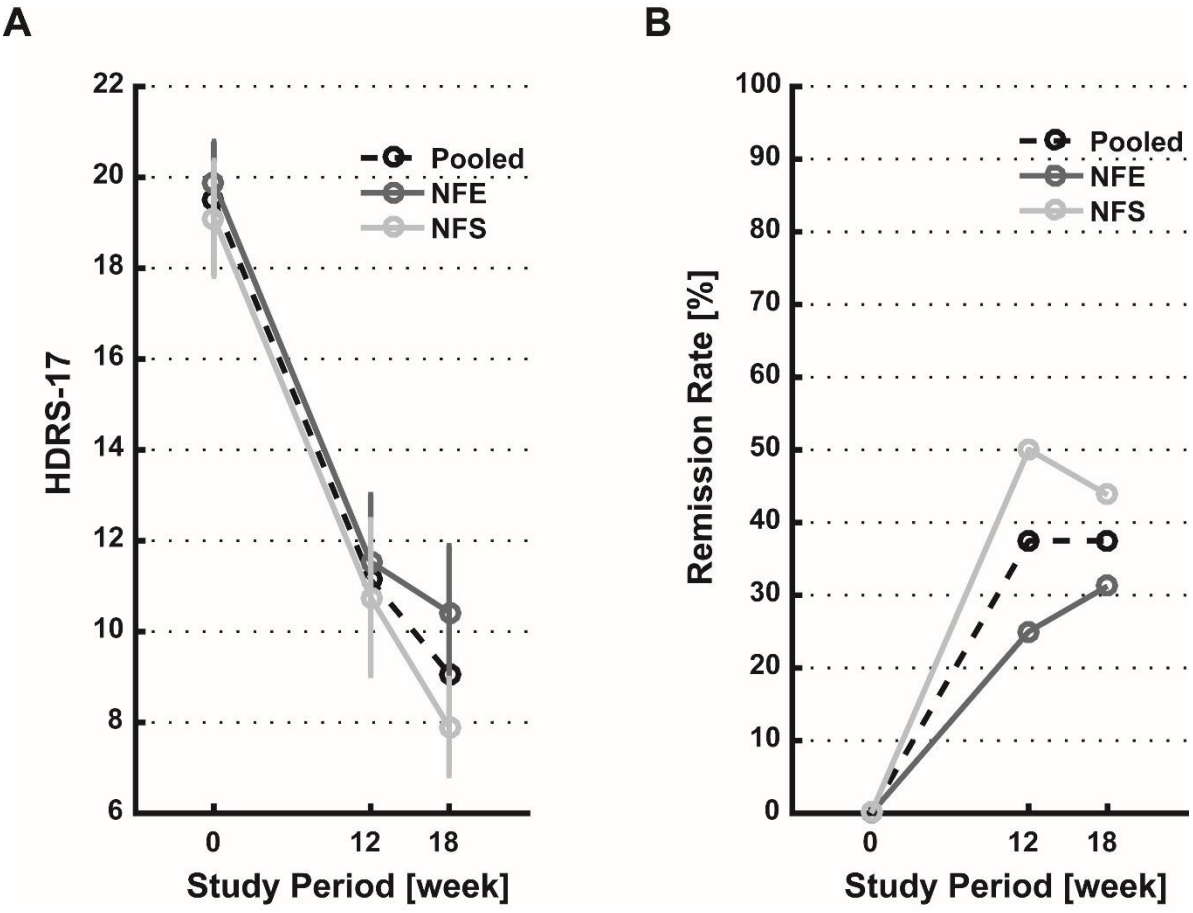

Figure 4.2 Change in primary outcome measure HDRS (A) and remission rate (B) for baseline, primary endpoint (12 weeks) and FU assessment (18 weeks). Adapted from (Mehler et al., 2018).

4.3.4 Secondary (registered) clinical outcome measures

Group differences for other clinical and psychological outcome measures are shown in Table 4.2. There was no significant difference between groups on the HDRS-17 (B = -1.121 [95% CI -4.948 to 2.706], p = 0.548) or any of the secondary clinical outcome measures at follow-up. Both groups improved compared to baseline on the HDRS-17 (NFE: -9.65 [95% CI -6.25 to -13.06]; NFS: -11.53 [95% CI -8.61 to -14.45]), such that scores decreased on average by about 48% in the NFE group, and 59% in the NFS group. The overall effect size for HDRS-17 improvement at follow-up was g = 1.88 [95% CI 1.24 to 2.54]; (NFE: g = 1.57 [95% CI 1.04 to 2.14] vs. NFS: g = 2.05 [95% CI 0.96 to 3.72]). Figure 4.2 B shows the remission rates at

follow-up, which remained at 38% on average across groups (31% remission in the NFE group and 44% remission in the NFS group). Pre-post differences were observed for multiple outcomes, a detailed list of descriptive statistics including effect sizes and their confidence intervals is documented in Table 4.3 for the primary endpoint (and Appendix A Table A1 for FU assessment).

Outcome measure	B	SE	t	95% CI	p value
HDRS-17 (week 12)	-0.415	2.147	-0.193	-4.847 to 4.016	0.848
	-1.121	1.835	-0.611	-4.948 to 2.706	0.548
HADA (A) (week 12)	-0.997	1.124	-0.887	-3.322 to 1.328	0.384
	-0.006	1.50	-0.004	-3.145 to 3.134	0.997
HADS (D) (week 12)	-0.860	1.631	-0.527	-4.235 to 2.515	0.603
	-0.865	1.843	-0.485	-2.963 to 4.753	0.633
BIS (week 12)	0.852	0.858	0.993	-0.924 to 2.627	0.331
	0.219	1.354	0.162	-2.615 to 3.053	0.873
BAS (week 12)	3.091	1.769	1.747	-0.569 to 6.751	0.094
	5.625	2.589	2.173	0.207 to 11.043	0.043
QoL (week 12)	1.510	3.438	0.439	-5.603 to 8.623	0.665
	2.635	4.0	0.659	-5.737 to 11.007	0.518
TCQ (week 12)	-2.617	3.301	-0.793	-9.447 to 4.212	0.436
	-4.273	4.051	-1.055	-12.752 to 4.207	0.305
TCAQ (week 12)	-5.532	4.515	-1.225	-14.871 to 3.807	0.233
	-1.396	4.750	-0.294	-11.337 to 8.545	0.772
SES (S) (week 12)	3.350	3.344	1.002	-3.568 to 10.268	0.327
	5.763	3.656	1.576	-1.888 to 13.415	0.131
SES (G) (week 12)	3.355	1.425	2.355	0.408 to 6.302	0.027
	4.206	1.814	2.319	0.410 to 8.001	0.032
EQ-5D-5L (week 12)	0.985	0.953	-1.034	-2.988 to 1.017	0.315
	-0.251	0.678	-0.370	-1.697 to 1.195	0.717

Table 4.2 Group mean difference estimates, based on a linear regression model that included minimisation factors used for randomisation as nuisance regressors, shown for primary endpoint (week 12) and follow-up (week 18, listed in row below). Adapted from (Mehler et al., 2018).

Outcome Measures	**Change from baseline to primary endpoint (12 weeks) for NFE / NFS**				
	M	95% CI	% -age	Hedge's g	95% CI
HDRS-17	-8.35/-8.34	-4.92 to -11.77/ -5.81 to -10.87	-42/ -44	-1.23 /-1.67	-1.92 to -0.63/ -2.48 to -0.94
HADS (A)	-1.81/ -3.19	0.07 to -3.69/ -1.44 to -4.93	-14/-25	-0.49 / -0.92	-.97 to 0.09/ -1.26 to -0.53
HADS (D)	-3.06/ -3.81	-0.87 to -5.25/ -1.35 to -6.28	-23/-31	-0.71 / -0.78	-1.09 to -0.24/ -1.71 to -0.01
BIS	0.75/ 0.94	-1.23 to 2.73/ -0.06 to 1.94	6/6	0.19 / 0.47	-0.33 to 0.61/ -0.04 to 0.95
BAS	-2.69/ 0.94	-5.71 to 0.33/ -1.09 to 2.96	-8/3	-0.45 / 0.23	-1.11 to 0.10/ -0.03 to 0.95
QoL	7.63/ 8.94	2.7 to 12.55/ 3.83 to 14.05	14/16	0.78 / 0.88	0.31 to 1.24/ 0.25 to 1.54
TCQ	5.25/ 2.19	-0.5 to 11/ -1.85 to 6.23	9/4	0.46 / 0.27	-0.07 to 0.79/ -0.22 to 0.74
TCAQ	9.19/ 3.75	1.58 to 16.79/ -3.03 to 10.53	17/7	0.61 / 0.28	0.04 to 1.15/ -0.25 to 0.86
SES (G)	3.56/ 5.81	-1.53 to 8.66/ 0.61 to 11.02	8/12	0.35/ 0.56	-0.13 to 0.86/ 0.13 to 1.05
SES (S)	0.69/ 3.94	-1.29 to 2.67/ 1.86 to 6.01	4/28	0.18/ 0.96	-0.33 to 0.64/ 0.25 to 1.64
EQ-5D-5L	-0.38/ -1.00	-2.25 to 1.48/ -2.18 to 0.18	-3/-10	-0.21/ -0.38	-0.78 to 0.45/ -0.38 to 0.80

Table 4.3 Changes over time in raw mean scores and standardised effect sizes both including 95% confidence interval estimates at primary endpoint assessment. Change in percentage (%-age). Adapted from (Mehler et al., 2018).

4.3.5 Other (registered) outcome measures

There were no significant group differences except for SES (S) at session 5 (B = 3.350 [95% CI 0.408 to 6.302], p = 0.027) and at FU (B = 4.206 [95% CI 0.410 to 8.001], p = 0.032), as well as for BAS at FU (B = 5.763 [95% CI 0.207 to 11.043], p = 0.043. However, after FDR correction for multiple testing, these differences did not remain significant.

With regards to pre-post changes of the POMS $_{TMD}$, no effect of group ($F_{1,23}$ = 0.357, p = 0.556, η^2_p = 0.015) was found. A Mauchly's test of sphericity (W = 0.393, p = 0.018) was significant for the within subject effect (session) and thus Greenhouse-Geisser correction applied. No significant effect of session was found ($F_{2.83,65.02}$ = 0.183, p =0.899, η^2_p = 0.008). Yet, a significant intercept ($F_{1,23}$ = 4.70, p = 0.041, η^2_p = 0.17) and interaction between session and baseline scores ($F_{2.83,65.02}$ = 5.569, p = 0.002, η^2_p = 0.195) was found. Taken together, data suggested that both groups showed comparable improvements on the POMS $_{TMD}$. Further, the significant interaction indicated that improvements over session covaried with differences in baseline POMS $_{TMD}$ between the first and last session.

With regards to pre-post changes of the subscale POMS $_{Depression}$, no group effect ($F_{1,24}$ = 1.065, p = 0.312, η^2_p = 0.042) or session effect ($F_{4,96}$ = 0.50, p = 0.736, η^2_p = 0.020) was found. Also, for this subscale a significant intercept ($F_{1,24}$ = 11.501, p = 0.002, η^2_p = 0.324) and interaction between session and baseline difference scores was found ($F_{4,96}$ = 15.854, p = 0.001, η^2_p = 0.398). Taken together, results suggested that both groups improved to a similar extent and that POMS $_{Depression}$ improvement decreased over sessions as a function of reduced POMS $_{Depression}$ baseline scores. Moreover, an effect of gender was found ($F_{1,24}$ = 4.360, p = 0.042, η^2_p = 0.154), which was confirmed by a corrected follow-up test (Mean difference = 3.405, 95% CI [0.04, 6.770], $F_{1, 24}$ = 4.360, $p_{Bonferroni}$ = 0.048, η^2_p = 0.154), suggesting that male patients (7.642, 95% CI [4.982, 10.301]) in the current sample showed overall larger improvements on the POMS $_{Depression}$ scale compared to female patients (4.327, 95% [2.288, 6.186]).

Frequentist paired samples t-tests indicated that both measures improved significantly (Table 4.4) and Bayesian tests suggested very strong evidence for such effect (POMS $_{Depression}$: BF_{10} = 12537.6; POMS $_{TMD}$: BF_{10} = 570.3).

						95% CI for Cohen's d	
		t	df	p	Cohen's d	Lower	Upper
pre-Depression $_{S1}$	- post-Depression $_{S5}$	5.988	30	< .001	1.076	0.626	1.514
pre-TMD $_{S1}$	- post-TMD $_{S5}$	4.794	30	< .001	0.861	0.442	1.269

Table 4.4 Paired samples t-Test between pre-test score at session 1 and post-test score at session 5 for POMS $_{Depression}$ and POMS $_{TMD}$, respectively.

Descriptive data indicated that both POMS $_{Depression}$ and POMS $_{TMD}$ scores decreased gradually over sessions (Figure 4.4 A and B), except from the last two sessions, which showed almost identical group mean pre-test scores. Further, data also indicated that group mean post-test scores were below pre-test scores of subsequent training sessions, suggesting that within-session improvement partly deteriorated until the next session.

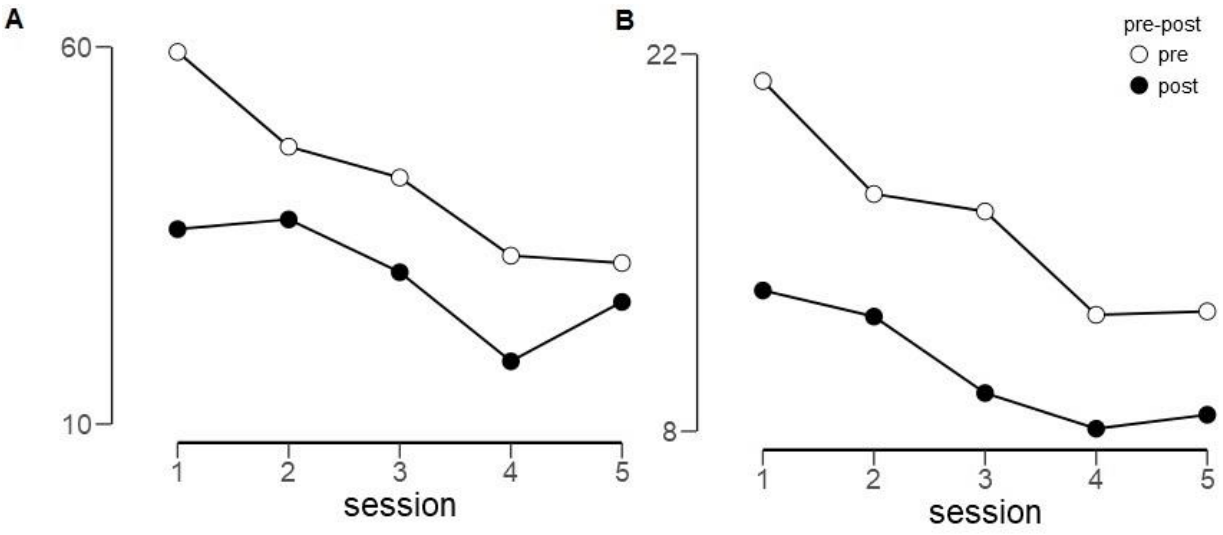

Figure 4.3 Group mean scores over sessions for pre-and post-tests for POMS $_{TMD}$ (A) and POMS $_{Depression}$ (B), respectively.

Descriptive data suggest that the largest within-session decrease was observed during the first and fourth session for POMS $_{TMD}$ scores (Table 4.5), and during the first and third session for POMS $_{Depression}$ scores (Table 4.6).

session	Mean	SD	N
1	23.367	32.42	30
2	10.700	27.52	30
3	8.967	25.09	30
4	14.067	23.14	30
5	5.367	21.50	30

Table 4.5 Descriptive statistics of within-session improvement for POMS $_{TMD}$.

sessions	Mean	SD	N
1	7.774	11.060	31
2	4.548	8.663	31
3	6.742	9.352	31
4	4.226	7.504	31
5	3.839	6.674	31

Table 4.6 Descriptive statistics of within-session improvement for POMS $_{Depression}$.

4.3.6 Exploratory analyses

4.3.6.1 Intention-to-treat analyses

Among non-completers, group allocation was balanced, with a proportion of five patients from NFE and 6 patients from the NFS group (0.455 vs. 0.545, p =1; BF_{01} = 2.707). The ANCOVA found no evidence that completion status explained baseline HDRS-17 scores ($F_{1, 35}$ = 0.361, p = 0.552) or duration of illness ($F_{1, 35}$ = 0.467, p = 0.499). Likewise, a frequentist independent t-test found no evidence for a difference in baseline HDRS-17 between completers and non-completers ($t_{14.72}$ = -0.486, p = 0.634, Cohen's d = -0.178), while a Bayes Factor remained inconclusive (BF_{10} = 0.808). Likewise, for duration of illness, no difference was found ($t_{28.82}$ = -0.302, p = 0.765, Cohen's d = -0.093) and Bayes Factor remained inconclusive (BF_{10} = 0.768). Taken together, the ITT suggested no significant differences between completers and non-completers with regards to group allocation, or key baseline characteristics. Bayesian test results provided anecdotal evidence for the absence of biases between completers and non-completers of the intervention.

4.3.6.2 Testing for null effects

Due to the lack of a group effect on the main clinical outcome measure HDRS-17, follow-up tests were conducted to test whether there was evidence for the absence of an effect. A Bayesian linear regression analysis was performed that controlled for randomisation factors. Moderate evidence was found for a null effect of group at both the primary endpoint (BF_{01} = 6.01) and at follow-up (BF_{01} = 3.51). Further, frequentist equivalence tests were conducted to test whether a SESOI could be rejected. However, these tests remained undetermined at both the primary endpoint ($t_{27.62}$ = -0.849; p = 0.202, raw score lower limit 90% CI = -3.40 and upper limit 90% CI = 3.40) and at follow-up ($t_{24.68}$ = - 0.104; p = 0.541, raw score lower limit 90% CI = -5.478 and upper limit 90% CI = 1.718), indicating that the current sample was too small to reject a clinical effect within the range of the SESOI. Lastly, based on the confidence intervals of the empirical effect size it was determined that only effect sizes larger than Cohen's d = |0.6| in either direction could have been rejected in the current sample. Taken together, the Bayesian analysis whose prior was scaled based on the effect size from the original power calculation suggested evidence for the absence of a group effect. In contrast, equivalence tests could not rule out group differences of the size that have been described for add-on treatment, but only effects that were at least twice as large. To summarise, the Bayesian ANCOVA provided evidence for the null model compared to a model that contained treatment group as a factor. This result indicated that the null model was preferred over the alternative model. Equivalence tests suggested that the present sample was too small to rule out group differences of the size that have been reported for anti-depressive medication as the gain in reduction of depression severity compared to mere placebo treatment (Cohen's d = 0.3).

4.3.6.3 Test against placebo effect reported for high-technology intervention in depression

A survey of placebo effects reported for high-technology based interventions in depression found that sham whole-body hyperthermia treatment was associated with an improvement of 5.5 points on the HDRS-17 (Janssen et al., 2016). A two-sided t-test against a test value of 6 found that our effect was significantly larger than that expected placebo response (t_{31} = 2.385, p = 0.023, Cohen's d = 0.422 [95% CI 0.057 to 0.780]).

4.3.6.4 Relationship between self-efficacy and HDRS-17

The exploratory robust regression analysis suggested that changes in self-efficacy predicted residualised depression scores at the primary endpoint (R^2 = 0.18, adjusted R^2 = 0.15, β = -

0.187 ± 0.073, Figure 4.3), such that increase in self-efficacy was associated with less depression severity ($t_{30} = -2.551$, $p = 0.016$).

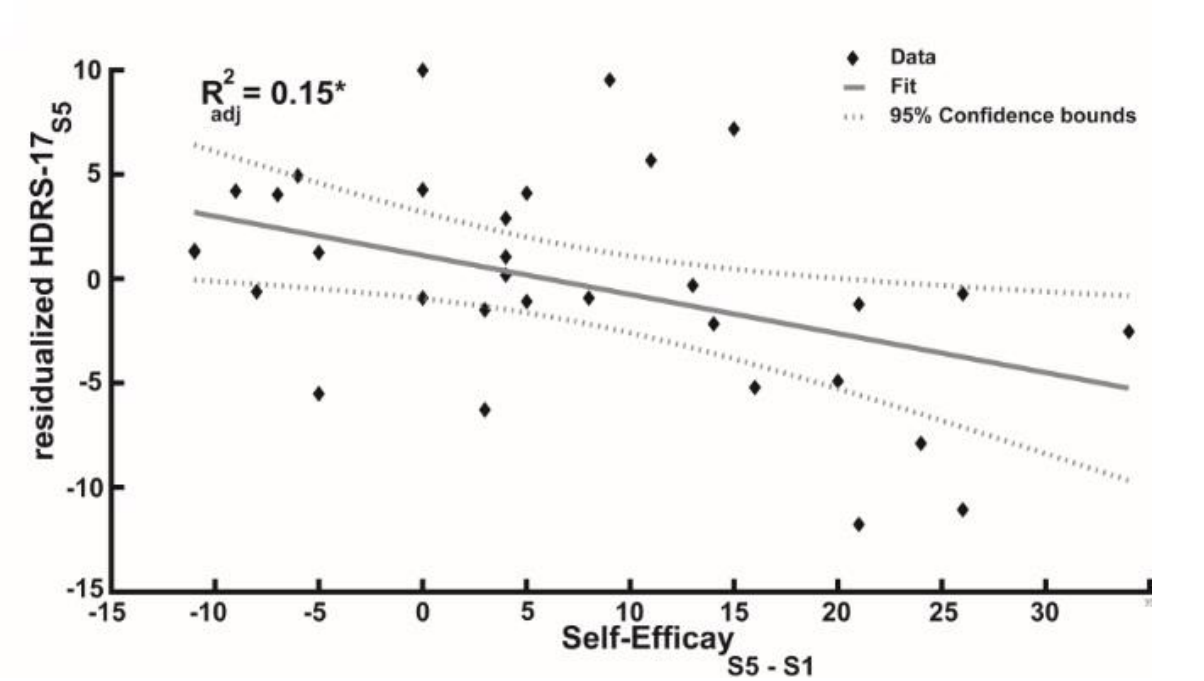

Figure 4.4 Robust regression analysis between change in total self-efficacy scores and residualised HDRS-17 scores at primary endpoint. *p<0.05. Adapted from (Mehler et al., 2018).

4.3.6.5 Changes in HDRS item scores

With regards to statistical tests for items that have been identified as indicators for side effects in response to psychopharmacological (SSRI) treatment, data indicates that scores improved significantly for gastrointestinal symptoms ($p_{FDR} = 0.023$) and genital symptoms ($p_{FDR} = 0.023$). Also scores for weight-loss improved ($p_{FDR} = 0.25$) but remained insignificant.

4.3.6.6 Relationship between POMS and HDRS-17

Both the TMD and Depression subscale of the POMS, showed a moderate correlation with the HDRS-17 at baseline (POMS $_{TMD}$ vs. HDRS-17: Spearman's rho = 0.353, p = 0.024; POMS $_{Depression}$ vs. HDRS-17: Spearman's rho = 0.306, p = 0.04). With regards to correlations between measures based on difference scores (i.e. changes over time), results were only significant for POMS $_{TMD}$ (Spearman's rho = 0.309, p = 0.045) and at trend level for POMS $_{Depression}$

(Spearman's rho = 0.266, p = 0.074). Taken together, baseline measures and pre-post changes on the POMS $_{TMD}$, and to a slightly lesser extent also on the POMS $_{Depression}$ subscale, reflected the primary outcome (HDRS-17).

4.4 Discussion

The current RCT demonstrated feasibility and clinical efficacy of fMRI neurofeedback training in moderate to severely depressed medicated patients with a longstanding history of clinical depression. Neurofeedback training of emotional areas was not superior to the control neurofeedback intervention. Nonetheless, both intervention groups showed a clinically significant improvement of the primary outcome measure with over 40% reduction in the HDRS-17. The overall remission rate was about 38%. Importantly, both measures were maintained, if not slightly improved at follow-up. Hence, this trial reproduced and extended clinical effects reported for the intervention group of a recent placebo controlled fMRI-NF RCT that was conducted in unmedicated patients (Young et al., 2017).

The absence of a group effect was supported by a Bayesian regression analysis. However, frequentist equivalence testing failed to reject the SESOI. Together with the non-significant ANCOVA, results thus remained inconclusive, indicating that the sample was likely too small to detect an effect of the prespecified size (Cohen's d = 0.3; Kirsch et al. 2008). Secondary and other outcome measures also suggest similar improvements in both groups. Overall, attrition rates were relatively low and comparable to other non-invasive intervention methods tested for MDD (Hollon et al., 2014; Janssen et al., 2016), suggesting that the technique was well accepted by patients. This interpretation is supported by substantial pre-post improvements found for both groups for secondary clinical outcome measures of anxiety and depression and reported positive effects on their quality of life. Further, besides changes on the HDRS-17 sum score, changes of item-based data indicated that symptoms related to gastrointestinal (GIT) and sexual dysfunction (GS) improved. In contrast, these symptoms represent side effects of psychopharmacological treatment and thus worsening of these symptoms is usually reported, in particular for increasing dosage or augmentation of existing medication (Hieronymus et al., 2016). Side-effects have been identified as one main reason for non-compliance and discontinuation (Demyttenaere, 1997; Johnson, 1981) and their reduction may thus not only contribute to patients' quality of life, but also to long term treatment adherence. Given the overall strong clinical response, fMRI-NF training may thus provide an attractive option as an add-on treatment in patients who remain depressed despite psychopharmacological treatment.

The current trial design with 5 training sessions allowed the assessment of immediate pre-post changes based on the POMS as a surrogate measure for depressive symptoms (Depression

subscale) and mood states (TMD). Indeed, a moderate correlation between both POMS scores and the HDRS-17 at baseline was found, and changes between baseline and primary endpoint correlated partly. These results suggest that changes in depression are captured to some extent by these secondary measures. It hence remains possible that the clinical improvement on the observed HDRS-17 was also gradual from session to session. POMS $_{TMD}$ and POMS $_{Depression}$ data indicated that both groups experienced immediate mood improvements to a similar extent, replicating and extending earlier findings (Linden et al., 2012). However, it should be noted that the POMS rather represent an implicit measure of mood and depression, in contrast to the HDRS-17, which is based on a structured, physician-led interview. Yet, although reported correlations are relatively small to moderate (around r = 0.3), they are largely comparable to some reports for more similar measures of depression severity such as the self-rated BDI and physician rated HDRS-17 (Schotte et al., 1997). Moreover, immediate changes partly lasted largely until the next feedback session, indicating that the impact of fMRI-NF on patients' mood was not merely a temporary effect. Yet, group mean post-test and pre-test scores of subsequent sessions indicated that within-session effects partly deteriorated. Moreover, group mean pre-test scores between the last two sessions were largely comparable, indicating that treatment effects on this implicit measure of mood and depression had levelled off. Notably, a gender difference was found for improvements on the POMS $_{TMD}$, suggesting that men responded more than women to this implicit measure of depressive symptoms and general mood state. Overall, these analyses show how POMS data can complement clinical assessments data because it captures immediate improvements in mood at individual sessions. Measures such as the HDRS-17 usually require longer temporal intervals (e.g. 2 weeks) and hence may not be suitable for weekly assessments (Geddes et al., 2012; Hamilton, 1980, 1960; Montgomery and Asberg, 1979) Of interest, data indicated that in particular clinical improvements of the fourth neurofeedback session were retained until the next feedback session. This finding may be of interest for two reasons: first, visual feedback was reintroduced during this session after it had been withdrawn from patients during the third transfer session. As such, the experience over sessions two to four can be seen as a slow cycle of intermittent feedback presentation, which may render the training experience more rewarding. Second, the fifth training session was on average one 3-4 weeks later than the other training sessions (Chapter 4.2.3), indicating that clinical changes were stable on group level after several weeks. Given that these analyses were declared as exploratory they require replication in larger samples.

A key question is why the active control neurofeedback group (NFS) also showed a substantial clinical improvement. The clinical efficacy of the NFS group may be explained by a similarly rewarding experience and the type of mental imagery training patients engaged in. Both groups up-regulated target ROIs to a similar extent, as shown by equivalence tests. Hence, participants in the NFS group also experienced success of self-regulation and they were positively reinforced for task success. Generic psychological or physiological factors could thus be important for any clinical efficacy of neurofeedback training in depression. We have earlier suggested that self-regulation experience during neurofeedback training may be associated with an increase of self-efficacy and explain therapeutic effects of fMRI-NF training in depression (Linden, 2014; Linden et al., 2012). This conjecture is supported by an exploratory robust regression analysis between changes in self-efficacy and residualised HDRS-17 scores (Figure 4.2 C). Self-efficacy has been shown to mediate vulnerability to develop depressive symptoms (Maciejewski et al., 2000), as well as the effect of depression on developing detrimental behaviour including nicotine abuse (Minnix et al., 2011). The reported changes in self-efficacy may thus yield secondary benefits including less vulnerability to relapse and enhancement of cognitive resources to cope with challenges (Bandura, 1982). However, larger trials of longer follow-up periods (Rance et al., 2018). Moreover, they should include transfer tasks that involve implicit and explicit measures of self-efficacy to test whether this psychological mechanism underlies observed therapeutic effects of fMRI-NF training. We further note that the task of NFS group – imagining relaxing visual scenes – may also have had a therapeutic effect. Relaxation therapies are used in clinical practice to treat depression (Bowers, 1990; Golding et al., 2017). A Cochrane review found that patients who completed relaxation interventions reported reduced depression severity compared to patients in wait-list or no treatment control groups (Jorm et al., 2008). Compared to patients in the wait-list condition, relaxation therapy showed moderate clinical effects for self-reported measures (SMD = -0.59, 95% CI [-0.94, -0.24]), and physician-rated depression scores showed non-significant differences in the same direction (SMD = -1.35 ([95% CI -3.06 to 0.37]). Taken together, our active control group likely received therapeutic self-regulation training and thus showed substantial improvement. It is hence possible that our control intervention was too conservative with respect to our aim of showing superiority of targeting the affective brain. Chapter 5 will present the neuroimaging data, providing insights into the neural correlates of the intervention that may provide further mechanistic explanations for the observed clinical effects.

Two (partly related) remaining questions are whether the reported clinical findings are specific, and whether they are comparable to effective active interventions or usual care. Because the current design did not include a sham-feedback or a care as usual control group, we will argue based on epidemiological data that the clinical effects reported here 1) exceed spontaneous remission rates, 2) are comparable to clinical effects reported for first line treatment as well as current best augmentation treatment alternatives, and 3) exceed placebo effects reported for sham controlled high-technology therapies.

First, the improvement of 40% on the HDRS-17 exceeds the minimal clinically relevant improvement of about 27% (Qids et al., 2003) and thus represents a substantial clinical effect. Findings also exceed expected regression to the mean effects for chronically depressed and partly treatment resistant patients (the average duration of illness was 19 years), for which epidemiological data only suggests a rather sluggish improvement of 10-15% (Keller et al., 1992). Further, given that patients in our trial had received stable medication for at least three months, the main response to pharmacological treatment had probably already occurred before enrolment (Rush et al., 2006).

Second, the effects reported here across both groups for the primary endpoint – about 8.3 points improvement and 37.5% [95% CI 22.9 to 54.8] remission on the HDRS-17 – are comparable to initial responses found for first line pharmacological treatment (8 to 9 points improvement on the HDRS-17 and 36.8% remission) in patients with depression of comparable severity and chronicity (Fournier et al., 2010; Rush et al., 2006). Indeed, the confidence interval around our remission rate excludes the estimated remission rate (22%) for placebo groups in first level pharmacotherapy treatment (Meister et al., 2017). However, these comparisons seem rather conservative given that patients received fMRI-NF as an augmentation, or even as a second augmentation treatment (one third of patients were already receiving combined pharmacological treatment), because augmentation treatments usually yield gradually diminishing clinical improvement. Notably, the presented clinical effects are similar and partly larger than effects reported for other accepted augmentation strategies: the ‘Sequenced Treatment Alternatives to Relieve Depression’ (STAR*D) trial reported 35% average remission for first augmentation, with about 36% for non-SSRIs and 29.4% for behavioural therapy, and remission rates of only 20.5% for secondary augmentation therapy (Rush et al., 2006).

Third, our clinical findings are also comparable to reported effects of other high-tech interventions for depression. We note that the size of non-specific treatment effects depends on patients' expectancies, beliefs, as well as the psychosocial context (Price et al., 2008), and non-specific treatment effects may thus be particularly large for fMRI-NF. However, a recent sham-controlled RCT of fMRI-NF of similar size in (unmedicated) depressed patients only found a marginal improvement of 2 points (HDRS-21) and 6% remission in a sham control group (Young et al., 2017). In contrast, the treatment group in their study that engaged in similar fMRI-NF training as our NFE group improved by about 11.5 points and showed 32% remission. Similarly, our results are comparable for treatment groups and outperform placebo groups in other sham-controlled high-technology interventions tested for depression: meta-analyses of transcranial magnetic stimulation (TMS) trials suggest about 35% remission for treatment groups, while sham TMS groups show only 5-10% remission (Berlim et al., 2014, 2013). Further, one whole body hyperthermia trial reported an improvement of 8.3 points (HDRS-17) and a remission rate of 40% remission for their treatment group, while sham treatment only led to a placebo response of 5.5 points and 0% remission (Janssen et al., 2016). Importantly, we could reject a large placebo response of 6 points in an exploratory analysis, thereby showing the superiority of our clinical findings. Altogether, this survey of the size of expected placebo responses suggests that the clinical effects found in our trial exceed a mere regression to the mean or placebo effect. Rather, our findings are comparable to treatment effects reported for accepted augmentation therapies and other high-technology therapies currently being investigated for depression.

Besides control for placebo effects through effective control conditions, blinding of participants, assessors and intervention represents an important design aspect of RCTs for reducing biases (Chapter 3.4.3). In the current trial, assessors were blinded to group assignment. However, they were not blinded to the fact that the intervention was a fMRI-NF trial and overall clinical findings may be overestimated. Hence, to address potential criticisms (Thibault et al., 2017a; Thibault et al., 2018b, 2018a), future trials could blind assessors to the general intervention that is being conducted. Blinding of patients was very limited in the current trial because both groups received veridical feedback from target regions that were matched to mental imagery instructions. As a result, both groups likely had a similar experience in terms of neurofeedback success, which formed the main rationale behind the choice of the active control group. Self-regulation effects between groups will be explored in Chapter 5 based on the neuroimaging data. Lastly, the researchers who provided the fMRI-NF training were not

blinded to the assignment, which was not feasible in the current setup because they operated the scanner including the localiser scan, chose target regions, and instructed patients with regards to the mental imagery task. Nevertheless, this circumstance cannot explain why no group differences were observed. If the lack of blinding on the intervention confounded results, rather the opposite (an increase in group differences) would be expected. Overall, the choice of control group remains a crucial aspect in trial designs of fMRI-NF interventions to demonstrate specificity of observed clinical effects. For instance, a randomised no-treatment control group captures effects such as regression to the mean, spontaneous remission and delayed effects of pharmacotherapy. However, a no-treatment group cannot control for therapeutic effects that may result from mere mental imagery training, or for various other psychological and neurobiological factors that likely play a role in complex interventions such as fMRI-NF training (Thibault et al., 2016; Thibault and Raz, 2016). Further, the choice of the control intervention will also determine the difference that one may hope to detect between groups, and thus the sample size needed to achieve sufficient statistical power (Chapter 3.4.1.2). Potential control designs for future fMRI-NF trials and associated methodological challenges will be reviewed in the general discussion (Chapter 9).

In conclusion, the present trial did not find statistically larger effects for self-regulation training of areas involved in processing positive emotions compared to training of areas that process merely higher visual information. Hence, in contrast to the initial hypothesis, no superiority of emotion-based fMRI-NF training was found. Importantly, both groups showed a clinical improvement that was comparable to effects reported for accepted augmentation therapy and that exceeded expected regression to the mean and commonly observed placebo effects. Hence, this trial demonstrated that if moderately to severely depressed patients can up-regulate a brain area that is not immediately linked to affective processing (NFS group), they can experience similar clinical benefits compared to patients that completed emotion-focused neurofeedback training (NFE group). Exploratory analyses suggested that the experience of brain control and the positive reinforcement of mental imagery may be necessary components for therapeutic effects of neurofeedback and should therefore be considered for future designs. Overall, neurofeedback was well tolerated, but it remains advisable for clinical neurofeedback protocols to be clinically supervised and to monitor mental strategies and psychological effects. This will allow patients to discuss their individual experience so training protocols can be adjusted if needed. Further, data suggest that further exploration of the clinical efficacy of real-time fMRI neurofeedback protocols and neural as well as psychological mechanisms are needed. To

understand underlying mechanisms, further exploration of the clinical efficacy fMRI-NF and its relation to neural as well as psychological mechanisms is needed (Chapter 5).

5 Neuroimaging Correlates of Neurofeedback Training in Depression

5.1 Introduction

Proof-of-concept (PoC) studies have shown that patients who suffered from mild to moderate depression can learn to up-regulate brain areas that were localised with positive affective stimulation, such as the anterior insula, lateral prefrontal areas, and the amygdala (Linden et al., 2012; Young et al., 2014). A recent double-blind RCT corroborated these findings, showing that unmedicated, moderately depressed patients learned to up-regulate the amygdala using positive mental imagery, while a control group that trained in up-regulating the intraparietal sulcus did not activate the amygdala (Young et al., 2017). Moreover, results suggested that the control group did not succeed in up-regulating the control area either until the last neurofeedback run, when relatively small activation was achieved (about 1/5 of the activation that was achieved in the amygdala training group. Clinical results showed that a) only patients who trained in up-regulating the amygdala showed significant clinical improvements on standardised rating scales such as the HDRS-17, and further that b) these changes were superior to changes in the control group. However, given that patients in their control group did not attain self-regulation success, groups were not matched for the reward experience. Therefore, it remains unclear whether the reported clinical superiority of amygdala fMRI-NF training is specific to feedback from emotion-regulating regions, or whether the general experience of self-regulation during fMRI-NF may be in itself therapeutic. Moreover, it remains to be tested if similar findings can be replicated in a sample of medicated patients. fMRI-NF training is designed to reinforce therapeutic cognitive strategies used in CBT, (e.g. positive mental imagery and reappraisal) in a reward learning based paradigm. Therefore, psychometric measures that are relevant to depression – such as self-efficacy – are of particular interest to explore potential predictors of neurofeedback success and mechanisms that may mediate therapeutic effects (Chapter 4.3.6.5). The present RCT included medicated patients with moderate to severe depression and compared up-regulation of brain areas involved in emotion processing to an active control procedure that involved up-regulation of brain areas involved in higher visual processes. This design gave patients in the control group the opportunity to experience self-regulation success and reward experience as well. It was hypothesised that neurofeedback of areas involved in processing positive emotions would produce clinical improvements exceeding those seen in the active control group. Lastly, relation between up-regulation success, clinical improvement and changes in psychometric measures of self-efficacy were explored.

5.2 Methods

5.2.1 Participants

Patients who had completed the intervention at the primary endpoint (session 5) were included in neuroimaging analyses. In total, 32 patients were counted as completers, with 16 patients respectively in the NFE and NFS group. Participants were matched for age, gender, depression severity, duration of illness, medication, and handedness (see Table 4.1).

5.2.2 MR image acquisition

A 3 Tesla whole body MRI system (General Electric, Milwaukee, USA) with an 8-channel head coil was used at the Cardiff University Brain Research Imaging Centre (CUBRIC). An echo planar imaging (EPI) sequence (TR = 2s, TE = 45ms, flip angle =80º, 30 slices, field of view = 192 mm, image matrix 64 x 64, in plane voxel size = 3 mm x 3mm, slice thickness = 4 mm, gap of 1 mm) was used for fMRI, and a high-resolution 3D T1-weighted image (TR 7.9s, TE 3.0ms, TI 450ms, flip angle 20°, matrix size 256 x 256, 1 mm isotropic voxel resolution) was acquired for anatomical co-registration.

5.2.3 Physiological recordings

To control for physiological confounding factors of the BOLD signal (Heunis et al., 2018; Murphy et al., 2013), heart rate (HR) and respiration volume per time (RVT) were measured using pulse oximetry and a respiratory belt, respectively, and recorded with Spike2 (version 5.21, Cambridge Electronics Design Limited, Cambridge, UK). In-house MATLAB (Mathworks Inc) scripts compiled by Dr Kevin Murphy and modified by David Mehler at CUBRIC (Cardiff University Brain Research Imaging Centre) were used for quality control and to compute the average heart rate (HR) and respiration volume per time (RVT) during each obtained volume.

5.2.4 fMRI-NF setup

Acquired EPI data was submitted to real-time motion correction and spatial smoothing (FWHM 4mm). For the localiser scan, real-time statistical analyses were carried out via an incremental general linear model (GLM) using Turbo-BrainVoyager (TBV) (Version 3.0, Brain innovation, Maastricht, The Netherlands). Target regions of interest (ROIs) in the respective groups were identified during a localiser scan based on the t-statistic of the contrasts of interest, which were defined as positive vs. neutral pictures in the NFE group and scene vs.

face pictures in the NFS group. Target ROIs in the NFE group were limited to limbic and frontal portions of the anterior cerebrum based on models of emotion processing in the human brain (Phillips et al., 2008). This focus also helped to exclude areas involved in early visual processing. For the localiser scans of the NFE group, we used a previously described procedure (Linden et al., 2012) using pictures rated as positive, negative and neutral from the International Affective Pictures System (IAPS) (Mikels et al., 2005). We aimed for a control condition that would entail similar up-regulation success – and thus reward experience – in both groups. Thus, the NFS group was presented with visual stimuli showing faces, scenes and animals to localise higher visual brain areas. Patients in both groups were presented with four series of four pictures (1.5s each) per category in pseudorandom order with alternating presentation and fixation baseline blocks. The parahippocampal place area (PPA) was chosen as the main target region in the NFS group. The PPA has been identified with the presentation of neutral scene stimuli (Epstein et al., 2003; Epstein and Kanwisher, 1998), but it is also activated during mental imagery of scenes (O'Craven and Kanwisher, 2000), and can be up-regulated with fMRI-NF (Habes et al., 2016).

To identify target ROIs for neurofeedback training, each session began with a functional localiser (325 volumes, first 6 volumes discarded to ensure T1 equilibrium magnetization) scan, which was followed by six neurofeedback scans (100 volumes, first 6 volumes discarded to ensure T1 equilibrium magnetization). Neurofeedback scans contained four 20s up-regulation blocks alternating with 20s rest blocks (the first rest block was 40s long). Neurofeedback was provided with a visual thermometer display projected onto the screen in the scanner and provided continuously (i.e. updated with each volume) as described previously (Linden et al., 2012). Real-time statistical analyses were carried using Turbo-BrainVoyager (TBV) (Version 3.0, Brain Innovation, Maastricht, The Netherlands). The signal intensity of target areas was measured as the percent signal change (PSC) relative to baseline. Neurofeedback was provided based on the Percent Signal Change of the target ROI and presented using PsychoPy (Peirce, 2007). Specifically, the feedback was calculated as the running average of the pre-processed BOLD signal over the three preceding TRs, normalized by the mean BOLD signal of the last five TRs from the preceding baseline block. The maximum feedback signal was set to 1 percent signal change, the thermometer contained ten bars such that each bar represented 0.1 percent signal change. The thermometer was continuously updated and shown during the neurofeedback and rest period. The third session served as a transfer session during which neurofeedback was not provided to test if patients

could up-regulate target ROIs based on successful strategies learned during the previous two neurofeedback sessions. The last neurofeedback session occurred approximately four weeks after the fourth session to test if patients retained similar up-regulation success.

5.2.5 Mental imagery instructions

Both groups were instructed to increase activation in their target areas during the up-regulation periods. Patients were not restricted in their mental strategies and could use any strategy that would enable them to achieve this. They did not receive specific instructions but were informed that using imagery of positive stimuli (NFE) or imagery of scenes (NFS) might be a potential starting strategy. During rest periods patients were instructed to engage in simple mental calculations (counting backwards from 99 in steps of 3). Patients were made aware of the temporal delay (6s) associated with the visual feedback on the thermometer, resulting from the nature of the BOLD signal.

5.2.6 Offline fMRI analysis

The neuroimaging data was analysed offline using Brain Voyager QX 2.8.4 (Brain Innovation, Maastricht, the Netherlands) and custom-written MATLAB scripts (Version 8.2, The MathWorks Inc., Natick, MA, 2000). T1 images were inhomogeneity corrected and normalized to Talairach space. EPI data was slice time corrected. To correct for head movements, spatial alignment of all volumes to the first volume of the localiser was performed. The fMRI data was co-registered to T1 images based on anatomical landmark points, and co-registration was manually corrected if necessary. The data was normalized to Talairach space using rigid body transformation and scaling. Linear drifts were removed and a temporal high-pass filter (2 cycles) was applied to remove non-linear drifts in the time series. Data were spatially smoothed using a Gaussian kernel with 6-mm full width at half maximum and temporally smoothed with a Gaussian kernel with 3s full width at half maximum. Motion parameters and physiological measures (heart rate, respiration volume per time) were z-transformed and included as nuisance regressors in the General Linear Model. To quantify the homogeneity of target areas within and distinctness of target areas between the groups' probability maps were calculated based on the Talairach coordinates of chosen target areas.

5.2.7 Comparisons between size of target regions

As an implementation check the sizes of target regions between groups was compared. Given the small sample and hence low power to detect potential interactions, the number of voxels of

data regions from all sessions were plotted for visual inspection using boxplots that indicated the median, interquartile range, and potential outliers. To test if the size of target regions differed between sessions or groups during neurofeedback, the number of selected voxels from all neurofeedback sessions were submitted to a 4 (session) x 2 (group) mixed ANOVA.

5.2.8 Region of interest analysis

ROI GLM analyses included time courses of six motion head motion parameters (three translation and rotation, respectively), and two physiology parameters (RVT and HR) as nuisance regressors. To correct for serial correlations, ROI analyses were performed with a 2nd order temporal autoregressive model (AR2), using Brain Voyager QX 2.8.4 (Brain Innovation, Maastricht, the Netherlands). AR2 models have been shown to outperform AR1 models in reducing biased error terms, and hence yield more reliable t-statistics (Lenoski et al., 2008). To test whether patients successfully up-regulated target areas within a session, ROI analyses of patient's target regions were performed for each session. Obtained t-test values of ROI analyses were submitted to two-sided t-tests for each session to test for activation on group level. P-values were adjusted based on the false discovery rate (FDR) to correct for multiple testing (Benjamini and Yekutieli, 2001). To test for effects of time, t-values were also submitted to a repeated mixed 2 (group) by 5 (session) ANOVA with the between factor group (NFE vs. NFS) and within factor session. Since the visual feedback provided to patients was based on the PSC during the task relative to rest, PSC values per session were also submitted to t-tests. This data indicated whether patients had on average managed to fill the thermometer display. For this manipulation check test statistics were not corrected for multiple testing as they were also not during real-time feedback. Further, inspection of these PSC values allowed a numerical comparison with the PSC value that was used to calibrate/scale the visual feedback (PSC = 1) and check whether ceiling effects had potentially occurred (i.e. if patients' PSCs were on average above this value, it would indicate that feedback was likely ineffective because patients would not have been informed about any signal change occurring above this cut-off).

5.2.9 Exploratory analyses

5.2.9.1 Equivalence test of target region up-regulation

Given the absence of a group effect in up-regulation and the similar clinical effects observed in both groups we also conducted equivalence tests (Lakens et al., 2018) based on participants' median ROI t values across the four neurofeedback training sessions. The test was conducted

with a Welch t-test (corrected for unequal variances) and with a SESOI = -0.7 to 0.7 [with raw score lower limit 90% CI = -1.799 and upper limit 90% CI = 1.799].

5.2.9.2 Self-regulation success of target region and clinical improvement (HDRS-17)

Two neurofeedback success measures were explored with regards to their relation to the main clinical outcome measure, the improvement on the HDRS-17 between baseline and session 5: 1) the average ROI up-regulation measured by the median t-value over all sessions served as a measure of centrality for the brain activation that participants had achieved. 2) the difference in t-values between session 5 and 1 served as a proxy for individual learning effects. Although no effect of time was found with regards to up-regulation, it remained possible that some patients increased up-regulation over time. Spearman's correlations were computed between each measure and HDRS-17 between sessions 1 and 5, respectively. In contrast to Pearson's correlation, which assumes linearity, Spearman's rho estimates monotonic relationships irrespective of linearity. Such relaxation of assumptions seemed adequate given the long duration of the trial and the logarithmic shape of response functions documented for psychopharmacological (Keller et al., 1992), psychotherapeutic treatment, and their combination in depression (Team, 2007). Results of Pearson's correlations were also reported for comparison.

5.2.9.3 Other (registered) neuroimaging related outcome measures

In the registration protocol, neuroimaging parameters and follow-up analyses were not further specified. The protocol states that the relation between self-regulation success and certain psychometric measures (SES, TCQ, TCAQ, BIS/BAS) will be analysed. However, the exact analysis plan and success measure for self-regulation, were not specified, leaving many degrees of freedom in analysing the data. Indeed, exploring the relation between changes in self-regulation and measures of self-efficacy was of particular interest given the earlier reported relationship between pre and post differences in self-efficacy and clinical improvement (3.3.7.4). Yet, to limit multiple testing and thereby preserve statistical power, two measures of neurofeedback success were first explored with regards to their relationship to the main clinical outcome (HDRS-17). To minimise multiple testing, this contrast was then subsequently entered in a multiple linear regression model, which allows interpretation of regression weights. Regressors were first tested for multi-collinearity and then entered as predictors.

5.2.9.4 Whole brain analysis

To identify task specific brain activity patterns, whole brain analyses were computed offline for both groups separately using a random effect General Linear Model and a task versus baseline contrast was defined. To test for effects of time, group and their interaction, an ANCOVA random effect General Linear Model was calculated and contrasts for intervention group (NFE vs. NFS) and time (linear contrast over all five sessions) tested. To limit multiple comparison correction, a mask was used that covered the cerebrum and upper brain stem and upper cerebellum. Active brain regions were identified based on the nearest grey matter coordinates using the Talairach Daemon atlas tool (Lancaster et al., 2000). Several contrasts were computed: 1) To test for task effects in the two groups, separate task vs. rest contrasts were computed. 2) To test for group differences, a contrast NFS>NFS was computed. 3) To explore effects of time, a linear time contrast across both groups S5>S1 was computed. All statistical parametric maps were multiple comparison-corrected using the Monte Carlo simulation implemented in Brain Voyager and recommended statistical thresholds (primary threshold $p < 0.001$; cluster-extent threshold $p < 0.05$) (Eklund et al., 2016). For the between-group contrast this primary threshold corresponded to FDR correction ($p < 0.05$). The time contrast was corrected at a more liberal threshold to increase sensitivity (primary threshold $p < 0.01$; cluster-extent threshold $p < 0.05$) (Eklund et al., 2016).

5.2.9.5 Relation between changes POMS score and up-regulation

POMS scores were assessed before and after each neurofeedback session and thus provided a psychometric measure that captured immediate changes in mood. Previous studies have tested for changes in POMS scores before and after fMRI-NF training (Linden et al., 2012; Young et al., 2014), although the relation between changes in POMS scores and neurofeedback success, defined as activation strength for up-regulation paradigms, remains to be explored. Specifically, two research hypotheses were of interest: 1) Is there a positive relationship between up-regulation and pre-post changes in mood during neurofeedback sessions, and 2) is there a positive relation between target region activation when no feedback is provided, and the respective pre-post changes in mood during the transfer session 3. POMS $_{TMD}$ pre-post difference scores were correlated with target region t-values across both groups at every session using Pearson's correlation. Specifically, a family of tests was carried out for all neurofeedback

sessions (session 1, 2, 4, and 5), and a separate analysis was carried out for the transfer session (session 3). Although related, both hypotheses were qualitatively different and hence multiple testing was adjusted accordingly: For the case of significant test results for the first set of tests (for session 1, 2, 4 or 5) or for both the first and the second set of tests, p-values were adjusted based on all five tests that were carried out using FDR adjustment. For the case that only the second hypothesis test returned a significant result, Bonferroni correction was applied based on the number of tested sets of hypotheses (N = 2, adjusted alpha level = 0.025).

5.2.9.6 Relationship between changes in insula activation, self-efficacy and HDRS-17

Lastly, the relation between baseline measures and changes in activity in the insular cortex were also explored regarding their relationship with clinical response (HDRS-17) and changes in self-efficacy, respectively. The (left) anterior insular cortex has been identified as a key region that is involved in neurofeedback training across paradigms (Emmert et al., 2016) and involved in reward processing during the training experience (Sitaram et al., 2017). Furthermore, the right anterior insular cortex has shown promise as a biomarker to predict response to CBT and pharmacotherapy treatment (Dunlop and Mayberg, 2014). Hence, exploratory ROI analyses were carried out for the left [-27, 21, -3] and right [27, 21, -3] insula (using a sphere of 443 voxel, respectively, around the coordinates). Two-sided exploratory correlation analyses were carried out based on differences of t-values. First, it was tested whether changes in self-efficacy correlated with changes in insula activity. Second, insula activity during the first session was correlated with clinical change to test whether it could serve as a potential predictor for treatment response. All analyses were carried for both the left and right insula. In addition to frequentist correlation analyses, also Bayesian analyses with a default prior were carried out.

5.3 Results

5.3.1 Target region size

Data of target regions was first explored for potential differences between sessions or groups and submitted to mixed model ANOVA. No significant effect of session ($F_{3,90}$ = 1.844, p = 0.149), or group ($F_{1,30}$ = 2.479, p = 0.126), or interaction of session and group ($F_{3,90}$ = 1.539, p = 0.210) was found.

Further, visual inspection of box plots (Figure 5.1) suggested that the number of voxels was similar overall between groups and sessions. However, a closer inspection suggested that between session 1 and 2 of the NFE group the number of voxels appeared to differ significantly, as indicated by no overlap of the indentation of the boxplot during session1 and the upper quartile at session 2 (Figure 5.1 A). Moreover, the interquartile range during the first session appeared to be larger for the NFE group compared to the following sessions, as well as compared to the data of the NFS group (Figure 5.1 B).

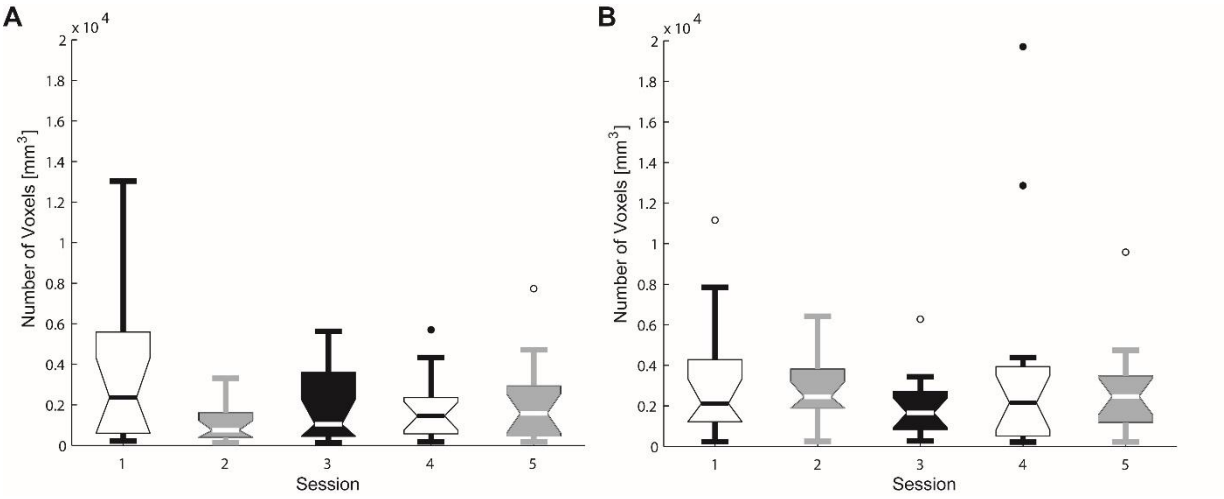

Figure 5.1 Boxplots of number of voxels for A) NFE and B) NFS group.

Regarding the spread of target region locations, Figure 5.2 shows a probability map of the main target region loci for the NFE and NFS group, respectively, collapsed across patients from both groups and all training sessions. A whole brain overview of target region selection is provided in Appendix B (Figure B1). Patients in the NFE group received fMRI-NF training mainly from anterior brain areas (e.g. insular cortex and striatum), which are in line with a set of regions that has been suggested as suitable target areas for interventions in depression (Linden, 2014). Patients in the NFS group received fMRI-NF training mainly from the PPA, which constituted

the primary target region of this group. Taken together, these results indicate that anatomical specificity of target region selection was largely maintained. Where patients in the NFE group trained mainly on areas related to processing emotions, patients in the NFS group mainly trained on higher areas of the visual system related to processing visual scenes. Further, these results indicate that target regions in the NFE group were more heterogenous compared to the NFS group.

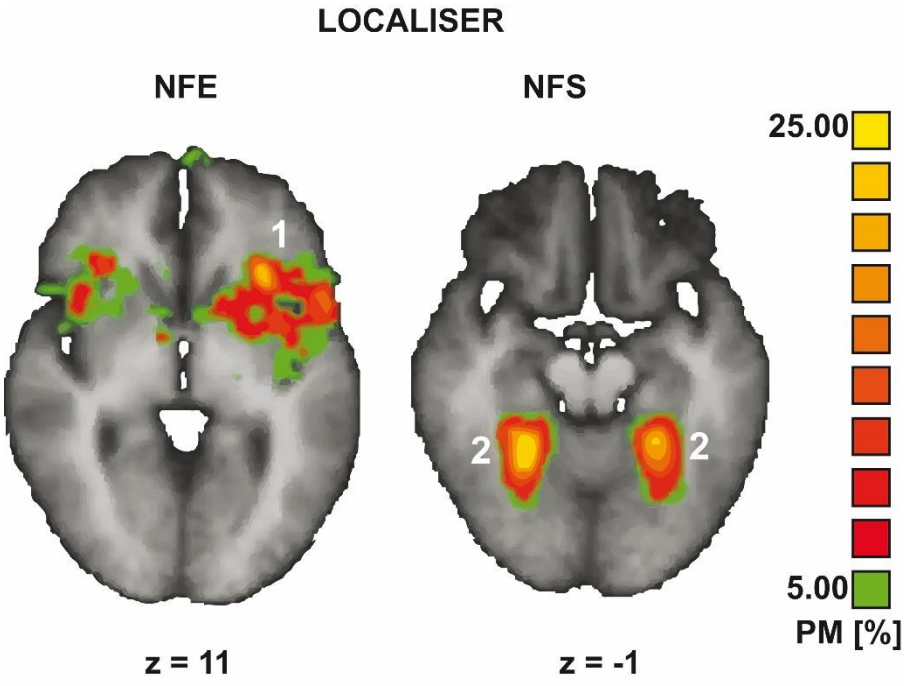

Figure 5.2 Probability map of target region placement centred around peak for NFE (left) and NFS (right) group. Adapted from (Mehler et al., 2018).

5.3.2 Region of interest analyses

As a first outcome variable the ability to up-regulate target regions (ROIs) during mental imagery compared to rest was assessed for both groups. T-tests indicated that patients up-regulated ROIs during all neurofeedback sessions (session 1: $t_{31} = 4.723$; $p_{FDR} < 0.002$, session 2: $t_{31} = 2.772$, $p_{FDR} = 0.011$; session 4: $t_{31} = 3.726$, $p_{FDR} = 0.002$; session 5: $t_{31} = 3.809$, $p_{FDR} = 0.002$; Figure 5.3), but not during the transfer session (session 3: $t_{31} = 1.404$; $p_{FDR} = 0.170$). Further, the difference between the average up-regulation across neurofeedback sessions and the transfer session was significant ($t_{31} = 2.397$, $p = 0.023$), indicating that providing feedback increased brain activation. As a manipulation check, PSCs were also computed and found positive for all sessions (session 1: 0.131 ± 0.033 [mean ± SEM], $t_{31} = 3.995$, $p = 0.001$; session 4: 0.199 ± 0.075, $t_{31} = 2.647$, $p = 0.013$; session 5: 0.163 ± 0.047, $t_{31} = 3.478$, $p = 0.002$), except

for the second (0.075 ± 0.040, t_{31} = 1.87, p = 0.071) and the transfer session (0.056 ± 0.047, t_{31} = 1.2, p = 0.239). These group average PSC values are largely in line with previous work (Young et al., 2017) and are below the calibration value (PSC = 1) that was used calibrate/scale the feedback display, suggesting that performance ceiling had likely not occurred.

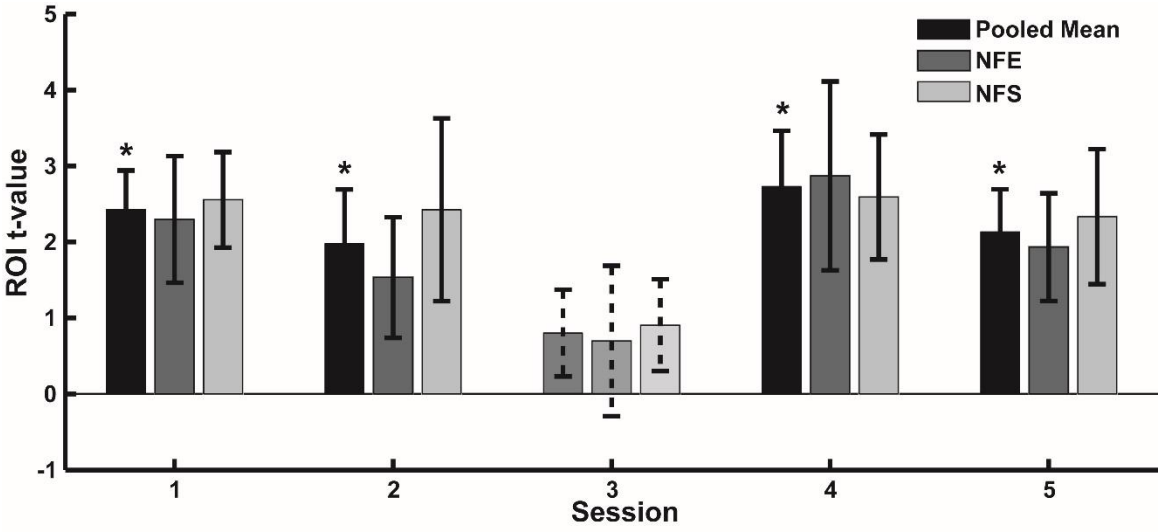

Figure 5.3 ROI analysis. Average t-value of target ROIs for each session. Pooled mean shown, as well as means for each group with error bars showing standard error of the mean (SEM). *) p < 0.05. Adapted from (Mehler et al., 2018).

To test for an effect of time, group or an interaction, ROI t-values were submitted to a mixed model ANOVA. Effects for session ($F_{4, 120}$ = 1.965, p = 0.104), group ($F_{1, 30}$ = 0.12, p = 0.73), and a group*session interaction ($F_{4, 120}$ = 0.160, p = 0.958) were found insignificant. This suggested that patients did probably not learn to increase activation over sessions.

5.3.3 Exploratory analyses

5.3.3.1 Equivalence test of target region up-regulation

ROI analyses suggested that both groups attained up-regulation to a similar degree. However, the lack of significance is an insufficient measure for the lack of an effect. Therefore, an equivalence test was performed to test if a smallest effect size of interest (SESOI) could be rejected as a group difference. Equivalence test results indicated that both groups indeed activated ROIs to a similar extent ($t_{29.8}$ = 1.702, p = 0.049, raw score lower limit 90% CI = -1.795 and upper limit 90% CI = 1.290).

5.3.3.2 Self-regulation success of target region and clinical improvement (HDRS-17)

No relationship between the average up-regulation success and clinical improvement was found on the HDRS-17 (Spearman's rho = - 0.08, p = 0.663). A positive correlation at trend level was found (Spearman's rho = 0.29, p = 0.051) between the change in up-regulation success from S1 to S5 and clinical improvement (Figure 5.4). A follow-up analysis split by group suggested that this correlation was mainly driven by the NFE group (NFE group: Spearman's rho = 0.58, p = 0.01; NFS group: Spearman's rho = - 0.11, p = 0.661). This interpretation was supported by a significant difference between correlations (z = 1.97, p = 0.024, one-tailed).

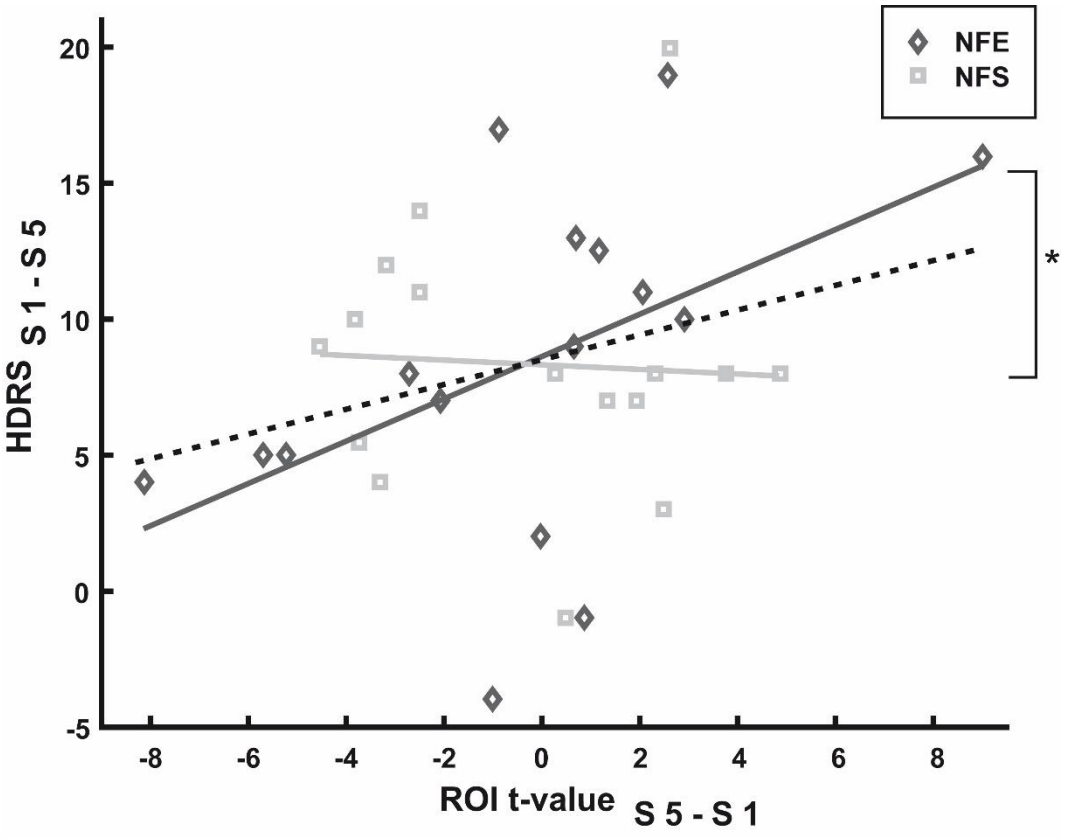

Figure 5.4 Correlation analysis between change of up-regulation (t-value) and change in clinical scores (HDRS-17) between baseline and primary endpoint. Split by group, black dotted line shows sample average. * p < 0.05

For comparison, analyses were also carried out using Pearson's correlation. Results were largely comparable for pooled data (r = 0.289, p = 0.109) and for individual groups (NFE: r =

0.485, p = 0.057; NFS: r = -0.055, p = 0.839), although differences in correlations between groups were slightly smaller and only at trend level (z = 1.49, p = 0.068).

Nevertheless, it was possible that correlations for the NFE group were driven by an overlap between target regions (Figure 5.2) and areas that are involved in self-regulation and reward processing such as prefrontal structures, parts of the basal ganglia and the insula (Emmert et al., 2016; Sitaram et al., 2017). Therefore, a post-hoc negative control analysis for the NFS was carried out using an ROI mask that was created based on the target regions of the NFE group. Specifically, it was tested whether a similar relationship between changes in activation and clinical scores could be reproduced in the NFS group. To do so, ROI analyses were carried out for the NFS group for sessions 1 and 5 and obtained t-values correlated with difference scores of the HDRS-17. Notably, this negative control analysis found no relationship (rho = -0.27, p = 0.312), supporting the specificity of the reported correlation for the NFE group. It is thus unlikely that the relationship reported for the NFE group stems mainly from the stimulation of positive feedback since both groups were largely matched for self-regulation success and thus reward.

5.3.3.3 Other (registered) neuroimaging related outcome measures

To limit the researcher's degree of freedom and hence multiple testing correction, this analysis was partly contingent on the previous analysis. In a previous analysis, change in up-regulation success between sessions 1 and 5 was identified as a correlate of clinical improvement (Figure 5.4). Thus, difference scores between these sessions for self-efficacy scale measures and neuroimaging data (differences of ROI t-values between session 1 and 5) were also used to explore potential relationships. To further limited multiple testing, a multiple linear regression model was used that allows for the inspection of loadings of individual regressor loadings. The regression model was not significant ($F_{5,31} = 0.722$, p = 0.613, $R^2_{adjusted} = 0.0$), and data suggested that none of the predictors was significant (Table 5.1).

Model	Unstandardized	Standard Error	Standardized	t	p
(Intercept)	0.770	0.793		0.971	0.341
TCAQ	0.047	0.056	0.182	0.841	0.408
SES	0.038	0.059	0.129	0.632	0.533

Model		Unstandardized	Standard Error	Standardized	t	p
	TCQ	-0.034	0.075	-0.089	-0.449	0.657
	BAS	0.133	0.137	0.192	0.972	0.340
	BIS	0.180	0.250	0.148	0.719	0.479

Table 5.1 Regression coefficients for self-efficacy measures.

5.3.3.4 Whole brain analyses

Figure 5.5 shows brain regions involved in the NF training (i.e. task vs. rest). For the NFE group, limbic and subcortical regions (e.g. the insular cortex, caudate and hippocampus) were activated while the dorsolateral prefrontal cortex was deactivated (Figure 5.5 A; Appendix B Table B1). For the NFS group, mainly higher visual areas were activated including the bilateral PPA as the main target region of this group (Figure 5.5 B; Appendix B Table B1).

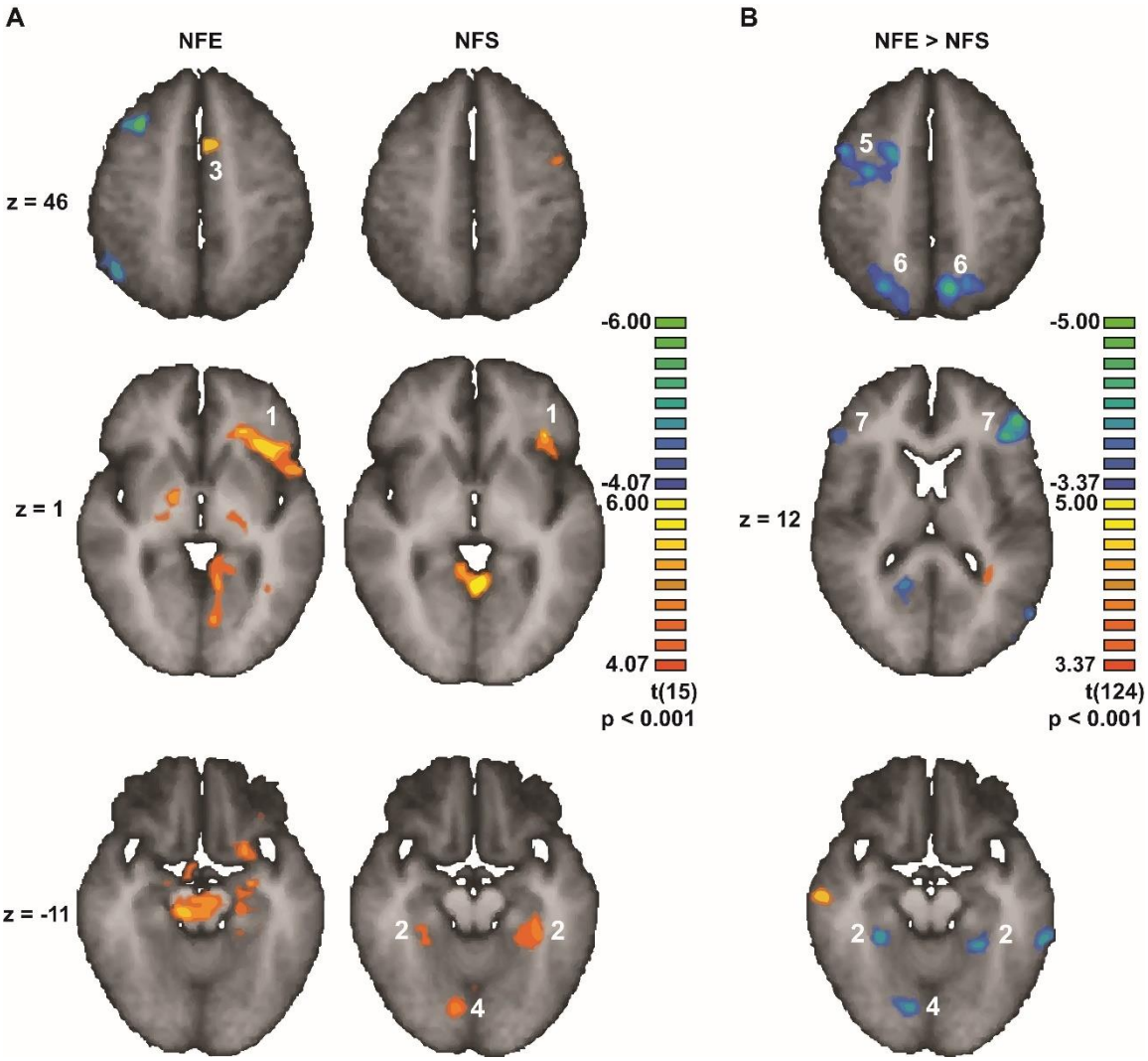

Figure 5.5 Whole Brain Analysis. A) Activity of intervention groups, shown separately for NF-Emotion and NF-Scene group. B) Group contrast. Key areas are labelled with numbers: 1) Insular cortex/ ventrolateral prefrontal cortex, 2) parahippocampal place area (PPA), 3) supplementary motor area, 4) lingual gyrus, 5) premotor cortex, 6) superior parietal lobule and 7) dorsolateral prefrontal cortex (DLPFC). Statistical maps cluster-threshold corrected ($p < 0.001$). Adapted from (Mehler et al., 2018).

A direct group comparison suggested that the NFS group showed more activation of the PPA and frontal regions (Figure 5.5 B; Appendix B Table B2). Regions that were activated in both groups included the anterior insular cortex and dorsal cingulate gyrus Figure 5.5 (see also Appendix B Table B2).

Lastly, a linear contrast of time indicated that participants across groups showed increased activation in prefrontal areas with a cluster extending into the dorsolateral prefrontal cortex, parts of the dorsal striatum (putamen and caudate body), and premotor cortices including the SMA (Appendix B Table B3).

5.3.3.5 Relation between changes POMS score and up-regulation

For neurofeedback sessions, no significant relationships were found, and correlation coefficients were rather small (session 1: $r = 0.058$, $p = 0.375$; session 2: $r = 0.057$, $p = 0.378$; session 4: $r = -0.172$, $p = 0.827$; session 5: $r = 0.167$, $p = 0.185$). However, a significant correlation was found between pre-post POMS $_{TMD}$ difference scores and target region activation during session 3 (session 3: $r = 0.358$, $p_{Bonferroni} = 0.048$). Likewise, a one-sided Bayesian correlation suggested moderate evidence for a positive relationship ($BF_{+0} = 3.671$, median posterior Pearson's rho = 0.319). This result suggested that participants who activated target regions more during the transfer run also showed larger pre-post improvements on the $POMD_{TMD}$.

5.3.3.6 Relation between changes in insula activation self-efficacy and HDRS-17

Similarly, no significant correlation was found between changes in left or right activity and changes in self-efficacy until the primary endpoint (left: $r = 0.058$, $p = 0.751$; right: $r = 0.002$, $p = 0.989$) and Bayesian hypothesis tests indicated moderate evidence for the absence of a correlation (left: $BF_{01} = 4.348$; right: $BF_{01} = 4.545$). Further, no significant correlation between left or right insula activity during the first session and clinical improvement until the primary

endpoint was found (left: r = 0.056, p = 0.762; right: r = 0.059 = p 0.750). Bayesian hypothesis tests indicated moderate evidence for the absence of such correlation (left: BF_{01} = 4.329; right: BF_{01} = 4.348).

5.4 Discussion

The analysis of neuroimaging data from this RCT demonstrated that moderately to severely depressed medicated patients could activate target regions located in limbic and higher visual areas by imagining either positive emotions (NFE) or visual scenes (NFS), respectively. Data further suggested that both groups up-regulated target regions when feedback was provided, but that they did not up-regulate localised target regions by purely engaging in mental imagery without feedback. Further, no significant difference between groups in up-regulation was found and post-hoc equivalence tests suggested that both groups up-regulated to a similar extent with regards to the chosen SESOI (Lakens et al., 2018).

Exploratory whole brain analyses further showed that brain activation was largely constrained to areas that are associated with the respective mental imagery instructions and target region selection. For instance, the NFE group mainly activated limbic regions including parts of the striatum, the insula, the retrosplenial cortex, and the hippocampus (Vann et al., 2009). These areas are involved in reward processing, as well as the recall of episodic memory. The NFS group mainly recruited higher visual regions, including the PPA as the main target region, as well as the lingual gyrus. These data thus suggested that participants in the NFE group engaged successfully in mentalising relaxing scenes. Further, a direct group contrast mainly identified regions that were more activated in the NFS group. These included higher visual regions such as the PPA and precuneus, as well dorsolateral prefrontal cortex (DLPFC). Hence, regions that were relatively more recruited in the NFS group mainly dominated a group contrast. One possible reason is that imagining natural scenes was a more salient and heterogenous form of visual mental imagery, which may have resulted in a more consistent set of neural correlates. In contrast, mental imagery of positive emotions may involve different aspects and modalities. As such, it may involve visual, haptic or acoustic aspects, and these may vary from patient to patient. Moreover, patients could imagine social as well as non-social situations, which probably recruited different brain structures in the context of emotion self-regulation. For instance, differences in neural correlates of emotion regulation have been reported for social vs. non-social contexts (Vrtička et al., 2011). Hence, the overlap of activated regions may have varied more between patients in the NFE compared to NFS group. To disentangle these potential interactions, patients' strategies could be explored in more detail. Overall, whole brain analysis results largely replicated neural correlates that earlier studies have reported for mental imagery of emotions (Johnston et al., 2011; Linden et al., 2012), and visual scenes (Habes et

al., 2016; O'Craven and Kanwisher, 2000), respectively. Moreover, neural correlates identified in the whole brain analysis are in line with key areas that have been reported for neurofeedback training and that are mainly involved in self-regulation, attention control and reward processing (Sitaram et al., 2017). For mental imagery of scenes, the PPA was recruited in particular, yielding larger activation for the NFS group compared to the NFE group (Figure 5.5). Notably, the PPA forms part of the medial temporal lobe together with other structures such as the hippocampus and entorhinal cortex that form the hippocampal formation and play an important role in declarative memory. Further, functional and structural disturbances within this network have been linked to depressive symptoms, which can be partly reversed with antidepressant treatment (Aminoff et al., 2013; Bar, 2009). The relationship between changes in mood and memory remains to an active research field, and it remains possible that training of the PPA may itself impact mood.

It has been suggested that successful training of emotional areas could mediate clinical effects (Linden, 2014; Linden et al., 2012). The whole-brain analysis of the NFE group showed activation of the insula, the supplementary motor area, and the ventral striatum, indicating that patients engaged in emotion self-regulation. These areas have been linked to positive reappraisal of emotions (Buhle et al., 2014), as well as in emotion regulation, as suggested by an activation likelihood estimation (ALE) meta-analysis that included data from 479 healthy participants across 23 studies (Kohn et al., 2014). Taken together, these whole brain analysis findings are encouraging because both emotion reappraisal and emotion regulation constitute important elements in psychotherapeutic treatment for depression, but both processes can be impaired in depressed patients (Zilverstand et al., 2017). The data presented suggest that rt-fMRI-NF training could indeed help patients in training this capacity. Of interest, activated voxels in the NFS and NFE partly overlapped, mainly in the anterior insula (see Figure 5B and Table S3). The (anterior) insula has been implicated in various cognitive tasks that may explain this overlap, including its involvement in the integration of visual information that is required for mental imagery of scenes (De Borst et al., 2012), self-regulation of affective as well as executive functions (Chang et al., 2013; Langner et al., 2018), and online monitoring of performance (Paulus and Stein, 2010), a factor for which groups were matched. Moreover, it is involved in allocating attentional resources and coordinating information processing between internal and external environments (Sitaram et al., 2017). Switching between these two modes is a central component of self-regulation training, which requires monitoring internal processes in the form of introspection during mental imagery, but also adapting strategies based on

external cues in the form of feedback signals. Without more fine-grained fMRI measurements, however, it cannot be determined if the overlapping clusters between groups reflect activation of the same neuronal populations. On a more general level, though, this overlap highlights the difficulty of selecting strictly separated target networks for different neurofeedback conditions.

As outlined above, the (anterior) insular cortex is involved in various cognitive processes that are central to fMRI-NF self-regulation training, including self-regulation, error monitoring and allocating cognitive resources. It may thus not be surprising that is has been identified as a common neural correlate across fMRI-NF training across paradigms (Emmert et al., 2016). Furthermore, the right anterior insular cortex in particular has shown promise as a biomarker to predict treatment response to CBT and pharmacotherapy (Dunlop and Mayberg, 2014), as well as relapse after remission (Zaremba et al., 2018). Given the strong clinical improvements and their relation to changes in self-efficacy (Chapter 4.3.6.4), it was first explored whether there was a relationship between changes in self-regulation and self-efficacy. No significant correlations were found, and Bayesian hypothesis testing suggested moderate evidence for the absence of a relationship. Furthermore, it was explored whether the activity of the insula during the first session could predict clinical changes until the primary endpoint. No significant correlations were found for these exploratory tests and Bayesian results suggested moderate evidence for the absence of an effect. Taken together, despite remarkable changes in depression scores, self-efficacy and recruitment of the anterior insula across both groups, a relationship between neuroimaging and clinical/psychometric measures could not be established.

The relationship between changes in self-regulation and clinical changes was explored and only the main intervention group (NFE) showed a significant relationship (Figure 5.4). The interpretation of this finding remains difficult, given that both groups showed similar clinical improvements. One possibility is that potential additional therapeutic effects resulting from the training of areas involved in emotional self-regulation were too small to be detected with the present sample size. Another possibility is that the correlation was driven by the overlap of limbic target regions in the NFE group that represent areas that are implicated in self-regulation and reward processing (Sitaram et al., 2017), including the basal ganglia and the insula (Figure 5.2). To test whether the relationship identified for the NFE group could be reproduced in the NFS group for a similar set of brain regions, a negative control analysis was performed. This control analysis found no relationship between changes in activity levels of these brain regions and clinical changes for NFS group, supporting the specificity of the finding reported for the

NFE group. Taken together, these results may be indicative that the relation between a change in self-regulation and in clinical scores was rather specific to the training experience of the NFE group. However, it should be noted that the current sample was relatively small and may thus have overestimated this relationship (Algermissen and Mehler, 2018). Further, the interpretation of this finding remains difficult because no mean group differences for clinical changes were observed (Chapter 4.3.3). As surveyed earlier (Chapter 4.4) the overall strong clinical response in both groups is comparable to effects reported for augmentation therapy and superior to large unspecific effects reported for high-technology interventions and is likely to represent a genuine clinical effect in both groups. It remains possible that the NFE fMRI-NF intervention did indeed target the pathophysiology of depression in a more specific way as was hypothesised initially (Chapter 4.1), but that the current sample sizes were too small to detect group differences. Replication in larger samples and meta-analytic analyses using similar self-regulation metrics will allow the exploration of the relation between these variables for fMRI-NF training in depression, but also other psychiatric conditions.

With regards to the strong clinical effects observed in the NFS group, the neuroimaging data provide further potential explanations: First, both groups up-regulated target ROIs to a similar extent, as also shown by equivalence tests, and participants in the NFS group experienced similar success rates of self-regulation and positive reinforcement. Hence, patients in both groups were probably successfully matched for the rewarding experience of veridical self-regulation training. Both factors, success of self-regulation and positive reinforcement, could be critical for the clinical efficacy of neurofeedback training in depression. This view is supported by the positive relationship reported for changes in self-efficacy and clinical effects (Chapter 4.3.6.5). Moreover, it is also in line with data reported from a recent fMRI neurofeedback RCT that demonstrated the clinical superiority of targeting the affective brain compared to a sham condition (Young et al., 2017). The main group of their trial received feedback from the amygdala (a condition comparable to our NFS group) and their control group from an area (the intraparietal sulcus) that patients could not self-regulate as effectively. While their main group improved clinically to a similar degree as both groups did in this trial (Chapter 4), Young et al's control group did not show significant clinical improvement.

One central design aspect in fMRI-NF interventions is the choice of target and control regions and the selection procedure employed to identify suitable voxels. The current study selected target regions based on subjects' brain responses during the localiser run. Such an approach

has the advantage that it circumvents the danger of placing too much confidence in biomarker areas, which may be hard to determine at individual level. In particular functional biomarkers that have been suggested for depressive research have thus far shown little convergence between studies (Müller et al., 2017). Moreover, consistent selection of voxels within the same target region based on functional localiser scans may be challenging given the relatively low repeat reliability. For instance, putative fMRI biomarkers for emotional face processing show substantial between-session variability. Whereas average activation on a group level seems replicable, intraclass-class correlation for areas such as the amygdala and the subgenual anterior cingulate cortex are relatively low (Nord et al., 2017). Using anatomical masks to guide selection of target regions can help addressing this issue (Young et al., 2017). One advantage of using the same targets regions for all patients in one group and throughout all sessions is that such approach yields probably larger learning effects over time because patients receive feedback from a very similar set of voxels. In other words, the subject and session specific target region selection in the current trial may have slowed down potential learning effects and could partly explain why no significant effect of time for the up-regulation of target regions was observed. Another advantage of selecting very similar voxel sets for all participants within one group is that this approach facilitates the comparison of measures of real-time imaging quality control including the temporal signal-to-noise ratios (Heunis et al., 2018; Stoeckel et al., 2014). Moreover, it facilitates performing negative control analyses because ROIs can simply be performed across training conditions to test for potential dissociations (Young et al., 2017). These considerations illustrate that the choice of target region and selection procedure remain subject to further refinement in light of control theory aspects, but also the condition that is being treated (Arns et al., 2017; Stoeckel et al., 2014; Young et al., 2018b).

Apart from classical localiser scans based on affective visual stimulation, target region selection may also be informed by network dynamics of patients, for instance acquired during resting-state (Drysdale et al., 2016; Ramot et al., 2017). This approach has been successfully used to identify neurophysiological subtypes in depression that were predictive of treatment response to TMS (but see also Dinga et al., 2018). Moreover, network-based metrics provide an attractive alternative to target region-based training paradigms. Specifically, such approaches can be informed by cognitive models in depression and incorporate top-down control mechanisms that represent a relevant therapeutic neurocognitive pathway in the treatment of depression (Disner et al., 2011). The feasibility of connectivity based feedback has recently been successfully tested for emotion regulation during threat exposure (Zhao et

al., 2018). Specifically, the study used a within-subject sham feedback-controlled design where feedback was provided based on correlations between ventrolateral prefrontal cortex and the amygdala during the experimental condition. During the sham control runs feedback was provided based on correlations between areas that are not related to emotion regulation. The authors demonstrated PoC that correlation measures between target regions could be controlled with emotion regulation strategies for the experimental, but not the sham condition. Further, these changes lasted at follow-up sessions and in the absence of feedback and changes in self-reported anxiety scores were observed, that partly related to changes in connectivity metrics. Importantly, participants did not notice the difference between sham and veridical feedback conditions and were thus not unblinded to the sham manipulation. Changes in network activity between the amygdala and prefrontal areas have also been reported for amygdala based fMRI-NF in depressed patients (Young et al., 2018a). As explained in more detail earlier (Chapter 2.1.3), their trial employed positive emotion self-regulation training of the amygdala based on the recall of positive autobiographical memory and incorporated a transfer task to test changes in autobiographical memory performance. Gains in positive recall were related to connectivity changes in particular in the experimental (amygdala training) group, but surprisingly also in the control (intraparietal sulcus training) group. These results suggest that besides self-regulation success, mental imagery training itself may yield largely comparable changes in functional markers such as network connectivity. To model the relationship between clinical changes, self-regulation success and cognitive performance, the authors also performed a mediation analysis (Young et al., 2017), which suggested amygdala activity during a transfer run (when no feedback was provided) explained a significant amount of variance in the relationship between gains in memory performance and clinical outcome measures (but see criticism in Chapter 2.1.3). The trial presented in this thesis also explored the relation between self-efficacy measures and self-regulation success based on pre-post difference contrast. While this contrast established a relationship between self-efficacy and clinical effects, no relation between self-efficacy measures and self-regulation were found and thus a follow-up mediation analyses between these measures was not justified. Future designs of fMRI-NF RCTs may benefit from incorporating independent physiological or cognitive markers to underscore the generalisability of findings, explore therapeutic mechanisms and to test the specificity of certain training types.

To conclude, neuroimaging data suggested that both groups recruited brain areas according to their mental imagery strategies. Moreover, both groups up-regulated training areas to a similar

extent, suggesting that patients were matched for success in self-regulation and reward experience. Taken together with the reported clinical findings (Chapter 4.3.3), this trial demonstrated that if moderately to severely depressed patients can up-regulate a brain area that is not immediately linked to affective processing (NFS group), they can experience similar clinical benefits compared to patients who completed emotion-focused neurofeedback training (NFE group). This suggests that the experience of brain control and the positive reinforcement of mental imagery may be necessary components for the therapeutic effects of neurofeedback and should therefore be considered for future designs. Potential neural markers for self-regulation success and their relation to changes of self-efficacy measures and depression severity were explored but could only partly be confirmed. Further exploration in larger clinical trials, and possibly across fMRI-NF trials in different conditions is needed to further the understanding of therapeutic mechanisms.

6 MOTOR IMAGERY BASED NEUROFEEDBACK OF PRIMARY MOTOR CORTEX AND SUPPLEMENTARY MOTOR AREA

6.1 Introduction

Real-time functional magnetic resonance imaging neurofeedback (fMRI-NF) is currently being explored as a non-invasive technique to improve motor rehabilitation outcome in neurological conditions including Parkinson's disease (PD) (Subramanian et al., 2011) and stroke (Liew et al., 2016; Linden and Turner, 2016; Wang et al., 2017). Recently, the first randomized controlled trial for fMRI-NF in PD suggested that the technique may lead to clinically significant motor improvements when combined with physical exercise (Subramanian et al., 2016). Regarding stroke, a recent systematic review concluded that fMRI-NF training can lead to learned modulation of brain signals and may be associated with beneficial behavioural changes (Mihara and Miyai, 2016; Wang et al., 2017). Hence, there is an increased interest in improving fMRI-NF protocols to maximize the potential of the associated clinical outcomes (Stoeckel et al., 2014).

One central consideration when designing fMRI-NF interventions is the choice of suitable target brain region(s) from which feedback is provided during training. Although some techniques work with implicit feedback that does not require cognitive strategies (Watanabe et al., 2017), most fMRI-NF training protocols currently employ mental strategies such as visual imagery (Habes et al., 2016), affective imagery (Johnston et al., 2011; Linden et al., 2012; Young et al., 2017) or motor imagery tasks (Auer et al., 2016; Berman et al., 2012; Blefari et al., 2015; Chiew et al., 2012; Subramanian et al., 2011; Subramanian et al., 2016). Therefore, target regions are often selected based on their anticipated involvement in the respective mental imagery process.

For fMRI-NF of motor areas, most protocols instruct participants to perform motor imagery (Blefari et al., 2015; Sepulveda et al., 2016; Sitaram et al., 2012; Subramanian et al., 2016). FMRI correlates of motor imagery have been well documented in healthy individuals (Guillot et al., 2009; Jeannerod, 2001; Munzert et al., 2009; Sharma and Baron, 2013; Solodkin et al., 2004; Vingerhoets et al., 2002) and preliminary data from stroke survivors also showed robust neural correlates of motor imagery (Bajaj et al., 2015a, b; Kraft et al., 2015; Sharma et al., 2009; Wong et al., 2013) in the ventral premotor cortex (vPMC), dorsal premotor cortex (dPMC), and the supplementary motor area (SMA; (Hétu et al., 2013)) This finding has been corroborated by recent motor imagery based fMRI-NF studies (Blefari et al., 2015; Sepulveda

et al., 2016; Subramanian et al., 2016). Further, neuroimaging studies of motor imagery and motor execution indicate that both tasks share overlapping motor networks, although differences have been identified (Hanakawa et al., 2008; for a review see Hetu et al., 2013). For the primary motor cortex (M1), the motor imagery literature consists of mixed findings, and it is not clear whether M1 can be robustly *activated* during motor imagery training. It thus remains an open question whether M1 represents an effective target region for motor imagery-based fMRI-NF paradigms that entrain up-regulation.

The first aim of this study was thus to compare self-regulation abilities for M1 and a higher motor region (SMA) in the same participants. Its second aim was to explore the feasibility of graded neurofeedback of motor regions. The rationale for graded neurofeedback protocols, where participants are not just trained to up-regulate activity in the target region but to up-regulate it to different specified levels, is that they provide more degrees of freedom for adaptive neurorehabilitation programmes and for neural communication in a BCI framework. 'Gradual' fMRI-NF has been introduced recently (Sorger et al., 2016) (referred to as 'graded' in this study). It offers increased scope for evaluating training success, for example by assessing how well participants can attain discrete magnitudes of BOLD signal changes through supervised mental imagery. Given that it is well established that certain motor regions, including the SMA, are *activated* during non-supervised motor imagery, significant BOLD activation of a motor region during supervised motor imagery training (e.g. fMRI-NF training) does not by itself provide sufficient evidence of volitional control success. Graded fMRI-NF training, in contrast, allows one to gauge the degree of control gained over the activation level of a target region.

This study aimed to lay the foundation for the further development of motor-imagery based neurofeedback of areas in the motor network. The graded NF protocol separates the general effects of motor imagery from the more specific neurofeedback targeted effect of volitional self-control as reflected in the BOLD signal. Participants were required to target two discrete levels (*low* and *high* level), while feedback was provided from either the SMA or M1. This experimental design offers a distinct conceptual purpose, in that its factorial nature provides a degree of within subject control. Whenever a participant is provided with feedback from one region, signal changes in the other region are governed by the effects of motor imagery and attempted self-regulation, but crucially, lack the specific 'active ingredient' of feedback information from the target region. Thus, for each region of interest (ROI) two conditions were

included; one in which feedback information is provided (*active* condition), and one in which feedback from the other region is provided (*passive* condition). Essentially, the novelty of this design is that it allows this to be assessed, alongside the typical neurofeedback effect, within a single factorial framework by contrasting which of the two regions feedback information is derived from.

The purpose of this study is to provide the foundation for future development of fMRI-NF protocols for neurorehabilitation such as stroke and PD. Given that the majority of published studies have targeted cortical motor regions (Wang et al., 2017), typically premotor areas or primary motor areas, the aim was to determine which region is most suitable for motor imagery based NF. The graded NF design allows to address two important questions;

1) When combined with a kinaesthetic motor imagery strategy, are SMA and M1 robustly activated? This is a key prerequisite for further NF up-regulation training.
2) Do subjects show better separation between discrete target levels for a given region when provided with feedback from than region?

It was hypothesised that SMA based fMRI-NF training would show robust activation in the SMA, but no activation in M1. Given previous unsuccessful attempts of M1 fMRI-NF training (Berman et al., 2012; Blefari et al., 2015; Chiew et al., 2012), it was hypothesised that M1 based fMRI-NF training would not yield M1 activation, whilst the SMA as a region involved in motor imagery would still show activation. Lastly, differences in SMA activation between target levels when feedback was provided from the SMA (active condition), but not when feedback was provided from M1 (passive condition) were hypothesised. The main findings were that SMA was robustly activated during motor imagery, whereas negative BOLD responses were found in M1. There was a significant effect of graded self-regulation in SMA, which was mostly driven by the *active* condition, but NF provided only minimal benefit.

6.2 Methods

6.2.1 Participants

Twenty healthy participants were recruited from an internal experiment database. All participants gave written informed consent and the Cardiff University School of Psychology Ethics Committee approved the study. Data from three participants were excluded due to

technical difficulties with the feedback system during scanning. Data from the remaining 17 participants (8 female; age 26.6 ± 5.5 [Mean ± SD] years) were included in the data analysis.

6.2.2 Motor imagery questionnaire

Motor imagery can be performed in two different modalities. It can be mainly visual or mainly kinaesthetic. Visual motor imagery focuses primarily on visual mental imagery (either from a first or a third person perspective), whereas kinaesthetic motor imagery usually involves taking a first-person perspective and imagining the feeling and experience of movements without overt movement (Hanakawa, 2016). Previous literature suggests that only a minority (27%) of studies had specified the modality that participants used during motor imagery training (Hetu et al., 2013), indicating that participants received inconsistent or unspecific instructions between or across studies. Kinaesthetic and visual motor imagery are associated with distinct neural activity patterns (Guillot et al., 2009; Kilintari et al., 2016; Solodkin et al., 2004) and this may partly explain the inconsistent findings reported for M1in the motor imagery literature. In the present study participants completed a standard motor imagery questionnaire before the scanning session, which included self-ratings for both visual and kinaesthetic motor imagery tasks on a scale from 1 to 7 (Gregg et al., 2010). The questionnaire was used (1) to make participants more aware of the distinction between the two different motor imagery forms and (2) to assess their self-rated baseline ability of motor imagery. After scan sessions, participants were asked to fill in a questionnaire about their experience.

6.2.3 Instruction for localiser and neurofeedback training

Among the few studies that have reported the modality of motor imagery instructions and M1 activation, only kinaesthetic motor imagery activates M1 (Hetu et al., 2013; Solodkin et al., 2004; Stinear et al., 2006) and therefore participants used only kinaesthetic motor imagery strategies (Guillot et al., 2009). Specifically, participants were instructed to imagine the body sensation and experience of moving both of their hands without moving them and while remaining relaxed. Participants could choose the precise movement they imagined but were told to imagine a movement that they regularly perform (e.g. playing an instrument, housework or typing).

During the localiser run, participants were presented with an empty thermometer on the screen. Participants were instructed to perform a paced bimanual motor execution task (finger opposition) during task periods (as indicated by green arrows). Specifically, numbers from 1 to 4 were presented at a frequency of 1.33Hz to pace the movement during task periods. During

rest periods (as indicated by red arrows) participants were instructed to rest and relax. Participants were reminded that no feedback was presented during the localiser run.

For the neurofeedback runs, it was explained to participants that bars in the thermometer represented the activity level in the target region. Participants were instructed that their goal was to use kinaesthetic motor imagery to control the feedback by filling up the bars of the thermometer display. They were further instructed to attempt to maintain the activation at target levels by adjusting their motor imagery strategy (e.g. by changing the speed and/or intensity of the imagined movement). Besides these aspects, participants were not restricted in the content of motor imagery and could imagine any activity or sport that involved movements of both hands. Indeed, participants were encouraged to explore which motor imagery strategies were most effective in gaining control over the filling bars of the thermometer display. Participants were further explained that mental imagery strategy changes would not immediately affect the feedback because of the hemodynamic delay. Participants were also instructed and reminded between scans to remain still and relaxed and avoid muscle contractions during scanning.

6.2.4 MRI data acquisition

Imaging data were acquired using on 3T General Electric HDx with an eight-channel receiver head coil. Blood oxygenation level-dependent (BOLD) fMRI runs (see description below) were measured with a T2*-weighted gradient-echo echo-planar imaging (EPI) sequence. Each functional EPI volume contained 25 slices of 2.5-mm thickness, with 0.5-mm inter-slice spacing (in-plane resolution = 3 mm, matrix size = 64 × 64, FoV (field of view) = 192 mm, TR (repetition time) = 1550 ms, TE (echo time) = 30 ms, flip angle = 80°, orientation = transversal). High-resolution structural images were acquired before the first functional scan using a fast-spoiled gradient echo sequence (FSPGR) with 172 contiguous sagittal slices of 1-mm thickness (voxel size: 1 × 1 × 1 mm, TR = 7.9 s, TE = 3.0 ms, flip angle = 20°, FoV = 256 × 256 × 172 mm).

6.2.5 fMRI-NF setup

Reconstructed DICOM images were transmitted in real-time from the MR computer to a dedicated analysis computer. Turbo-BrainVoyager (TBV) software (BrainInnovation B.V., Maastricht, The Netherlands, version 3.2) was used for real-time online pre-processing and analysis of BOLD signals including motion correction (with respect to the first volume of the functional localiser) and spatial smoothing (4mm full width at half maximum; FWHM).

6.2.6 Physiological recordings

As with any BOLD fMRI study, magnitude changes can be confounded by non-neural sources of variance, such as motion and physiological noise (Bright and Murphy, 2017). Whereas most motor imagery based paradigms target a PSC of 0.5 to 1.0, extensive breath holds may lead to PSC of up to a PSC of 6.0 (Thibault et al., 2018b). In the context of fMRI-NF, a recent critical systematic review has highlighted the importance of testing for the effect of physiological confounding factors of the BOLD signal (Thibault et al., 2018b).Physiological processes may be modulated by underlying mechanisms that also respond to the cognitive demands of a task, such as arousal or concentration, e.g. increased heart-rate during arousal or subconscious changes in breathing during increased mental activity. Specifically, partial pressure end-tidal carbon dioxide ($P_{ET}CO_2$) acts as a strong vasodilator. Moreover, changes in the heart may also impact the BOLD signal (Murphy et al., 2013). Instead, physiological data were collected to test offline whether physiological parameters correlated with the task predictor. Recorded pulse waveforms were recorded using pulse oximetry, from which heart-rates (HR) were calculated. Also, the partial pressure end-tidal carbon dioxide ($P_{ET}CO_2$) was recorded using a nasal canula from most participants. Data was recorded with Spike2 (version 5.21, Cambridge Electronics Design Limited, Cambridge, UK) for 10 ($P_{ET}CO_2$) and 13 (pulse traces) participants, respectively.

6.2.7 fMRI training session

The sequence of scans is shown in Figure 6.1A. Subjects lay supine in the scanner with their heads fixed using foam cushions to minimise head motion and they were instructed to remain as still as possible during data acquisition. Scanning started with an anatomical scan, followed by a motor execution functional localiser run (LOC). The LOC run served to identify functionally relevant voxels and to calculate individual percent signal change (PSCLOC) to scale the visual feedback. The motor execution LOC run (272 volumes) consisted of four blocks of bilateral finger opposition, flanked by five rest blocks (each block consisted of 16 volumes and lasted for 24.8 seconds). Once target regions (M1 and SMA) were identified, the session proceeded with five neurofeedback runs (NF) with one ROI, followed by five NF runs with the other ROI, using a counterbalanced order across participants.

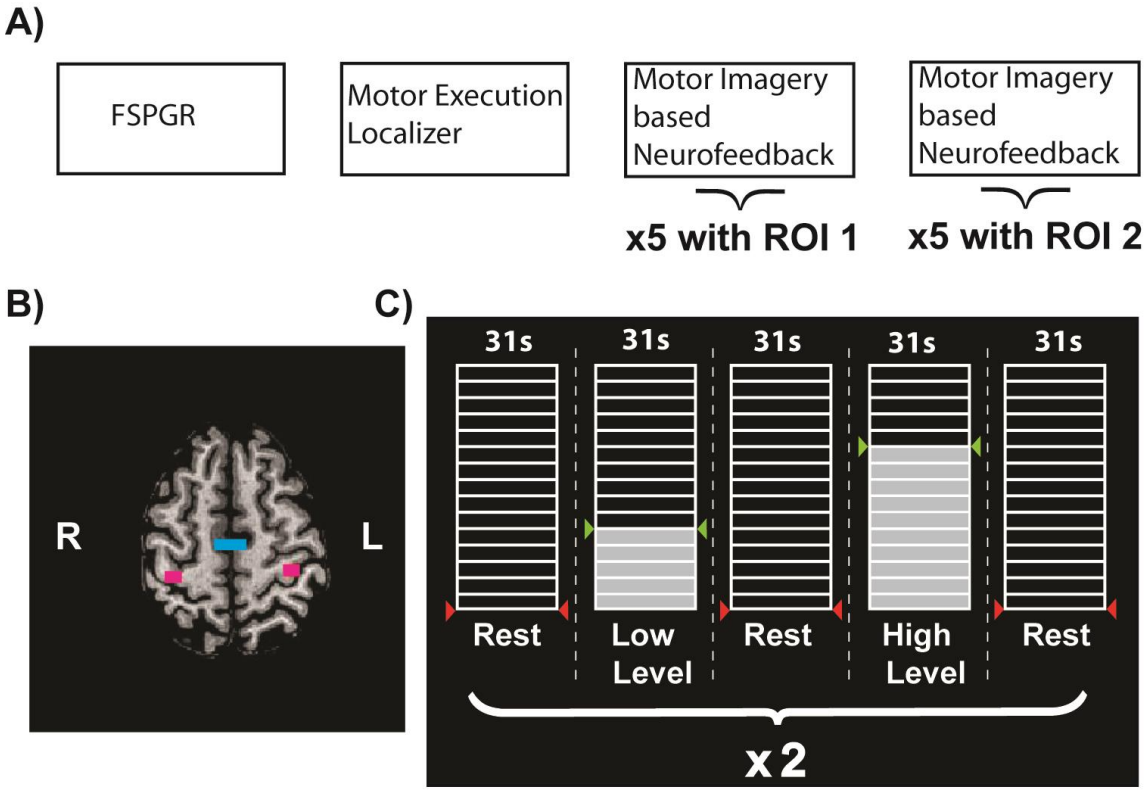

Figure 6.1 Experimental Setup. A) shows sequence of scans during train session, b) shows target region with SMA in blue and M1 in magenta for an exemplary participant. C) shows an exemplary sequence of the thermometer display with low level and high-level training blocks, flanked by rest blocks. Adapted from (Mehler et al., 2019).

If feedback information was provided from the SMA, it was labelled as the *active* condition for SMA and *passive* condition for M1, and vice versa. NF runs consisted of 180 volumes each, and contained two repetitions of two block types, a *low* and *high* neurofeedback level, which were flanked by rest blocks (each block consisted of 20 volumes and lasted for 31 seconds). Thus, the study had a factorial design, in which for each ROI there were two conditions (*active* and *passive*), each with two levels (*high* and *low*). Participants were instructed to use kinaesthetic motor imagery involving both hands during fMRI-NF and to avoid any movement or muscle contractions.

6.2.8 Target region selection and feedback normalisation

An incremental GLM was used for online analysis of both LOC and NF runs (Version 3.0, Brain innovation, Maastricht, The Netherlands); the model included a single task predictor for both LOC and NF. For LOC, a linear drift term was also included. The selection of voxels for the target regions was guided by use of t-statistic maps (with a variable t-threshold, but at least t = 5.0), as well as anatomical landmarks. For M1, this study aimed to restrict the ROI to the

hand knob area (Yousry et al., 1997). For the SMA ROI, anatomical selection aimed to target the proper SMA, and avoid selecting voxels in the pre-SMA. Local gyri and sulci have been shown to be rather variable between subjects and be less reliable, than for instance, white matter tracts (Behrens et al., 2006). To identify proper SMA, the vertical line traversing the anterior commissure was used, which is perpendicular to the anterior commissure/posterior commissure plane (Picard and Strick, 2001; Zilles et al., 1996). Voxels were selected in both hemispheres within 3 slices and with a similar number of voxels per hemisphere and per target region (in total 36 voxels with 18 voxels per hemisphere for M1 and SMA; 6 voxels per slice). Based on selected target regions, the percent signal change (PSC) during the motor execution localiser was calculated for both regions (M1 and SMA) separately. This allowed us to calibrate the feedback presentation based on PSC values from the respective target region and thereby to account for differences in the hemodynamic response (Obata et al., 2004).

6.2.9 Online fMRI analysis, neurofeedback calculation and presentation

For feedback presentation during NF runs, the mean raw signal value was extracted from the ROI at the analysis computer and transmitted to a stimulation computer at every TR via a direct TCP/IP network connection to minimise delays in data transmission. The percentage signal change (PSC_{NF}) was provided as a feedback signal and was computed based on equation 1,

Eq. (6.1): $$PSC_{NF} = \frac{(val - baseline) * 100}{baseline}$$

where *val* is the mean of last three consecutive ROI signal values, and *baseline* is the median ROI signal value during the second half (i.e. last 10 TRs) of the preceding rest period. PSC_{NF} were then normalized by PSC_{LOC} to map it on to the (15) segments of the thermometer display such that every segment represents 10% of the PSC_{LOC}. Values below 0 were rounded up to 0, values above 15 were rounded down to 15. The calculation was carried out in an in house written Python script (Python 2.7.10). The Open Source Python library Expyriment was used for online feedback presentation (Krause & Lindemann, 2014). The visual feedback display consisted of a thermometer with 15 segments and indications for either rest (i.e. red arrows at the bottom) or the target levels (Figure 6.1C). Target levels were defined by 50% (*low*) and 100% (*high*) of the PSC_{LOC} and indicated by green arrows. Both the *low* and *high*-level conditions were repeated twice per run and were interleaved by rest periods. During rest periods, no feedback was presented, and the thermometer remained empty. The order of the condition (*low* and *high* target level) was counterbalanced across runs and subjects.

6.2.10 Post-training questionnaire

After scan sessions participants were asked to fill in a questionnaire about their experience, that was designed to rate their perception of the following aspects of the experiment: (1) controllability of feedback for both ROIs, (2) the difficulty of filling up bars in the thermometer, and (3) the difficulty in maintaining the thermometer at a discrete level. For question 1, a score of 10 indicated high controllability, whereas for question (2) and (3), a score indicated high difficulty. Question (2) and (3) were added as exploratory items to the questionnaire after data collection had already started and were thus only recorded in 10 participants.

6.2.11 Offline fMRI analysis

All subsequent results are based on an offline analysis of raw data that was performed in AFNI (version 16.2.18). Data were motion corrected and spatially smoothed (4mm FWHM) with pre-processing parameters identical to the online analysis. Mean ROI time series were then extracted for all runs and subsequent analysis was restricted to these. Since ROI mean time series were extracted from both ROIs for each run, for each ROI there exists a condition in which it was the "active" target for NF training (active), as well as a condition in which NF training was performed on the other ROI (passive). Thus, for each ROI, the active condition contains the response to motor imagery when feedback was provided from that region, and the passive condition contains the response to motor imagery when feedback was provided from the other region. Thus, the passive condition served as a form of internal (offline) control condition for a given ROI, because it reflects the response to motor imagery in that ROI when NF training is based on the feedback signal from the other ROI.

For LOC and NF runs a GLM analysis of the data was performed. For LOC runs, the model included a predictor for activation blocks and a linear drift term. For NF runs a predictor that modelled the NF response across both levels was included, as well as a parametric predictor that modelled the difference between levels. No linear drift term was included, but separate pre-onset baseline period predictors were included to match the online analysis baseline period (i.e. 10 volumes preceding stimulus onset). Thus, the offline analysis replicated the online analysis, except for the addition of pre-whitening to account for temporal auto-correlation of the BOLD signal (Woolrich et al., 2001).

Offline PSC values were defined similarly to the online analysis, as a ratio between task

response and baseline estimation. In the offline analysis the LOC baseline period was simply represented by the intercept term in the model (as is the norm for standard analysis). For NF runs, PSC values were calculated for *low* and *high* levels separately, and the baseline period was represented by the intercept term plus the pre-onset baseline period that was used in the online analysis. In this way, the offline NF PSC values are representative of the average online NF PSC values.

6.2.12 Power calculation

No *a priori* power calculation was performed. It was aimed to include a number of participants (15-20) that were comparable to the average or participants included in most fMRI studies (Poldrack et al., 2017). Data analysis was not performed until the last participant was scanned.

6.2.13 Planned statistical analyses

For each subject, median PSC values were calculated across all runs for each ROI in both *active* and *passive* conditions, and for the *low* and *high* target level separately. For each ROI data were fit to a 2x2 ANOVA model (target level x feedback), with target level having *low* and *high* levels and feedback with *active* and *passive* conditions. Physiological traces (P_{ET} CO_2 and HR) were convolved with respective hemodynamic response functions and correlated with the task predictor (Pearson's r). Obtained correlation coefficients were Fisher-z-transformed for both measurements and ROIs, averaged across runs, and submitted to one-sample t-tests.

All frequentist t-tests were carried out one-sided, unless stated otherwise. Bayesian paired t-tests were conducted using JASP (version 0.8.5.1; Team, 2017) using informed half-normal priors. Priors that were used to test for M1 activity were scaled by 75% of the group M1 PSC value as measured during the functional localiser and priors to test for differences between *low* and *high* target level of SMA activity were scaled by 50% of the group SMA PSC as measured during the functional localiser, which reflects the scaling of target levels during online feedback (for more details on used prior distributions, see Results section).

6.2.14 Exploratory analyses

6.2.14.1 Training effect over time

To test for an effect of time in level separation, PSC values for the low and high target level of each run and participant were submitted to a repeated measures ANOVA.

6.2.14.2 Comparison of other self-regulation experience ratings

Graded feedback training requires of participants to reach and maintain target levels on the visual thermometer bar feedback display. Besides general controllability of the feedback, a subset of participants also rated how well they could 1) fill up and 2) maintain the filling level of bars in the visual thermometer feedback display. Data for both ratings were submitted to two-sided frequentist and Bayesian t-tests.

6.2.14.3 Counterbalancing and potential order effects

This study employed a repeated measures within-subject design in which the order of training ROIs was counterbalanced across participants. While such design has the advantage that they increase statistical power because they eliminate between subject error, is bears the risk of carry-over/order effects. In the current study, the starting condition (i.e. the choice of the first target region) could have biased results. Hence, post-training questionnaire and ROI data were tested for potential order effects. Given the differences in self-regulation experience between M1 and SMA training runs, it is possible that the starting condition affected participants performance. For instance, participants' motivation may have differed such that individuals who started in the M1 condition may have been discouraged when they reached the SMA training session. Such carry over effects could confound results and for instance explain why no interaction between the feedback condition and target level was found for *active* SMA PSCs. While the current study did not operationalize a measure for participant's motivation, hypothetical differences in motivation between participants of different starting conditions would likely yield differences in participants' average evaluation of controllability. Average controllability ratings were split by starting condition and submitted to two-sided independent frequentist and Bayesian t-tests.

Besides perceptual effects, another possible source for an order effect could relate to a transfer of learned separation between low and high target BOLD levels: participants who started with SMA training may have been able to apply successful motor imagery strategies learnt during SMA training during M1 training runs. Such transfer effect would be reflected in a larger separation of target levels during *SMA passive* training runs for some participants, rendering the detection of an interaction between feedback condition and target level more difficult. To test whether SMA *passive* PSC values differed between starting conditions, difference scores of participants' median PSC values were split accordingly (by starting condition) and submitted to two-sided independent frequentist and Bayesian t-tests.

6.3 Results – planned analyses

6.3.1 Motor imagery questionnaire

Motor imagery questionnaire data were collected from 14 of 17 participants. Data suggested that participants reported competence in imagining movements visually (5.5 ± 0.2 points [Mean ± SEM]) and kinaesthetically (5.4 ± 0.2 points), as indicated by the average ratings. The sum scores for visual (38.4 ± 1.7; range 24-48) and kinaesthetic (35 ± 1.6; range 28-44) motor imagery were also comparable. During administration of the questionnaires, all subjects reported that they understood the difference between the two forms of motor imagery and provided examples for illustration. Hence, it can be concluded that participants in this sample were able to engage in kinaesthetic (upper-limb) motor imagery.

6.3.2 Percent signal changes in target ROIs

Group mean PSC values during the motor execution (finger opposition) localiser task were 1.48 ± 0.12 [SEM] for M1, and 1.10 ± 0.08 for SMA, respectively. The intra-class correlation between online and offline PSC values was high for both target ROIs, (SMA: r = 0.961, 95%CI [0.895, 0.986], p < 0.001); M1: r = 0.929, 95% CI [0.809, 0.974], p < 0.001), indicating an excellent test-retest reliability between online and offline analyses.

Figure 6.2 shows the mean BOLD responses across subjects and runs for each ROI in both the *active* and *passive* neurofeedback conditions. A robust positive BOLD response during the task period can be seen in the SMA for both feedback conditions, whereas the M1 shows a clear negative response. Event related time courses indicated that the SMA showed sustained *activation* during the task period, while M1 was *deactivated* irrespective of whether feedback was provided from M1 or the SMA (Figure 6.2 A, B). M1 *deactivation* was confirmed by a t-test based on participants' M1 PSC values averaged across the *low* and *high* feedback level during the task period (M1 *active:* $t_{16} = -2.196$, $p = 0.022$; Cohen's $d = -0.533$; M1 *passive*: $t_{16} = -3.552$, $p = 0.002$; Cohen's $d = -0.862$). This was supported by Bayesian t-tests (N(0, 1.11); M1 *active*: $BF_{-0} = 3.496$; M1 *passive*: $BF_{-0} = 44.241$).

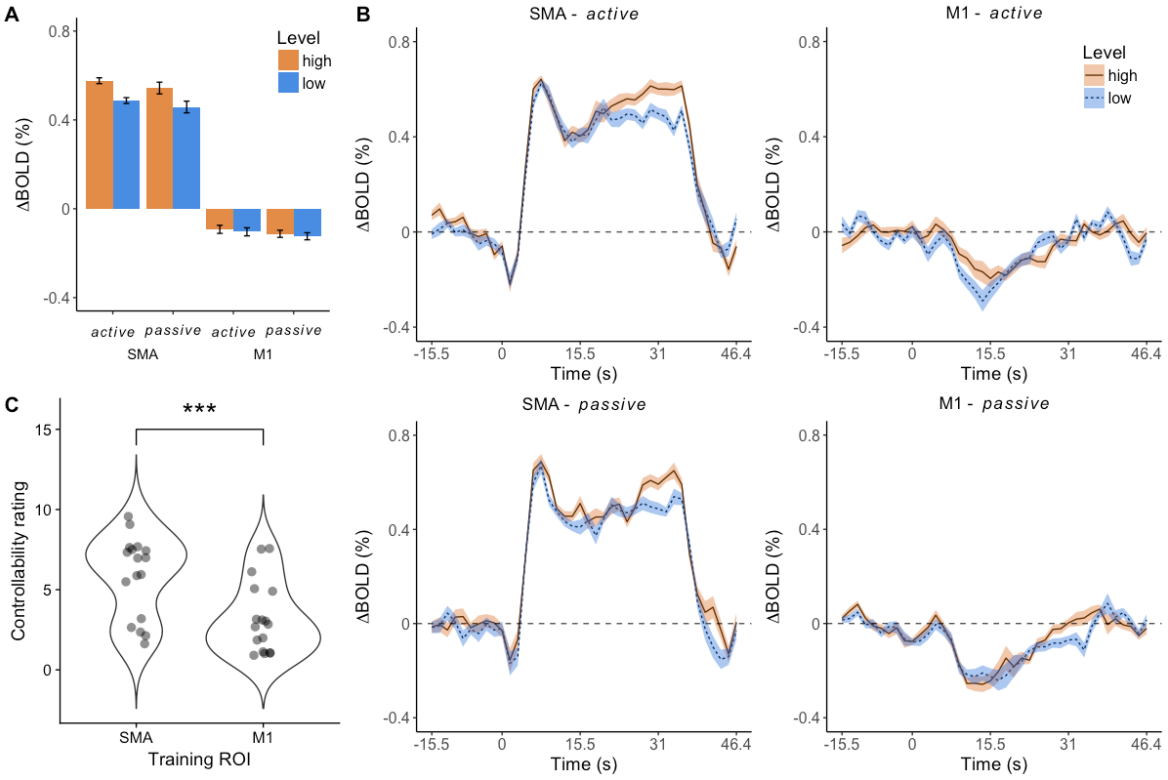

Figure 6.2 A) Bar plots for BOLD percent signal changes (PSC) in the target ROIs during the active and passive conditions (Mean ± within subject standard error). B). Event related BOLD activity in target ROIs during the active and passive conditions. Shown are group mean values and within-subject standard error around the mean (shaded). C) Violin plot showing the Controllability ratings for SMA and M1, * significant difference p<0.001. Adapted from (Mehler et al., 2019).**

Next, data were tested for differences in SMA *activation* during the low and high neurofeedback condition. SMA PSCs values during SMA (*active*) and M1 (*passive*) training were submitted to a 2 (*low* and *high* target level) by 2 (*active* and *passive* condition) repeated measures ANOVA. An effect of target level ($F_{1,16}$ = 7.334, p = 0.016, ω^2 = 0.26) was found. There was no effect for feedback condition ($F_{1,16}$ = 0.231, p = 0.637, ω^2 = 0.0), nor an interaction between feedback condition and level ($F_{1,16}$ = 0.005, p = 0.947, ω^2 = 0.0). Thus, only target level effects for SMA pooled across *active* and *passive* SMA PSCs were followed up. A significant difference of moderate strength was found (Mean difference: 0.087 ± 0.029 PSC, t_{16} = 3.006, p_{bonf} = 0.005, Cohen's d = 0.729). Likewise, a Bayesian t-test suggested strong evidence for a greater PSC during the *high* compared to the *low*-level condition (N(0, 0.55); BF_{+0} = 14.61). Yet, it should be noted that this effect was likely mainly driven by PSCs from the *active* SMA condition, as suggested by a descriptive comparisons of effect sizes (*active*: Cohen's d = 0.818, 95% Confidence Interval [0.257 to 1.361]; *passive*: Cohen's d =

0.393, 95% CI [-0.107 to 0.882]). Taken together, this study found evidence for M1 *deactivation* irrespective of the feedback condition. Further, SMA PSC data suggested a main effect of the motor imagery level, which seemed largely driven by the *active* SMA neurofeedback condition.

6.3.3 Physiological measurements

As with any BOLD fMRI study, magnitude changes can be confounded by non-neural sources of variance, such as motion and physiological noise. Further, physiological processes may be modulated by underlying mechanisms that also respond to the cognitive demands of a task, such as arousal or concentration, e.g. increased heart-rate during arousal or subconscious changes in breathing during increased mental activity. Specifically, partial pressure end-tidal carbon dioxide ($P_{ET}CO_2$) acts as a strong vasodilator, and heart-rate (HR) impacts the BOLD signal (Murphy et al., 2013). To account for potential physiological confounds (Thibault et al., 2018b), collected $P_{ET}CO_2$ and HR traces were recorded, pre-processed (see methods) and correlated with the task predictor. In general, small negative relationships were found, except for HR in M1 (Table 6.1), although these were not significant after multiple comparison correction. Also, the Bayes Factor remained inconclusive for all measurements and ROIs, suggesting that correlations between physiological confounds and the task were rather negligible.

	z (Mean ± SEM)	**t**	**df**	**p**	**p_{FDR}**	**Cohen's d [95% CI]**	**BF_{10}**
HR SMA	-0.06 ± 0.04	-1.359	9	0.207	0.277	-0.43 [-1.07 to 0.23]	1.022
C02 SMA	-0.08 ± 0.03	-2.351	12	0.037	0.148	-0.65 [-1.24 to -0.04]	2.915
HR M1	0.02 ± 0.05	0.336	10	0.744	0.774	0.10 [-0.49 to 0.69]	0.540
C02 M1	-0.07 ± 0.04	-1.920	13	0.077	0.154	-0.51 [-1.06 to 0.05]	1.765

Table 6.1 Correlations between physiological parameters and motor imagery task predictor for heart-rate (HR) and pressure end-tidal carbon dioxide ($P_{ET}CO_2$) for both training ROIs. Descriptive and inferential statistics (one-sample t-test) of z-transformed correlation coefficients. Bayesian t-test were conducted with an informed prior N(0, 0.5). Shown are Mean and Standard Error of Mean (SEM) values and p-values before and after (FDR) correction. 95% CI = 95% Confidence Interval, BF = Bayes Factor. Adapted from (Mehler et al., 2019).

6.3.4 Post-training questionnaire

Post-training questionnaire data (question 1) was analysed to test whether differences between ROIs (SMA and M1) in up-regulation were reflected in reported experiences in general (based on ratings averaged across target levels for both target regions, respectively) and for level separation (based on rating differences between the low and high target level for both target regions, respectively). Specifically, it was assessed how participants rated the controllability of the feedback during M1 and SMA training in general. Ratings scores for the low and high target level were averaged and submitted to a paired t-test, which revealed that the controllability for the SMA was rated as higher compared to M1 ($t_{16} = 4.28$, $p < 0.001$, 5.8 ± 0.6 vs. 3.2 ± 0.5, Cohen's d = 1.0; Figure 6.2 C). The Bayes Factor suggested strong evidence (N(0, 0.5); $BF_{10} = 53.5$) for this effect. In contrast, difference scores or ratings for the high and low target level suggested no effect of target region ($t_{16} = 0.108$, $p < 0.915$, 0.12 ± 0.37 vs. 0.06 ± 0.40, Cohen's d = 0.026). A Bayes t-test suggested slightly more evidence for a null effect, but remained inconclusive (N(0, 0.5); $BF_{01} = 2.28$). Taken together, average ratings of controllability showed that participants perceived being more in control during the SMA neurofeedback training compared to M1 training. The data further suggested that differences between the low and high target level did not differ between target regions.

6.4 Results – exploratory analyses

6.4.1.1 Training effect over time

To test for an effect of time in level separation, low level and high-level PSC values of the SMA were submitted to a repeated measures ANOVA for each *active* and *passive* SMA run. A Mauchly's test indicated violation of sphericity for the run level ($W = 0.183$, $p = 0.024$) and hence Greenhouse-Geisster correction was applied. No main effect of time was found ($F_{2.21, 28.72} = 2.281$, $p = 0.116$, $\omega^2 = 0.082$), while a main effect of target level was preserved ($F_{1,13} = 14.612$, $p = 0.002$, $\omega^2 = 0.476$), which also survived correction for multiple comparison after follow-up ($t_{16} = 2.540$, $p_{bonf} = 0.012$, Cohen's d = 0.679, mean difference high vs. low 0.066 ± 0.026 PSC). Importantly, the interaction of interest (run ✻ target level ✻ feedback condition) was not significant ($F_{2.32}, 30.15 = 0.491$, $p = 0.644$, $\omega^2 = 0.0$).

6.4.1.2 Comparison of other self-regulation experience ratings

In line with ratings regarding overall controllability, the SMA was rated as the better region with regards to the perceived ability to fill up the thermometer bars ($t_9 = 6.89$, $p < 0.001$, $3.5 \pm$

0.7, vs. 8.7 ± 0.4, Cohen's d = 2.2), and to maintain bars at target levels ($t_9 = 5.16$, $p < 0.001$, 5.2 ± 0.6, vs. 8.6 ± 0.4, Cohen's d = 1.6). For both effects Bayesian t-tests suggested strong evidence (N(0, 0.5); $BF_{10} = 52.08$ and $BF_{10} = 24.29$, respectively). Hence, also these additional measures of self-regulation experience confirmed that participants overall felt in high control of the feedback display when the SMA was the target ROI, which widely exceeded the perceived level of control during M1 training.

6.4.1.3 Counterbalancing and potential order effects

Random assignment yielded a similar number of participants (N = 9 vs. N = 8) in each starting condition (defined by the training ROI M1 or SMA, respectively, from which feedback was provided during the first training 5 training runs).

Next, average controllability ratings were split by starting condition and submitted to post-hoc independent two-sided t-tests, which were not significant for the SMA ($t_{15} = -0.076$, $p = 0.940$; M1 start: 5.78 ± 0.88 vs. SMA start: 5.88 ± 0.93) or M1 ($t_{15} = -1.185$, $p = 0.254$; M1 start: 2.61 ± 0.41 vs. SMA start: 3.88 ± 1.04). Bayesian t-tests suggested slightly more evidence for a null effect, but remained inconclusive for SMA (N(0, 0.5); $BF_{01} = 2.377$) and M1 (N(0, 0.5); $BF_{01} = 1.489$) ratings.

To test whether SMA *passive* PSC values differed between starting conditions, difference scores of participants' median PSC values were split accordingly (by starting condition) and submitted to an independent t-test, which remained inconclusive ($t = -0.552$, $p = 0.589$, Cohen's d = - 0.268; $BF_{-0} = 0.991$).

Taken together, no systematic differences between for self-regulation experience or level separation were detected between participants with different starting conditions and Bayesian tests remained inconclusive. Hence, relevant systematic biases due to an order effect could be ruled out.

6.5 Discussion

The main finding was that kinaesthetic motor imagery did not elicit a positive BOLD response in the M1 region of motor cortex, irrespective of whether of feedback information was provided. In fact, a significant negative BOLD signal change was observed for both conditions. This result has considerable implications for motor imagery-based fMRI-NF because it suggests that up-regulation of the M1 BOLD signal is precluded by a negative activation during motor imagery. The inability of participants to up-regulate M1 BOLD is also reflected in subjective ratings of controllability, which were low. This observation has wider implications for any fMRI-NF paradigms that employ specific cognitive strategies, because it highlights the critical importance of selecting regulation objectives that are achievable with a given choice of target region and cognitive strategy.

A second target area, the SMA, was also trained in separate runs, and consistent with previous reports (Subramanian et al., 2016, 2011), participants did show *activation* during motor imagery. However, this finding is not sufficient to demonstrate volitional control, because it is established that the SMA is *activated* during motor imagery in general, without feedback being necessary. The proposed benefit of neurofeedback as a therapeutic tool is based on the rationale that participants can self-regulate their own brain activity, and that this volitional control can be used to modulate brain states that can lead to a clinical benefit. This requires demonstrating genuine neurofeedback effects that extend beyond mere motor imagery training effects, and mere psychosocial effects (Thibault et al., 2018b). The strength of the present study lies in the use of a graded NF paradigm, because the targeting of discrete BOLD response magnitudes presumably requires some degree of self-regulation that would benefit from a feedback loop. A main effect of feedback level on SMA PSC values was found, which demonstrated significantly higher BOLD responses during in *high* compared to *low* levels, but no interaction between target level and feedback condition. Moreover, this study addressed the concern of potential physiological confounding factors (Thibault et al., 2018b) and showed that P_{ET} CO_2 and HR did not correlate significantly, and numerically only to a small degree with self-regulation blocks (Table 6.1). Furthermore, besides reported effect sizes, the present study examined the shape of the activity profiles for all tested ROI. Thus, this approach does not only go beyond mere reporting of statistical results, which has been criticized with regards to fMRI work in general (Chen et al., 2017). Although not explored in this analysis, this approach allows for the evaluation of BOLD dynamics during the training period, which may for instance reveal

further information about training success and can be explored with regards to their relationship with reported mental imagery strategies.

The role of M1 in motor imagery is likely complex and context dependent and subject of an active debate (for a review, see Hétu et al., 2013; Munzert et al., 2009). For instance, an Activation Likelihood Estimation (ALE) meta-analysis based on 73 published fMRI motor imagery studies found no conclusive results for M1 *activation* and noted that only 22 out of 122 neuroimaging studies (fMRI and PET) reported M1 *activation* (Hétu et al., 2013). The authors concluded that imprecise motor imagery instructions may explain some of the inconsistency of M1 findings reported in the motor imagery literature, and it further been suggested that only kinaesthetic motor imagery may elicit M1 *activation* (Blefari et al., 2015; Solodkin et al., 2004). This study controlled more rigorously for both these factors by providing clear instructions and instructing participants to exclusively use kinaesthetic motor imagery of actions that involved both hands. The use of SMA as a separate ROI, and the demonstration of a robust positive BOLD response in this area, also served as a control condition, and strongly negates any notion that participants were simply unable to engage in motor imagery. It has been shown that M1 (hand knob) could not be *activated* despite reinforcing feedback, a finding that is in line with previous unsuccessful M1 fMRI-NF attempts (Berman et al., 2012; Blefari et al., 2015; Chiew et al., 2012). However, the present study only included a single session and it thus remains possible that participants can *activate* M1 and gain volitional self-regulation with more training experience, for instance after multiple sessions of M1 up-regulation fMRI-NF.

One potential confounding factor of motor imagery paradigms are movements or subtle muscle contractions, which could lead to spurious activations within regions of interest (Thibault et al., 2018b). The few studies that have reported M1 *activation* during motor imagery mostly did not control (Lotze et al., 1999; Perronnet et al., 2017; Sharma et al., 2008) for overt movements and muscle contractions, for which Electromyography [EMG] recordings are required, that remain technically challenging in the MR environment. Likewise, this study did not record EMG to rule out muscle activity and this limitation should be addressed in future studies. Noteworthy, the within subject design and comparison of separate ROIs limits the likelihood that participant motion (either voluntary or involuntary) contributes in any systematic way. Importantly, the main finding of M1 *deactivation* during motor imagery periods strongly suggests that the presented results were likely not confounded by participants' overt movement

production, because these would result in a positive BOLD response in the M1 target region whose voxels were selected based on a motor execution localiser.

One potential mechanistic explanation for the observed M1 deactivation during motor imagery could relate to the suppression of overt movements. Invasive electrophysiological stimulation studies have shown that M1 can have suppressive effects on muscular activity in motor control (Ebbesen and Brecht, 2017). In the current motor imagery paradigm participants were required to avoid muscular activity, and it is plausible that inhibition of movement is realised in the primary motor cortex, which acts as the final station where motor commands converge before being executed. Previous fMRI motor imagery studies have reported inhibitory projections from the SMA to M1 in healthy participants (Kasess et al., 2008; Solodkin et al., 2004) and stroke survivors (Bajaj et al., 2015), which could explain the M1 BOLD deactivation reported here. Moreover, one motor imagery study conducted in stroke survivors reported an inverse relation between (self-reported) motor imagery and activity of M1, indicating that individuals with higher motor imagery capacity showed more suppression of M1 activation (Confalonieri et al., 2012).

Electrophysiological studies suggest that M1 is involved in motor imagery and can provide further insight into the neural origin of M1 BOLD *deactivation*. For instance, one magnetoencephalography study reported activity in the hand area of M1 during motor imagery (Schnitzler et al., 1997). More direct evidence for the involvement of M1 in motor imagery stems from a electrocorticography (ECoG) study that reported decrease of power (K. J. Miller et al., 2010) within lower frequencies reported for motor imagery (Pfurtscheller and Aranibar, 1979; Pfurtscheller and Neuper, 1997), including the Mu-rhythm (8-13 Hz). Similar findings have been reported in BCI that could demonstrate training effects in motor imagery paradigms (for a review, see Cervera et al., 2017). A power decrease of the Mu-rhythm during unilateral hand motor imagery has further been associated with increased spinal motor neuron excitability (Takemi et al., 2015), as well as increased excitability of the corticospinal tract (CST) and intracortical GABAergic inhibition (Takemi et al., 2013). Simultaneous EEG-fMRI recordings have linked the Mu-rhythm to a decrease in BOLD *deactivation* (Mullinger et al., 2014). Taken together, some electrophysiological correlates of motor imagery anti-correlate with BOLD and may hence partly explain the presented M1 findings.

The second aim of this study was to test if participants could exert volitional control over the BOLD signal, in either SMA or M1 during kinaesthetic motor imagery, to target two discrete BOLD signal change magnitudes. A significant main effect of target level for SMA PSCs was found, but no interaction between target level and feedback condition (active vs. passive). Thus, although there was a significant difference between SMA PSC in high vs. low target level conditions, this effect was not significantly modulated by the feedback condition. However, a comparison of effect sizes suggested that the level effect was mainly driven by the active condition. It remains possible that the present sample size was too small to detect a rather small difference. Overall, it appears that graded fMRI-NF does aid participants in achieving discrete levels, but the specific neurofeedback effect of volitional control of BOLD signals was relatively small compared with the general motor imagery effect. Besides a larger sample, more neurofeedback training sessions may be required to demonstrate that participants benefit from feedback by utilizing this information to achieve self-regulation of discrete BOLD target levels. With regards to mere up-regulation training of the SMA (Sepulveda et al., 2016) or sensorimotor areas in general (Auer et al., 2016), previous work suggest that multiple training sessions of motor imagery based fMRI-NF are necessary to achieve pre-post differences for motor imagery transfer runs. Future work may compare the efficiency of such paradigms and graded fMRI-NF training regarding their efficiency in demonstrating superiority of training over control conditions.

Noteworthy, a previous study that investigated (3-level) graded fMRI-NF for different mental imagery tasks reported superior self-regulation effects within a single training session when comparing fMRI-NF to a pure mental imagery session (Sorger et al., 2016). However, one main difference between the study designs lies in the selection of target regions: whereas the present study used a motor execution task as a localiser, Sorger et al (2016) used a mental imagery task to identify target regions and maximum PSC to scale the feedback, and thus possibly targeted voxels that are more specifically involved in the respective mental imagery task. Further, LOC PSC values were on average larger than in the present study. Although it remains speculative, participants may in consequence have had more dynamic range to exploit of the BOLD signal. This notion seems supported by a recent BCI study that used a graded feedback approach and found a positive relationship between accuracy of self-regulation and maximal localiser PSC values that were used to scale the individual feedback (Krause et al., 2017).

This study has provided new insights into neural correlates of (supervised) motor imagery, which can inform future attempts of translating fMRI-NF to clinical applications. In particular for stroke rehabilitation, fMRI-NF may provide new means to enhance neuroplasticity (Wang et al., 2017). One possible target region could be M1, which has been intensively studied in brain stimulation based stroke rehabilitation (McDonnell and Stinear, 2017). Considering the presented findings, however, M1 is likely a suboptimal target region for motor imagery-based fMRI-NF up-regulation training, although it remains possible that M1 activation can be achieved with longer training. We note that our results are largely in line with findings from Hanakawa and colleagues who had compared motor networks during motor planning, motor imagery and motor execution (Hanakawa, 2016). The authors had classified brain areas as "movement-predominant", which included the M1 and SMA, and as "imagery-predominant", which included the SMA but not M1. Further, a direct comparison between tasks suggested that associated motor networks largely overlapped, but also that voxels that were more responsive to motor execution were located more anterior within the SMA and the precentral gyrus. In contrast, voxels that were more responsive to motor imagery were located more posterior within these areas. This observation may be of relevance for the interpretation of the presented findings, because training voxels were identified using a motor execution localiser. To achieve optimal voxel selection future studies may benefit from employing a motor imagery-based localiser scan. One alternative approach to test if M1 activation can be learned could be realised in a design that does not use explicit motor imagery strategies whilst controlling for movements. Another set of techniques that could yield control over M1 is multivariate decoded fMRI-NF (DecNef). This technique combines statistically more sensitive multivariate decoders and implicit fMRI-NF, allowing to decode and train brain without participants' conscious awareness (Koizumi et al., 2016; Watanabe et al., 2017). Such approaches may circumvent possible interferences due to a dual task situation and may allow attaining self-regulation of M1 activity. Studies that have employed multivariate decoding techniques to pure motor imagery tasks could decode different imagined actions from M1 (Park et al., 2015; Pilgramm et al., 2016; Zabicki et al., 2017). Apart from motor imagery-based paradigms, M1 remains a promising target region for motor execution-based fMRI-NF training for patients who possess sufficient residual ability to execute motor tasks. The literature on motor execution-based fMRI-NF training is currently very limited, but first evidence in healthy participants suggests that the M1 hand knob may provide a suitable target for such paradigms (Neyedli et al., 2017).

The presented findings are confined to young healthy participants and may not generalize to patient populations. For instance, stroke survivors show altered motor networks (Rehme et al., 2011b; Rehme and Grefkes, 2013), implying that motor areas in the contralesional hemisphere can partly compensate for motor impairment. Furthermore, inter- and intrahemispheric network activity alters over time, hence optimal target regions may also depend on the time after stroke, in addition to other factors such as the lesion location (e.g. hemisphere), and involvement of the CST (Stinear et al., 2007). For instance, individuals with right-hemispheric stroke seem to show more frequently impairment of motor imagery than individuals with left-hemispheric stroke (Kemlin et al., 2016).

With regards to the SMA as a target region, besides its role in motor control, it is involved in vigilance more generally (Cunnington et al., 2002; Hinds et al., 2013), as well as attention to timing (Coull, 2004; Macar et al., 2006). Both cognitive processes are likely recruited during the attempt to self-regulate brain activity. For instance, participants likely monitored their self-regulation performance and attended to the temporal delay of the visual feedback display. Of interest, a previous real-time fMRI study used activation in the SMA to trigger the presentation of stimuli that required subjects to respond with a button press (Hinds et al., 2013). The authors found that that stimuli preceded by SMA activation were associated with shorter reaction times compared stimuli that were triggered by a control target in the default mode network, underscoring the role of SMA in controlling vigilance. Besides these cognitive processes that may have occurred during self-regulation attempts, also cognitive strategies of some participants during rest periods may have engaged the SMA as an area that is generally involved in mental imagery. One recent study for instance found that the SMA could be up-regulated (albeit to a smaller extent, compare PSC 0.4 to 0.7 in the present study to on average only 0.1 to 0.15 in the G_F condition of Sepulveda et al., 2016) when no explicit imagery instructions were provided and participants used mental strategies comparable to those commonly used during rest blocks (Sepulveda et al., 2016), including relaxation and recall of autobiographical information. Taken together, although the current design controlled for general mental imagery effects during training blocks by comparing PSCs obtained from the SMA *active* and the SMA *passive* condition, other cognitive processes that involved the SMA and occurred either during the training or rest blocks may have interfered with training success. It thus remains to be tested whether self-regulation success in motor imagery based graded fMRI-NF could be increased by targeting other premotor areas (e.g. vPMC), using different a form (e.g. intermittent, Hellrung et al., 2018) or modality (e.g. auditory) of feedback

presentation, or employing a different feedback type (e.g. connectivity based feedback, Liew et al., 2015).

To conclude, this study has demonstrated that M1 shows a robust negative BOLD response during kinaesthetic motor imagery in the context of NF. This finding may explain previous unsuccessful attempts of M1 fMRI-NF, suggesting that it represents a suboptimal target region for *up-regulation* paradigms in which participants use explicit kinaesthetic motor imagery. Conversely, premotor cortical regions such as SMA are more suitable target regions because of their more active role in motor imagery in general. However, the robust reactivity of premotor areas like SMA to different forms of motor imagery also presents a confound in NF studies that aim to develop volitional self-regulation of neuronal activation, as additional evidence beyond simple positive BOLD responses is required to demonstrate volitional self-regulation.

7 GRADED FMRI-NF TRAINING OF MOTOR IMAGERY IN MIDDLE CEREBRAL ARTERY STROKE PATIENTS – A PREREGISTERED PROOF-OF-CONCEPT STUDY

7.1 Introduction (preregistered)

Ischemic stroke of the middle cerebral artery (MCA) is one of the most common forms of stroke (Aouad et al., 2013; Leys et al., 1992). The MCA is the main blood supply to the primary motor cortex, including the hand knob area, as well as the premotor cortex. Hence, MCA stroke often leads to severe upper limb impairment and compromises patients' quality of life (Langhorne et al., 2011; Miller et al., 2010). The efficacy of current rehabilitative strategies is limited and most patients remain impaired such that more than 70% of stroke survivors show residual upper limb dysfunction, and in many cases to a moderate-to-severe degree and despite intensive rehabilitation (Lawrence et al., 2001). Further, it is estimated that only 5-20% of patients regain full upper limb function, whereas 33-60% have not showed any recovery six months after stroke (Kwakkel and Kollen, 2013). This remaining limb dysfunction presents a major impediment to rehabilitation, activities of daily living and occupational prospects of stroke survivors, and has a considerable, negative effect on their wellbeing (Langhorne et al., 2011; Pollock et al., 2014). Therefore, there is a need for new non-invasive therapies to promote recovery of motor function in general and in particular for the upper limb after MCA stroke. Several motor imagery-based interventions have been suggested (Ietswaart et al., 2011; Sharma et al., 2009b), which may enhance neuroplasticity and thus potentially facilitate recovery (García Carrasco and Aboitiz Cantalapiedra, 2016; Sakurada et al., 2017). Indeed, feasibility studies suggest that stroke patients can still engage in motor imagery and that regular motor imagery training combined with physical therapy is associated with improved outcome (García Carrasco and Aboitiz Cantalapiedra, 2016). Motor imagery strategies can be primarily visual or kinaesthetic, and it is possible that the type of motor imagery influences the degree of improvement in motor impairment (Jackson et al., 2001; Sharma et al., 2009b). Whereas visual motor imagery focuses primarily on visual mental imagery, either from a first or a third person perspective, kinaesthetic motor imagery is defined as motor imagery from the first-person perspective and involves imagining the feeling and experience of movements without overt movement. Studies conducted with Transcranial magnetic stimulation (TMS) (Stinear et al., 2006) and functional Magnetic Resonance Imaging (fMRI) (Guillot et al., 2009; Sharma and Baron, 2013; Solodkin et al., 2004) suggests that kinaesthetic motor imagery recruits motor areas including the supplementary motor area (SMA), and that this type of imagery is thus of interest for motor rehabilitation. Moreover, robust involvement of the SMA in kinaesthetic imagery was confirmed by a meta-analysis using activity likelihood estimation (Hétu et al., 2013). The same meta-analysis also found that in majority of the 75 included motor imagery

studies did not specific the motor imagery type and that instructions were reported inconsistently between studies (Hétu et al., 2013). This renders comparability and interpretation of outcome measures for rehabilitation motivated motor imagery interventions difficult. Further, for the study of neural correlates during motor imagery after stroke, knowledge of the location is important to assist in the interpretation of activation patterns observed in individual patients. Most motor imagery studies conducted in stroke have included patients with a variety of lesion sites (Ietswaart et al., 2011; Kraft et al., 2015; Sharma et al., 2009b).

Besides type (e.g. visual and kinaesthetic), motor imagery may also be distinguished in terms of supervision (i.e. unsupervised vs. supervised motor imagery paradigms). For unsupervised motor imagery participants usually receive imagery instructions only, whereas for supervised motor imagery, participants are also provided with feedback on their performance. Brain computer interfaces (BCIs) including real-time (rt-) fMRI neurofeedback (-NF) training have been suggested to support motor imagery and mental imagery ability in general. For motor rehabilitation, feasibility of rt-fMRI-NF has been demonstrated for Parkinson's disease (PD), and case studies suggest that stroke patients can also learn to self-regulate motor areas using motor imagery (Liew et al., 2015; Sitaram et al., 2012; Subramanian et al., 2016, 2011).

In this proof-of-concept (PoC) study, we aim to address some of the limitations identified in the literature and aim to show that motor imagery-based rt-fMRI-NF training is feasible in MCA stroke patients. First, we will test if stroke patients with an MCA lesion can achieve sustained SMA activation using supervised kinaesthetic motor imagery (using rt-fMRI-NF). The SMA is an attractive target region of interest (ROI) in MCA stroke patients, because it is supplied by the anterior cerebral artery (ACA) (Brugger et al., 2015), and thus is not expected to have (substantially) compromised haemodynamic function following MCA stroke. We therefore expect that MCA stroke patients can activate the SMA. Second, we will test if MCA stroke patients can self-regulate the feedback signal to two discrete target levels. Such 'gradual' neurofeedback training (referred to as 'graded' in this study) has recently been demonstrated in healthy participants for various mental instructions including mental rehearsal (Sorger et al., 2016), mental arithmetic (Krause et al., 2017) and kinaesthetic motor imagery (Mehler et al., 2017c).

7.2 Methods (preregistered)

7.2.1 Design and materials

7.2.1.1 Participants

This will be a PoC study to ascertain if unilateral MCA ischemic stroke patients, with moderate to severe upper limb impairment, are able to 1) achieve sustained SMA activation during motor imagery and 2) demonstrate control over the feedback (by self-regulating the activation to different target levels), which we see as prerequisites for the use of graded neurofeedback training as motor rehabilitation therapy. Stroke severity will be determined by the local clinician and based on the Modified Rankin Scale Questionnaire Score of 3 to 4 (Bonita and Beaglehole, 1988). Potentially suitable participants, within 6 months of stroke, will be identified by clinical staff at Cardiff and Vale University Health Board. The inclusion criteria for patients are:

1) infarct of the middle cerebral artery territory
2) persisting hemiplegia/hemiparesis
3) patients will have either a) have been discharged from the acute stroke unit and received ambulatory physiotherapy (so called *early supported discharge*), b) are still undergoing or c) have already completed an inpatient rehabilitation programme
4) no receptive aphasia to ensure that task instructions will be properly followed

Exclusion criteria are any MRI contraindications. Patients will provide informed consent. [Amendment to original preregistration protocol: The study has been approved by local research ethics committee (Wales REC3, reference number 16/WA/0167) operated by Health and Care Research Wales.] For a detailed list of exclusion criteria that affect the data acquisition and/or quality, please see *Data exclusion* section below. Although exclusions have been anticipated, any further expulsions and reasons will be reported.

7.2.1.2 Data exclusion

Neurofeedback runs that contain too much head movement, defined by >30% volumes with a framewise displacement (FD) > 0.5mm (Power et al., 2014). If debriefing suggests that patients have misunderstood instructions, data of affected runs will be excluded. Further, if patients do not attend the second neurofeedback session and do not withdraw their consent, acquired data from the first neurofeedback session that meets the quality criteria as specified above will still be included in the Sequential Bayes Factor sampling and all other planned analyses.

If technical problems occur that either disrupt or delay the feedback presentation, data from affected runs will be excluded. Incomplete neurofeedback runs are excluded.
If patients interrupt a scan session but consent is not withdrawn, acquired data of this session will be included. However, in the debriefing it will be established when a patient stopped complying with the task and affected runs will be excluded. Other unanticipated circumstances effecting data quality can result in exclusions. In such circumstances, all exclusions and reasons will be reported.

7.2.1.3 Dependent variables and hypotheses

The first dependent variable of interest is the median percent signal change (PSC) of the SMA-ROI during motor imagery (taking an average across the low and high neurofeedback condition) > rest (details for calculation see *Analysis plan* below). The second dependent variable of interest is the difference in SMA-ROI PSC between the low and high neurofeedback condition (details for calculation see *Analysis plan* below).

H0_A: MCA stroke patients show no sustained SMA-ROI activation
H1_A: MCA stroke patients will show sustained SMA-ROI activation as measured by a positive PSC.

Such a finding would provide a basis for future motor imagery based rt-fMRI-NF training in MCA stroke patients.

H0_B: MCA stroke patients show no difference in SMA-ROI activation (measured by PSC) between low and high neurofeedback conditions.
H1_B: MCA stroke patients show a difference in SMA-ROI activation between low and high neurofeedback conditions.
Such a finding would suggest that rt-fMRI neurofeedback training enables MCA stroke patients to self-regulate the BOLD signal of the SMA activity to two discrete target levels using kinaesthetic motor imagery.

7.2.1.4 Sample size and sampling plan

Owing to feasibility and PoC, we intend following a Bayesian sampling strategy with a minimum of N = 5 patients and continue recruiting either until the Bayes factor for both hypotheses (A and B) is conclusive - i.e. either for the alternative with $BF_{10} > 10$ (indicating strong evidence for a positive effect) or for the null with $BF_{01} > 10$ (indicating strong evidence for a null effect) - or until the end of the data collection period (September 30, 2017) is reached.

Effect size	Power
0.6	0.30
0.8	0.44
1.0	0.58
1.2	0.71
1.4	0.82

Table 7.1: Power calculation based on sample size of N = 5 and alpha level of 0.05, one-tailed.

Although there is risk for inflated type-II error rates with the present small sample and min N = 5 (Table 7.1), we believe that preregistering the protocol of this PoC study outweight concerns about power. Importantly, preregistration of this stepwise Bayesian approach effectively addresses several concerns of standard null hypothesis testing by (1) reducing the risk of inflated type-II errors associated with small sample size, (2) eliminating publication bias, and (3) ensuring that only *a priori* hypotheses are declared as confirmatory.
Moreover, a preregistered PoC study, if successful, provides a solid basis for future replication in larger clinical trials. Lastly, we note that this is to our knowledge the first preregistered report of a rt-fMRI-NF study that is not a formal randomized controlled trial.

7.2.2 Procedure

7.2.2.1 Familiarization

Patients will be invited to two rt-fMRI-NF training sessions. Sessions will be conducted at the patient's convenience but within six months after their stroke. If necessary, patients will be made familiar with the MR scanner environment using a mock scanner before their scanning session.

Mental imagery performance has been shown repeatedly to be modulated by hand orientation (Jongsma et al., 2013) as well as body posture (de Lange et al., 2006; Ionta et al., 2012). Thus, patients will be asked to identify a kinaesthetic motor imagery strategy that involves both hands (see Chapter 7.2.2.6). Patients will be reminded to avoid any movements and muscle contractions. During the scanning, patients will be asked to use this motor imagery strategy during the localiser run. The localiser run is used to identify voxels activated during the task within an atlas-based ROI of the SMA (see Chapter 7.2.2.3).

7.2.2.2 MRI acquisition and online processing

Imaging data will be acquired using a 3 Tesla Siemens Prisma scanner at the Cardiff University Brain Research Imaging Centre (CUBRIC). Blood-oxygenation-level-dependent (BOLD) signals during localiser and neurofeedback runs (see Procedure) are measured with a T2*-weighted gradient echo planar imaging (EPI) sequence that is synchronised to the onset of the stimulus presentation. Functional EPI volumes of 24 slices of 2.5-mm thickness, with 0.5-mm inter-slice spacing will be used (in-plane resolution = 3 mm, TR = 1500 ms, TE = 30 ms, flip angle = 80°). High-resolution structural images will be acquired before the first functional scan using a magnetization-prepared rapid gradient-echo sequence (MPRAGE) T1 weighted image with 172 contiguous sagittal slices of 1-mm thickness (voxel size: $1 \times 1 \times 1$ mm, TR = 7.9 s, TE = 3.0 ms, flip angle = 20°, FoV = $256 \times 256 \times 172$ mm). Turbo BrainVoyager (TBV) software (BrainInnovation, Maastricht, The Netherlands, version 3.2) will be used for online analysis of BOLD signals including motion correction with respect to the first volume of the functional localiser and spatial smoothing (4mm full width at half maximum; FWHM). The functional EPI data will be co-registered with the T1-weighted anatomical scan and transformed to Talairach space to allow for standardised ROI selection during the localiser. The functional localiser run (180 volumes) will consist of 4 blocks (30sec) kinaesthetic motor imagery, flanked by rest blocks (30sec).

7.2.2.3 Functional localiser

For the localiser, an incremental GLM will be used including a task predictor and linear drift term (Version 3.0, Brain innovation, Maastricht, The Netherlands). The localiser run serves to identify most active voxels and to calculate the individual percent signal change (PSC_{LOC}) that is used to scale the visual feedback.

Previous real-time neurofeedback studies that were conducted with neurological patients have either used motor execution (Subramanian et al., 2016, 2011) or action observation (Sitaram et al., 2012) task during localisers runs. However, given that motor execution is impaired in stroke patients in general and to different degrees across patients, a motor execution localiser would be difficult to implement and most likely provide unreliable estimates for PSC_{LOC}. We therefore decided to use a motor imagery localiser, which further allows to identify voxels that are most activated during the task of interest (i.e. motor imagery). The voxel selection will be constrained to an anatomical ROI of the SMA (see details in *Anatomical SMA-ROI*). To do so, statistically significant voxels are selected using a t-contrast of motor imagery > rest with a variable t-threshold to ensure that a sufficient a number of voxels are available within the SMA for the following selection. The 40 most active neighbouring voxels (determined by their t-value) in four neighbouring slices will be identified based on this t-contrast using a custom-made plugin (*Best Voxel Tool* with settings 4-4-1-0-40; plugin will be available on OSF). To give participants time to disengage from motor imagery during the rest and allow time for the BOLD to recover to baseline, only the second half of each rest period will be considered for the calculation using a custom-made PSC plugin too (*PSC calculation Tool*).
We will use 80% of the maximum PSC_{LOC}. However, because it is possible that patients initially struggle to engage in motor imagery when no feedback is provided yet, a lower bound $PSC_{LOC} = 0.7$ will be used to avoid underestimating an appropriate PSC_{LOC} that will scale the feedback. Further, because the maximum PSC value could be biased by outliers or spikes in the time series, an upper bound $PSC_{LOC} = 1.4$ will also be used. These values are comparable to PSC_{LOC} default values of 1% that have previously been used in neurological patients (Subramanian et al., 2016, 2011).

The PSC_{LOC} value will determine the 100% level of the thermometer feedback and the rest of the thermometer is linearly scaled accordingly (see *Neurofeedback calculation and presentation*). In the event that the functional localiser run fails (e.g. due to technical reasons or because the patient is not able to perform the task initially without any visual feedback), we will select 40 adjacent voxels within the anatomically defined SMA as a default option and set the $PSC_{LOC} = 1$.

7.2.2.4 Anatomical SMA-ROI

To guide voxel selection and standardize the SMA-ROI across participants, we decided to use an anatomical ROI template, which requires a custom-built workflow. An atlas template

(Automated Anatomical labelling; AAL) (Tzourio-Mazoyer et al., 2002) will be extracted from the PickAtlas tool (Maldjian et al., 2004, 2003). Further, the corresponding T1-weighted image (i.e. collin27) that these structures have been co-registered to will be extracted from the SPM Anatomy toolbox (Eickhoff et al., 2005). The resulting NIfTI images require conversion into a data format that is readable by TBV. Because the conversion via the Brainvoyager NIfTI-1 converter plugin (v1.09) requires identical image dimensions (i.e. bounding boxes), the SMA-ROI will first be sampled to the T1-weighted image using Nilearn (version 0.3.0; resample_to_img function) (Abraham et al., 2014) and then converted to BrainVoyager anatomical format. Next, images are transformed from MNI to Talairach space using BrainVoyager (BrainInnovation, Maastricht, The Netherlands, version 20.4). Lastly, the obtained anatomical file is transformed into an SMA-ROI and thresholded (at minimum value range = 50) using Brainvoyager for use in TBV.

7.2.2.5 Neurofeedback calculation and presentation

The localiser run is followed by 5 neurofeedback runs (180 volumes) which contain two repetitions of two block types, a low and high neurofeedback level which are interleaved by rest (30 sec). For feedback presentation, the mean raw BOLD value will be extracted from the SMA-ROI ($BOLD_{SMA\text{-}ROI}$). The feedback will be computed based on equation 1,

Eq. (7.1): $$PSC\_NF = \frac{(val - baseline) * 100}{baseline}$$

where *val* is the mean of three consecutive $BOLD_{SMA\text{-}ROI}$ values, *baseline* is the median $BOLD_{SMA\text{-}ROI}$ during the second half (i.e. last 10 TRs) of the preceding rest period and PSC_NF is the resulting percent signal change. PSC_NF is then normalised by PSC_{LOC} to map it on to the (15) segments of the thermometer display such that every segment represents 10% of the PSC_{LOC}. Values below 0 are rounded up to 0, values above 15 will be rounded down to 15. [Amendment to original preregistration protocol: Given the relatively small number of data points (N=10), the median was considered more robust than a mean to protect against potential outliers due to signal variations (e.g. in consequence of abrupt head motion)]. The calculation is carried out in an in house written python script (Python 2.7.10). The Open Source Python library Expyriment will be used for online feedback presentation (Krause and Lindemann, 2014). Figure 7.1 shows the visual feedback display which consists of a thermometer with 15 segments and indications for either rest (i.e. red arrows at the bottom) or the target levels. Target levels are defined by 50% (low) and 100% (high) of the localiser SMA-ROI PSC and

indicated by green arrows. Both the low and high-level conditions are repeated twice per run and are interleaved by rest periods. During feedback periods activation of the SMA-ROI compared to rest results in filling of the bars. During rest periods, no feedback is presented, and the thermometer remains empty. The order of the condition (low and high target level) will be counterbalanced across runs and subjects (determined by subject number and run number).

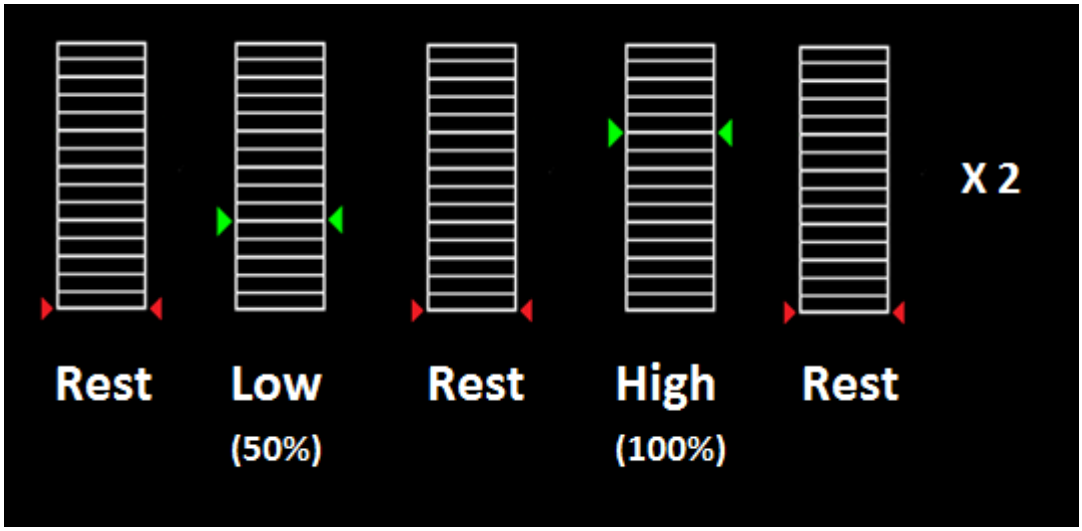

Figure 7.1 Example sequence of conditions. Rest indicated by red arrows, target levels during neurofeedback indicated by green arrows. Empty fields fill with grey bars (not shown here) when Feedback (Eq. 7.1) is a positive value.

7.2.2.6 Instructions

Before the scan, patients will be asked to identify a kinaesthetic motor imagery strategy that involves both hands and that they can comfortably, vividly and consistently perform for about 30 seconds (e.g. an activity of daily living) while lying in a supine position comparable to the actual scan session. Patients are further instructed to avoid any movements and muscle contractions.

During the localiser run, patients will be asked to use this motor imagery strategy. Patients will be presented with an empty thermometer on the screen which contains coloured arrows at both sites (Figure 7.1). Patients are instructed to remain still in the scanner and rest while shown arrows are red and to perform vivid kinaesthetic mental imagery of an action that involves both of their hands during indicated task periods. Patients are reminded that no feedback will be presented yet.

For both the localiser and neurofeedback runs, patients are further instructed that during scans they should 1) remain still and relaxed, 2) avoid movements and muscle contractions and 3) use only kinaesthetic motor imagery that involves both hands. Besides these aspects, patients are not restricted in the content of motor imagery (e.g. a particular type of activity or sport).
For the neurofeedback runs, it is explained to patients that bars in the thermometer will represent the activity level in the target region and that their goal is to use kinaesthetic motor imagery to control the feedback by filling up the bars contained in the thermometer display. They are also instructed to maintain the activation at target levels by adjusting their mental strategy (e.g. changing speed and/or intensity of the imagined movement).

7.2.3 Analysis plan

7.2.3.1 offline fMRI analysis

The present study is a repeated measure within-subject design. The dependent variable of interest is the PSC of the SMA-ROI in the contrast task - i.e. supervised kinaesthetic motor imagery based on rt-fMRI-NF from the SMA-ROI > rest. The open source software AFNI (version 16.2.18) will be used for offline fMRI ROI analyses using the same pre-processing parameters (motion correction, 4mm spatial smoothing) as used for the online feedback. Based on existing in-house scripts, AFNI functions 3dDeconvolve and 3dREMLfit function will be used. 3dDeconvolve will be used to calculate task and baseline predictors (see below) and pre-processed BOLD time series and design matrices will be submitted 3dREMLfit to compute SMA-ROI PSCs. This approach will largely replicate the analysis carried out online and additionally correct for temporal auto-correlation (AR1) of the BOLD time series. For the functional localiser, the intercept, drift and task are modelled, and the PSC is defined as the ratio between task and intercept parameter estimates. For the neurofeedback runs, analyses will be carried out on a concatenated time series for both the low and high target level with their respective preceding rest periods such that five coefficient estimates are returned for: the 1) intercept, 2) low level task block, 3) high level task blocks, 4) baseline blocks preceding low level task blocks and 5) baseline blocks preceding high level task blocks (with order of 2-5 depending on randomization). Based on these coefficient estimates, PSCs are calculated using equation 6.2

Eq. (7.2): $$PSC = task * \frac{100}{intercept+rest}$$

for the low and high neurofeedback level, respectively.

Returned values from the neurofeedback runs will be normalized by the offline calculated PSC_{LOC}. Hence, for instance a value of 0.5 would indicate that 50% of the PSC_{LOC} have been achieved during neurofeedback, the desirable target level for the low neurofeedback condition. Both SMA activation (H1_A) and difference between target levels (H1_B) will be tested on the group level as well as in individual patients. Group analyses will be carried out based on patients' median PSC values calculated across all runs from both neurofeedback sessions. [Amendment to original preregistration protocol: The use of the median (rather than the mean) as a measure of centrality was motivated by the relatively small number of runs (N = 5) since the median is more robust against potential outliers.] To test for H1_A, the grand median is calculated from PSC (grand mdPSC) values across the low and high target level conditions. To test for H1_B, the median PSC is calculated separately for the low and high target level condition (mdPSC). In addition to the group analyses, both H1_A and H1_B will be tested within subjects based on all PSC values available for a patient.

7.2.3.2 Sequential Bayesian Sampling

Sampling will be stopped either when the period of the study is exceeded (30 September 2017) or when conclusive Bayes factors are reached for both hypotheses (A and B), i.e. either for the alternative with $BF_{10} > 10$ (indicating strong evidence for a positive effect), or the null with $BF_{01} > 10$ (indicating strong evidence for a moderate null effect).
For both, H1_A and H1_B, Bayes Factor calculations will be performed using a uniform prior (Baguley, 2010; Dienes, 2014) with the lower bound set to zero. For H1_A, we will test for activation vs. no activation and thus conduct a one-sided one-sample Bayesian t-test with the upper bound being determined by the group median of the PSC_{LOC}. For H1_B, we will test for a difference between the low and high neurofeedback level and thus conduct a one-sided paired Bayesian t-test. In the optimal case the difference would be 0.5 (see paragraph above), hence to test for H1_B the upper bound of the uniform prior will be set to 0.5.

Besides these Bayesian t-tests that will be used for the stopping rule, we will also report respective Bayesian t-tests on group level with a default Jeffreys-Zellner-Siow (JZS) prior (Cauchy r = 0.707) (Rouder et al., 2009). These will be reported together with the median of the posterior and its 95% credible interval. Given that Bayes factors are sensitive to priors, a prior robustness check (with r = 0.5, r = 1 and r = 1.4142) will be conducted to assess the robustness of the outcome (Rouder et al., 2016; Schönbrodt and Wagenmakers, 2016).

7.2.3.3 Frequentist statistical analysis

Besides Bayesian hypothesis testing, frequentist hypothesis testing will also be conducted. H1_A will be tested with a one-sided one-sample t-test and H1_B with a one-sided paired t-test for both the group analysis and individual subject analyses. Because we have directed hypotheses in the cases of H1_A and H_1B (i.e. positive differences), all tests will be carried out one-tailed. Both hypotheses will be tested on the group level as well as in individual patients. Tests conducted on group level will be carried out at a significance level of 0.05. Tests at subject level will be Bonferroni corrected for multiple comparisons and carried out at a significance level of 0.05/number of tested patients. For all measures, effects sizes (Cohen's dz) will be reported.

7.3 Changes applied to preregistered methodology

To address challenges in the data acquisition and analysis that were not anticipated when this study was preregistered (Mehler et al., 2017a), the following changes to the preregistered methodology were applied:

7.3.1.1 Participants and sessions

The patient recruitment was slower than expected. Therefore, the recruitment phase was extended from the end of September 2017 until the end of February 2018 with approval from the ethics and research governance committees. Further, inspection of real-time motion parameters during data acquisition suggested substantial head motion. Therefore, two patients completed one additional scan session to compensate for expected data loss, which was an option included in the study protocol.

7.3.1.2 Co-registration and functional localiser

Brain lesions and head motion corrupted data quality and anatomical landmarks (Appendix C Figures C1 and C2). In consequence, co-registration attempts failed. Hence, all but neurofeedback runs were performed in native space. Therefore, the created anatomical SMA mask in Talairach could not be applied and instead target voxel selection was guided by visual inspection of functional brain slices in a similar way as described earlier (Chapter 5.2.8). Target regions were mostly constrained to the 40 most active adjacent voxels as described in 6.2.2.3, however, this approach failed for two scanning sessions due to technical problems and a larger set of voxels was selected.

7.3.1.3 Co-registration and functional localiser

When this study was preregistered, data from Chapter 5 was used to inform decisions about the choice of prior distributions. Specifically, custom-written MATLAB (Mathworks Inc) scripts (Dienes, 2014) were used that included priors based on a uniform and normal distribution. These prior distributions were not yet implemented in earlier software versions of JASP during the piloting phase. However, more recent JASP software versions included user set normal priors and thus allowed comparing results between custom-made scripts and results obtained with the JASP software. Comparison between results based on normal prior distributions deviated and with respect to the Bayes Factor. Hence, to ensure reproducibility it was decided to switch to JASP (version 0.8.4.0; Team, 2017) and use normal priors. These were scaled based on localiser PSC values as described in the preregistered methods. Given that the scaling was maintained, the switch from uniform to normal priors likely results only in marginal differences (Rouder et al., 2009). This assumption was further tested for reported analyses using sensitivity analyses with JZS priors and a range of scaling factors (Chapter 3.4.2.2).

7.3.1.4 Supplementary information about head motion quantification

Head motion was quantified using framewise displacement (FD), with

Eq. (7.3): $$FD_i = |\Delta d_{ix}| + |\Delta d_{iy}| + |\Delta d_{iz}| + |\Delta \alpha_i| + |\Delta \beta_i| + |\Delta_{\gamma i}|,$$

where $_i$ indicates the volume, $\Delta d_{ix} = d_{(i-1)x} - d_{ix}$ for translation rigid body motion parameters, and $[\alpha_i\ \beta_i\ \gamma_i]$ are the three head rotation parameters (roll, pitch, yaw) that were converted to millimetres using a projection onto a sphere with a radius of 50mm (Power et al., 2014).

7.4 Results

7.4.1 Recruitment

The present study was a PoC study and thus one main outcome is the feasibility and efficiency of MCA stroke patient recruitment. The recruitment started in December 2016 and was initially planned to be completed by August 2017 to allow completion of the study by September 2017. The first data collection began in late May after the preregistration was completed and uploaded to the Open Science Framework. Due to slow recruitment rates we extended the study end date until February 2018, when the last patient completed the study. During the period December 2016 – February 2018, 44 patients were screened for their eligibility (Figure 7.2). In total, 37 patients (84%) had to be excluded because they did not meet inclusion criteria. Leading reasons were overall poor health conditions (13 patients, 30%), either as a complication of the stroke, or due to pre-existing co-morbidities, as well as cognitive impairment (12 patients, 27%). Another five patients (11%) declined to participate for personal reasons. In total, seven patients (16%) provided consent to be included in the study, of which five patients completed the study (11%). One consented patient withdrew consent later due to personal reasons and one patient deceased before starting the study.

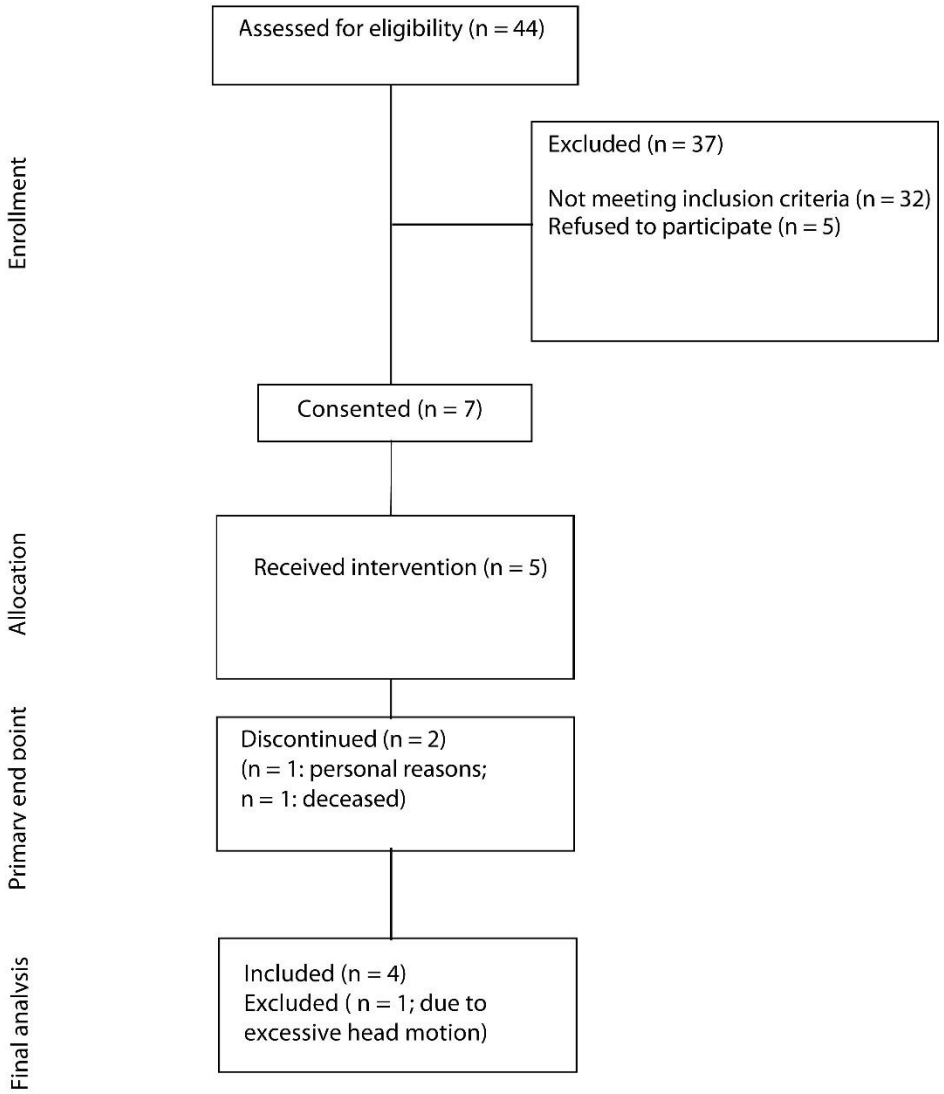

Figure 7.2 CONSORT flow diagram.

7.4.2 Patient demographics

The patients that completed the study mainly suffered from right hemispheric MCA stroke, resulting in left-sided hemiparesis (Table 7.2). Overall, patients were relatively young, with a median age of 49 years (range 38 to 68 years), broadly gender balanced (three female and two male patients) and mostly identified themselves as right-handed (four right handed, one left handed). Patients further mostly suffered from right-hemispheric MCA stroke (four right hemispheric stroke, one left-hemispheric stroke).

Subject	Fugl Meyer (max 66)	Median ± interquartile range/ Median percentage of excluded scan volumes
P1	42 (moderate)	0.26 ± 0.13 mm/ 14.5%
P2	13 (very severe)	1.19 ± 0.03 mm/ 97.8%
P3	43 (moderate)	0.27 ± 0.08 mm/ 6.7%
P4	66 (recovered)	0.32 ± 0.11 mm/ 14.5%
P5	52 (mild)	0.167 ± 0.02/ 1.1%

Table 7.2 Patients' motor impairment and head motion.

7.4.3 Preregistered analyses

FMRI data pre-processing and motion correction were applied as described in 7.2.3.1 and data exclusion criteria based on patient's head motion.

All data acquired from patient 2 (P2), as well as two runs from patient 1 (P1), were excluded because of severe head motion. For P2, general discomfort due to joint pain and obesity prevented closure of the acquisition coil. Hence, head padding was suboptimal, which likely resulted in larger head motion. The remaining data were processed as described in 7.2.3.2 and 7.2.3.3. In total, about 33% of all acquired volumes contained head motion above the set threshold of 0.5mm. The exclusion of entire runs when more than 30% of volumes passed this threshold resulted in discarding 15 of 50 (30%) acquired functional imaging runs. For the remaining PSC data, subject's median PSC values across runs were calculated to conduct further analyses as described in Chapters 7.2.3.2 and 7.2.3.3.

The groups' median localiser PSC value was 1.12 However, in three sessions the lower boundary value of 0.7 was used because the PSC estimation procedure yielded a value that was too low. This indicated that patients overall struggled to activate the SMA during the kinaesthetic motor imagery localiser run.

To test H1_A (net activation of the SMA), data were submitted to a right-tailed one-sample Bayesian t-test. The results indicated only anecdotal evidence for SMA activation during fMRI-NF blocks compared to rest (N(0, 1.12), BF_{+0} = 2.316, median posterior Cohen's d = 0.794, 95% Credibility Interval (95% CrI) [0.074, 1.819], Figure 7.3A). The prior sensitivity analysis that was carried out with JZS priors and a range of scaling factors were overall comparable and suggested anecdotal evidence for SMA activation (Figure 7.3B). This convergence between analyses indicated that the results were relatively invariant to the prior distribution being used (Rouder et al., 2016).

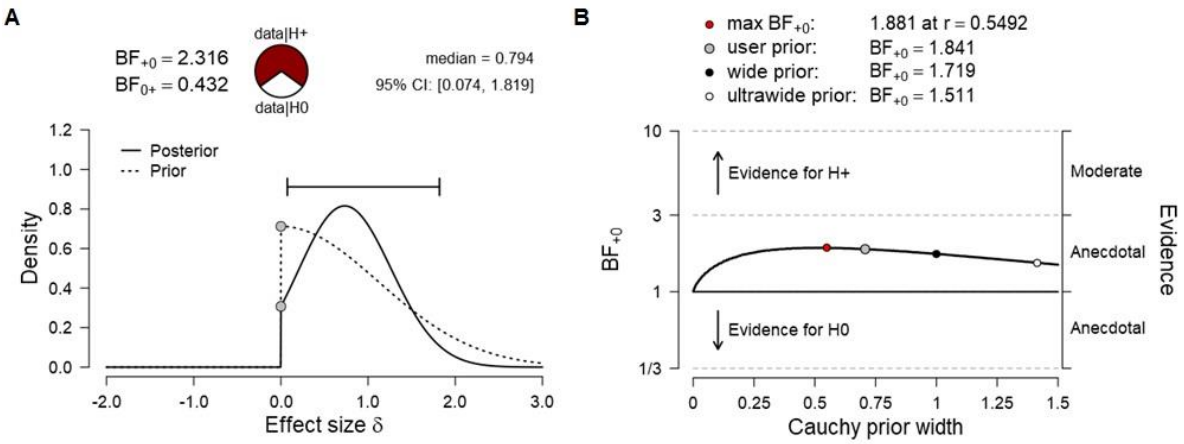

Figure 7.3 A) Bayesian t-test and B) Prior sensitivity analysis for H1_A (net activation of the SMA).

Data was also tested with a right-tailed one-sample frequentist t-test, which remained inconclusive (t_3 = 1.789, p = 0.086, mean difference to zero: 0.142, 95% CrI [-0.045, ∞], Cohen's d = 0.895, 95% CrI [-0.161, ∞]). Given the small sample, assumptions of normality of the data were tested. A Shapiro-Wilk test indicated that these were fulfilled (W = 0.808, p = 0.118).

To test H1_B (more SMA activation for high vs. low target level condition), data were submitted to a right-tailed paired Bayesian t-test. The results indicated anecdotal evidence for the absence of a level effect (N(0, 0.5), BF_{0+} = 1.829, median posterior Cohen's d = -0.198,

95% CrI [-0.009, -0.715], Figure 7.4A). The prior sensitivity analysis suggested moderate evidence for SMA activation for various scaling factors and the evidence for the null increased monotonically with increasing prior width (Figure 7.4B). A frequentist right-tailed t-test remained inconclusive (t_3 = -0.483, p = 0.669, mean difference = -0.014, 95% CI [-0.083, ∞], Cohen's d = -0.242, 95% CI [-1.06, ∞]).

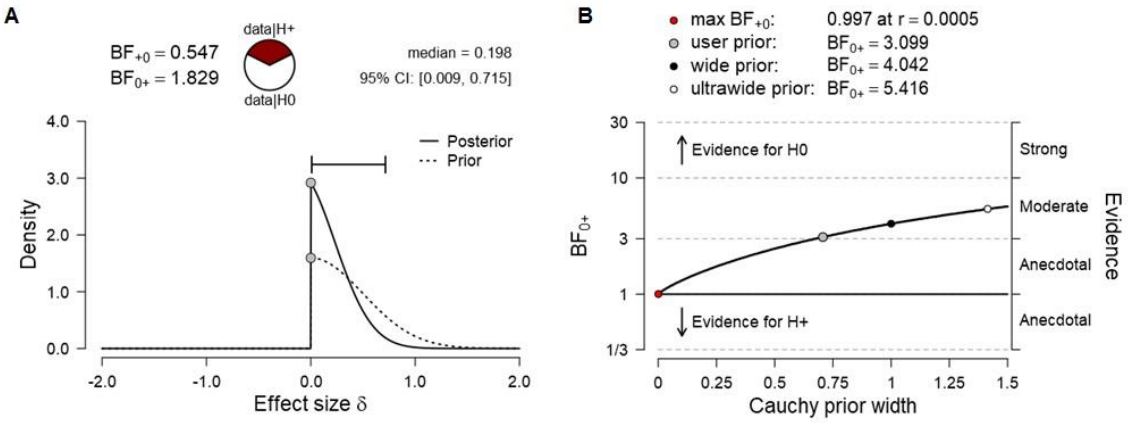

Figure 7.4 A) Bayesian t-test and B) Prior sensitivity analysis for H1_B (more SMA activation for high vs. low target level condition).

Next, H1_A and H1_B were tested on subject level using frequentist t-tests. First, a test for non-normality indicated no deviation for any patient with regards to their average PSC values (collapsed across the low and high level; Table 7.3).

	W	**p**
P1	0.904	0.435
P3	0.917	0.513
P4	0.907	0.452
P5	0.909	0.463

Note. Significant results suggest a deviation from normality.

Table 7.3 Test of Normality (Shapiro-Wilk) for H1_A on subject level.

When split by target level (H1_B), no deviation from normality was detected (Table 7.4).

	W	**p**
P1 high - low	0.973	0.892
P3 high - low	0.886	0.337
P4 high - low	0.883	0.322

	W	p
P5 high - low	0.981	0.940

Note. Significant results suggest a deviation from normality.

Table 7.4 Test of Normality (Shapiro-Wilk) for H1_B on subject level.

To test for net SMA activation, one-sided one-sample t-tests for patients individually. Tests for P1 and P3 remained inconclusive, and for P1 there was a trend opposite to the expected direction (SMA deactivation). Tests for P4 and P5 reached significance when uncorrected, however, when corrected for multiple testing using Bonferroni, only the effect for P5 remained significant (Table 7.5).

					95% CI for Location Parameter				95% CI for Effect Size	
	t	df	p	Location Parameter	Lower	Upper	Effect Size	Lower	Upper	
P1	-0.42	4	0.653		-0.03	-0.191	∞	-0.189	-0.921	∞
P3	1.34	4	0.126		0.04	-0.024	∞	0.599	-0.241	∞
P4	2.598	4	0.030		0.153	0.028	∞	1.162	0.130	∞
P5	5.488	4	0.003	*	0.401	-0.116	∞	2.454	0.827	∞

Table 7.5 Frequentist one sample t-tests (H1_A) on subject level. *) significant at Bonferroni corrected alpha = 0.0125.

Bayesian one-sample t-tests that were informed by patients' individual activation during the motor imagery localiser scan indicated moderate evidence for P4 and P5 (Table 7.6).

	BF_{+0}
P1	0.525
P3	1.759
P4	3.579
P5	6.906

Table 7.6 Bayesian one-sample t-tests (H1_A) on subject level.

Paired t-tests for high vs. low target level remained inconclusive for P1, P3 and P4, with mean differences indicating a trend opposite to the expected direction (i.e. $PSC_{low} > PSC_{high}$). Tests were only significant for P5 before, but not after correction for multiple testing (Table 7.7).

		t	p	Mean Difference	SE Difference	Cohen's d	95% CI for Cohen's d Lower	Upper
P1 high	- low	-0.75	0.753	-0.047	0.062	-0.336	-1.076	∞
P3 high	- low	-0.92	0.794	-0.066	0.072	-0.409	-0.156	∞
P4 high	- low	-2.22	0.949	-0.060	0.029	-0.946	-1.809	∞
P5 high	- low	2.53	0.032	0.057	0.022	1.132	0.122	∞

Table 7.7 Frequentist paired sample t-tests (H1_B) on subject level.

Bayesian t-tests suggested anecdotal to moderate evidence for an absence of a level effect for P1, P3 and P4, and moderate evidence for the presence of a target level effect for P5 (Table 7.8).

	BF_{0+}
P1 high - low	2.201
P3 high - low	2.337
P4 high - low	3.009
P5 high - low	0.288

Table 7.8 Bayesian paired samples t-tests (H1_B) on subject level.

Taken together, data suggested anecdotal or moderate evidence for SMA activation in two patients, respectively. With regards to level effects, one patient showed moderate evidence for an effect, whereas the remaining patients showed anecdotal to moderate evidence for the absence of an effect.

7.4.4 Exploratory analyses

The following analyses were not predeclared in the pre-registration protocol and served to explore hypotheses that were declared post-hoc.

7.4.4.1 Head motion and qualitative observations

Head motion-based data exclusion criteria indicated that about one third (33.02%) of the acquired volumes contained head motion beyond the threshold. This data exclusion compromises significantly the power of the current study, in particular in combination with a second threshold that discards entire runs if more than 30% of volumes are affected.

In comparison, when applying the same threshold to head motion data acquired from a similar paradigm in healthy participants (Chapter 6), only about 3.4% of volumes showed too large motion. Head motion in stroke may result from different factors, including joint pain and general discomfort (reported by P2, who showed consistently large head motion and was excluded from the analysis), compromised ability to focus on a task (observed in P3), and in general a lack of or compromised somatosensory afference. Median head motion ranged from 0.167mm to 1.19mm between patients, with a group median of 0.32mm. With regards to excluded volumes, the variability for the stroke patients in this sample was even more drastic, ranging from 1.1% to 97.8%. In comparison, when the same threshold was applied to data acquired in young healthy participants in a largely similar fMRI-NF paradigm, median head motion ranged from 0.10 mm to 0.25mm between participants, with a group median of 0.15mm, less than 50% found in the patient sample.

7.4.4.2 Evidence for negative over positive level effect

One unexpected outcome of the present study was the trend towards level differences opposite to the expected direction. For the following analysis Bayes factor's transitivity was exploited (Chapter 3.4.2.2) to calculate the evidence for a negative over a positive level effect BF_{-+}. Specifically, a new BF_{-+} by was calculated as the ratio between Bayes factors resulting from two one-sided Bayesian t-tests of opposite direction (i.e. tests for a negative and a positive effect, respectively). A test for a positive level effect (H1_B) yielded $BF_{+0} = 0.547$, a test for a negative level effect yielded $BF_{-0} = 0.726$, their ratio resulted in $BF_{-+} = 1.327$. Hence, data indicated anecdotal evidence for a level effect that was reversed to the expected direction.

7.4.4.3 Relation between lesion site and motor imagery ratings

Patients evaluated their motor imagery capacity through a standardised motor imagery questionnaire, the Kinaesthetic and Visual Imagery Questionnaire (KVIQ). The KVIQ includes test items for visual and kinaesthetic motor imagery for the left and right upper limb, respectively, which are rated on a scale from 1 to 5 (Francine Malouin et al., 2007). The

questionnaire was used to familiarise patients with motor imagery, as well as to evaluate patient's self-ratings. For this study, only items 3 to 5 of the upper limb section were assessed, which included motor imagery tasks concerning the shoulder, elbow and fingers. Scores for each body side and modality were calculated, such that a maximum of 15 points could be scored. To evaluate differences in self-rated motor imagery, scores obtained of the affected side were subtracted from the scores of the non-affected side. Positive scores thus reflect a higher rating for the non-affected bodyside, an outcome that one may expect based on reports of motor imagery in stroke patients (Malouin et al., 2007; McInnes et al., 2016). However, results suggested that patients either mainly rated higher scores for the affected side or rated sides indifferently, as indicated by zero and negative difference scores (Table 7.9). Likewise, no clear relationship with respect to patients' handedness was evident.

Subject	Dominant body side	Affected body side	VI left	VI right	VI Diff	KI left	KI right	KI Diff
P1	L	L	5	4	-1	7	7	0
P2	R	L	10	9	-1	12	9	-3
P3	R	L	12	8	-4	15	15	0
P4	R	L	6	7	1	10	10	0
P5	R	R	9	9	0	13	9	4

Table 7.9 Side of dominance and stroke, alongside self-rated motor imagery sum scores for visual imagery (VI) and kinaesthetic imagery (KI) for the left and right upper limb. Difference scores are calculated with respect to the non-affected side. Positive difference scores reflect larger ratings for the non-affected side, negative difference scores reflect larger ratings for the affected side.

7.5 Discussion

This chapter presented data from a preregistered PoC study of kinaesthetic motor-imagery based real-time fMRI-NF training in first-time MCA stroke patients. On group level, data suggested only anecdotal evidence for the two main preregistered hypotheses. Analyses on subject level indicated that most patients did not activate the SMA when multiple testing correction was applied. Similarly, results suggested that only one patient separated target levels significantly, but for three patients, PSC values were marginally larger during the low compared to the high-level condition. This reversed level effect was supported by Bayesian hypothesis tests that indicated anecdotal to moderate evidence for the absence of the hypothesised positive level effect in three out of four patients.

The marginal evidence for SMA activation during motor imagery may be surprising in light of earlier reports of SMA activation during motor imagery in chronic stroke patients (Confalonieri et al., 2012; Sharma et al., 2009b, 2009a; Sharma and Baron, 2013). Besides larger sample sizes of earlier reports, differences in patient selection procedures likely play a role. First, previous studies mostly included well recovered patients, whose lesion was often limited to small subcortical structures (Confalonieri et al., 2012; Sharma et al., 2009b, 2009a). Further, Sharma and colleagues used a rigorous screening procedure to assess motor imagery abilities at baseline (Sharma et al., 2009b, 2009a). Thereby, these studies ascertained that included patients were skilled in performing motor imagery. However, this form of rigorous screening also excluded 30-50% of patients (Sharma et al., 2009b, 2009a; Simmons et al., 2008). Screening for baseline motor imagery capacity may also be one main reason why samples of previous motor imagery studies featured mainly, or even exclusively left hemispheric stroke patients (Sharma et al., 2009a) Compared to right hemispheric stroke, which most patients had suffered in the present sample, left hemispheric stroke patients show less deficits in motor imagery tasks. For instance Malouin and colleagues found that right-sided parietal stroke was associated with slower motor imagery performance (Malouin et al., 2012). Another study found that right hemispheric stroke patients did not exhibit congruence between executed and imagined hand movements (Kemlin et al., 2016). Damage to the fronto-parietal network has been hypothesised as a potential pathophysiological mechanism underlying these deficits (Buch et al., 2012; Dettmers et al., 2015), For instance, the network is involved in cognitive processes that involve spatial attention and body-environment interactions (Ptak, 2012), functions that are likely involved in different forms of motor imagery (Hanakawa,

2016). Of interest, one electrophysiological study that paired motor imagery with TMS pulses over the motor cortex found motor facilitation effects in stroke patients with lesions in the left, but not in patients with lesions in the right hemisphere (Stinear et al., 2007). While the present study used a motor imagery questionnaire that has been validated in stroke patients to assess self-rated motor imagery capacities (Francine Malouin et al., 2007), it was not used to select patients based on their ratings. Noteworthy, data from questionnaires did not match with the side of the stroke and hence did not reflect the pathophysiology (i.e. the affected side was rated as equally high or even higher compared to the non-affected side). The aim of this PoC study was to test the technique within the constraints of an established health care setting to provide a measure of feasibility with an ecologically meaningful sample. Also given the limited recruitment success, further restriction on patients would have additionally decreased the statistical power and biased the sample towards patients with above average motor imagery capacities. Taken together, the presented imaging findings are limited given the small sample size that has likely compromised the statistical power, but also lesion locations that may have impaired patients' motor imagery abilities.

The choice of the SMA as the target region and the design of motor imagery based graded fMRI-NF were informed by an experiment previously conducted in young healthy participants (Chapter 6). However, a few modifications were introduced to meet anticipated requirements for stroke patients with motor impairment: for instance, a motor imagery localiser was used to target voxels that responded to the task of interest, and to obtain an estimate for baseline PSC voxels within the SMA during the task of interest: motor imagery. In the event that patients did not achieve stable SMA activation during the localiser scan, preregistered lower and upper bounds (Chapter 7.3.2.3) were used similar to an approach used in studies with neurological patients (Subramanian et al., 2016, 2011). The choice of the target region (SMA) as a target region for fMRI-NF training in MCA stroke patients was informed by the outcome of a previous study (Chapter 5), but also based on neurovascular criteria: Cortical MCA stroke affects motor regions including M1 and the dPMC. In contrast, the SMA is supplied by the anterior cerebral artery (ACA) (Brugger et al., 2015). Hence, grey matter around the SMA was expected to be largely preserved, providing a more reliable basis for target voxel selection. Further, it was expected that the neurovascular coupling and perfusion were less affected compared to regions supplied by the MCA. Moreover, the length of training blocks was adapted for stroke patients such that the block length lasted 50-100% longer (30 seconds vs. 15-20 seconds) compared to previous motor imagery based fMRI-NF studies (Berman et al., 2012;

Chiew et al., 2012; Sitaram et al., 2012; Subramanian et al., 2016, 2011) in order to account for potential hemodynamic delays. Global and local hemodynamic BOLD delays, in particular around the lesion territory but also relative to the non-lesioned hemisphere, have been reported for patients suffering from cerebral micro- or macroangiopathy. Such vessel pathologies may manifest in reduced vascular reactivity, which regulates hemodynamic responses to changes in pCO2/pH of brain tissue, as well as reduced neurovascular coupling, which adapts perfusion based on changes in metabolic demands (Wolf, 2015). Hence, future fMRI-NF studies may benefit from estimating hemodynamic delays in individuals to optimise their setup.

In fact, hemodynamic delays in resting-state fMRI data can be used to localise perfusion deficits (Lv et al., 2013) and this information can be either used to inform target voxel selection, or to adapt training block lengths. Furthermore, given that the BOLD contrast depends on coupling relationships between the metabolic rate of oxygen, cerebral blood flow (CBF), and volume (CBV), changes in the metabolic rate of oxygen (e.g. due to inflammation) may attenuate BOLD responses. Besides hemodynamic delays, reductions in BOLD amplitudes have been observed in stroke patients, which may result from reduced neurovascular coupling (Lin et al., 2011; Schroeter et al., 2007), or an inability of the neurovascular system to increase local perfusion (Siegel et al., 2017). Both aspects are of particular relevance for fMRI-NF in stroke patients, because neurovascular coupling likely represents a key physiological mechanism in learning self-regulation of the BOLD signal (Birbaumer et al., 2013). Pathological changes of BOLD dynamics in stroke patients render the interpretation of the present findings, i.e. only marginal evidence found for SMA activation, difficult, and can challenge for successful implementation of fMRI-NF as well as functional near-infrared spectroscopy (fNIRS-NF) training for certain stroke patients. Hence, future attempts of BOLD based neurofeedback training should include vascular measurements such as breath-hold functional scans that allow assessing vascular reactivity in stroke patients (Geranmayeh et al., 2015; Murphy et al., 2011), as well as CBF and CBV weighted data that can help understanding null findings related to BOLD (Blicher et al., 2012). These measures allow assessing intracerebral vascular reactivity and can thereby help to dissociate factors that could contribute to reduced or absent BOLD responses as reported here for the SMA. Moreover, they may allow to identify patients whose brain vessels are still largely intact, and who may show most training success. Lastly, looking at the role of neurovascular coupling though the lenses of rehabilitation, retraining this capacity after cerebral stroke may represent a promising outcome variable in its own right.

Another main challenge related to the imaging is related to patients' head motion. In the present study head motion severely compromised data quality such that MRI co-registration failed at setup, mainly because anatomical image quality was insufficient (for illustration, see Appendix C, Figure C1). Therefore, co-registration during fMRI-NF sessions was not feasible. Besides head motion, other factors such as deformations and large lesions likely also contributed to failed co-registrations. These issues remain a challenge for offline data analysis pipelines (Siegel et al., 2017). As a result, fMRI-NF training was performed in native space and SMA target voxels were selected manually as described in 6.2.8. Head motion also corrupted the quality of functional data. During offline analyses, this was accounted for by quantifying framewise displacement (FD) and excluding entire runs that contained too much head motion as assessed by predefined criteria (>30% of volumes containing >0.5mm head motion). These criteria resulted in the exclusion of an entire patient and two runs of another patient (25% of total data). Whether the applied criteria may have been too conservative remains an open question for stroke research. For comparisons, FD > 0.5mm is considered a lenient threshold in resting-state fMRI for healthy participants (Power et al., 2014, 2012). In the context of stroke, this threshold likely represents a reasonable compromise in the trade-off between sufficient statistical power and rigorous quality control. Another study that applied a threshold of FD >0.5mm to resting-state data acquired in acute stroke patients (in scans of similar run length, and excluding entire runs if >50% of volumes were affected) discarded similar amount of data as the presented study (30% of data exclusion, 14% of patients entirely excluded; Siegel et al., 2017)). Likewise, their reported estimates of head motion (0.4 – 0.5mm) were comparable to the median head motion reported here (0.36mm). Taken together, head motion and required data exclusion severely compromised statistical power in the current sample. Effective thresholds for data exclusion criteria and alternatives remain to be explored.

Factors that may complicate adequate data exclusion criteria include potential interactions between severity of head motion, severity of motor impairment, and patient's cognitive capacities. Although this has not been systematically evaluated yet for stroke patients, this notion is supported by one fMRI study that used an N-back task in patients suffering from multiple sclerosis (Wylie et al., 2014). The authors reported larger head motion in patients compared to healthy controls, and this difference scaled with difficulty of the task, and was inversely related to cognitive capacities. Hence, more challenging conditions resulted in more head motion, which affected data from individuals with lower cognitive capacities in particular

(Wylie et al., 2014). With regards to real-time fMRI studies including fMRI-NF training, such interaction should be considered when selecting patients and formulating data exclusion criteria. Moreover, besides data discarding, alternative techniques such as prospective motion correction may provide more effective solutions to maintain statistical power and prevent sampling biases. With regards to the role of head motion in published motor imagery studies conducted in stroke patients, quality control measures for head motion have not been documented (Liew et al., 2015; Sharma et al., 2009b; Sharma and Baron, 2013). This suggests that besides the reported realignment of acquired volumes, imaging data were not adequately controlled for head motion. Therefore, it is likely that reported clusters may partly have resulted from spurious correlations due to head motion, in particular with regards to reported changes in functional connectivity (Liew et al., 2015; Sharma et al., 2009a). Confounding through head motion may also explain the counterintuitive findings reported by Liew et al, who applied fMRI-NF training based on connectivity estimates between the thalamus and primary motor cortex (measured by Pearson's correlation between times series of these two target regions). The authors reported that connectivity estimates increased over a two-hour scan session mainly for more severely impaired stroke patients, who presumably showed most head motion. Spurious correlations in connectivity estimates due to head motion have been well documented for healthy participants (Power et al., 2014, 2012) and stroke patients (Siegel et al., 2017). Further while data in the present study were controlled for head motion in offline analyses, the real-time fMRI BOLD signal was still subjected to significant head motion. Hence, feedback shown to patients likely captured spurious activity that was not related to the task of interest (Appendix C FigureC2). Taken together, future neurofeedback studies will likely benefit from a systematic evaluation of head motion in stroke, comparing acute, chronic and recovered stroke patients with age-matched healthy control subjects, as well as sufficient online control for head motion, which will be discussed in the general discussion (Chapter 9).

One main challenge that this study encountered was the rather sluggish recruitment of MCA stroke patients that fit inclusion criteria and did not have to be excluded due reasons such as general health problems or cognitive impairment. The present sample of five patients was recruited out of 44 screened patients, the main reason to exclude patients was impaired cognitive capacity and lack of general well-being (Chapter 7.5.1). In consequence, the trial period was extended by 5 months (100% of the originally planned period) to reach the minimum target of 5 patients. Further, due to excessive head motion a large proportion of data (including all data from one patient) was discarded. Thus, the statistical power of the present

study was likely compromised and below the anticipated 30-40% (Table 7.1). However, it should be noted that even though the present sample was relatively small, its size is comparable with previous PoC studies of fMRI-NF in neurological populations (Liew et al., 2015; Sitaram et al., 2012; Subramanian et al., 2011). Notably, for stroke research, these small sample sizes may reflect overall recruitment challenges for a relatively complex intervention in a patient group that suffers from multiple medical and neurological complications following ischemic stroke (Johnston et al., 1998). For comparison, one registered trial at a leading centre of biomedical research (National Institutes of Health) reported a recruitment of seven patients out of an anticipated sample size of up to 50 patients over a four year period and also required extending the trial termination date (https://clinicaltrials.gov/ct2/history/NCT02089776). Besides low statistical power, the recruitment history further suggested that the present sample was biased towards stroke patients that showed above average cognitive capacities and general health. These factors are also reflected in the relatively young median age of 49 years, an age group among which the incidence of stroke only represents only a fraction compared to the incidence found for higher age groups (Feigin et al., 2003). As noted earlier, the sample mainly consisted of patients suffering from right hemispheric stroke, which may be surprising from an epidemiological perspective given that strokes across hemispheres occur at similar rates, if not that left hemispheric stroke is more common (Hedna et al., 2013). Although it remains speculative, one reason for this unfavourable sampling bias may be that left hemispheric stroke has also been implicated with more severe clinical outcome and higher mortality (Hedna et al., 2013). Moreover, left hemispheric stroke is more often associated with forms of aphasia. Hence, left hemispheric stroke patients may have been more likely excluded because they did not meet all inclusion criteria, resulting in a sampling bias of right hemispheric stroke patients.

Besides being the first fully preregistered fMRI-NF study, the presented work also introduced Bayesian Sequential Sampling (BSS) as an innovative sampling method to the field of neurofeedback training. BSS allows to accumulate data until a desired threshold of evidence either for the alternative or the null hypothesis is reached (Schönbrodt et al., 2015). As such, BSS employs a stopping rule that is not affected by type-I and type-II error rates (Chapter 3.4.2.2), although false positive and false negative rates need to be considered for parameter choices to yield optimal designs (Schönbrodt and Wagenmakers, 2016). The preregistration protocol contained all relevant parameters that were relevant for data pre-processing and analysis, including prior distributions that were informed by subject's localiser scan as described in a predefined procedure, as well as criteria for stopping based on a measure of

evidence. This rigorous approach increased the validity of presented results. For instance, the criterion for the alternative hypothesis was set as BF_{10} >10 (indicating strong evidence for a positive effect) and for the null hypothesis it was set as BF_{01} > 10 (indicating strong evidence for a null effect). These criteria were likely too conservative and mainly implemented to assure that the trial did not stop too early. Bayes Factors that are based on prior distributions whose effect sizes for the alternative deviate considerably from the empirical effect size may overestimate the evidence against the alternative, resulting in relatively large Bayes Factors in favour of the null hypothesis. In particular small samples tend to initially overestimate null effects (Schönbrodt et al., 2015). To improve the generalisability of presented results and assess their robustness, sensitivity analyses were included based on a set of default JZS prior distributions with different scaling parameters. Given the slow recruitment rate, the potential of BSS could not be exploited in the current trial. However, the design may serve as a template for future larger (multi-arm) trials. Lastly, with regards to the preregistered secondary frequentist tests that were carried out, it should be noted that the Bonferroni adjustment of the alpha level for ROI analysis findings across participants may have been too conservative. However, uncorrected p-values were reported, allowing to interpret the results without multiple comparison correction.

To summarise, the present study tested the feasibility of motor imagery based graded fMRI-NF neurofeedback training in a small sample of MCA stroke patients. Results suggested only anecdotal evidence for preregistered hypotheses and replication in larger samples is required. Various potential factors were discussed, including the relatively small sample size, a sampling bias for right hemispheric stroke, which is associated with motor imagery impairment, and pathophysiological changes underlying the BOLD signal that may have attenuated responses in the target region. These measures may inform future attempts to stratify and screen patients in a way that can improve patients' learning ability in self-regulating the BOLD signal using motor imagery. The study was the first fully preregistered fMRI-NF study deposited at the Open Science Framework platform that included pre-specified hypotheses and analysis plans. As such, it ensured transparent reporting and documentation of unanticipated challenges and null findings, that may have remained unreported as often observed with traditional closed science practices. In contrast, the reported work can inform future improvements and replication attempts. Further, the study documented details of recruitment, helping to assess the feasibility of fMRI-NF within a conventional stroke health care setting. The importance of control for head motion in neurological populations has been highlighted and trade-offs

between rigorous and more lenient approaches discussed. Altogether, this PoC study thus represents an experiment in an experiment: experience with preregistration for translational clinical studies remains scarce, recommendations for future attempts will form part of the general discussion in Chapter 9.

8 EEG-NF FOR PARKINSON'S DISEASE – A TRANSLATIONAL CHAPTER

Electrophysiological correlates of motor symptoms and treatment response in Parkinson's disease (PD) have been intensively studied over the last three decades. In particular DBS treatment in PD has provided insights into the relationship between neural oscillations of different frequency bands and motor symptoms in patients (Chapter 1.3.3.1). Pharmacological studies and lesion studies in PD animal models have yielded further mechanistic insights, probing the relationship between dopamine depletion and behavioural deficits typically observed in PD patients. Brain circuit-based models help understanding the pathophysiology underlying PD motor symptoms and provide promising targets for non-invasive interventions that aim to stimulate or entrain neural oscillations at particular frequencies (Chapter 1.3.3.2.3). EEG-NF training provides a non-invasive method to entrain neural oscillations at desired frequencies by the means of self-regulation training. This chapter reviews potential markers for EEG-NF training to improve motor symptoms in PD patients.

8.1.1 EEG-NF in PD and potential EEG target markers

This review focuses on three potential markers that have been considered for the design of an EEG-NF study protocol that aims to facilitate initiation of voluntary movement in early-stage PD patients. Previous fMRI-NF conducted in early-stage (HY stage I-II) PD provided feedback from the SMA (Buyukturkoglu et al., 2013; Subramanian et al., 2016, 2011), a premotor region that forms a key structure in the hyperdirect pathway (Nambu et al., 2002) between motor cortex and STN (Chapter 1.2.3.1). This indicates that SMA activity can potentially modulate STN activity, and vice versa. Given their initial promising clinical findings, Subramanian and colleagues hypothesised that activation of the SMA results in activation of the associated network, including the STN, which may have acted as a compensatory mechanism (Subramanian et al., 2011). The authors indeed found bilateral STN activation in their subsequent larger RCT (Subramanian et al., 2016). Although their study did not test for a relationship between changes in the network and the clinical effects reported for the trial, data from another study can complement these findings in form of correlations between improvements in upper limb bradykinesia and rigidity and the tract integrity between the STN and the SMA after STN DBS treatment (Akram et al., 2017). Taken together, these findings indicate that the hyperdirect pathway between the SMA and STN (Chapter 1.3.3.1) is likely involved in motor imagery training and that it may mediate therapeutic effects.

With regards to potential EEG signatures that could be trained to facilitate movement initiation in PD patients, various electrophysiological correlates of motor symptoms and treatment

success have been documented in the literature. However, to date only a small numbers of motor rehabilitation EEG-NF studies have been conducted in PD patients (reviewed in Esmail and Linden, 2014). With regards to activity of the STN during motor imagery, invasive electrophysiological studies have provided additional insights. For instance, it has been shown that motor imagery (but not mere visual mental imagery) could reduce beta oscillations in the STN (Kühn et al., 2006). These early findings were replicated and extended more recently in a paradigm that studied changes of STN activity in response to different levels of imagined movements. Specifically, beta amplitude in the STN could be modulated by participants when they imagined different insensitivity levels of imagined grip force (Fischer et al., 2017). Although the study did not involve neurofeedback, this finding suggests that graded neurofeedback could also be implemented in a parametric suppression of oscillations in the STN within the beta frequency band.

Beta band oscillations (13-30Hz) have received particular interest in PD research and have been linked to symptom severity and treatment effects. In healthy and PD patients, beta oscillation measured over central electrodes (C3/C4) show a desynchronisation during motor execution, which is reflected in a power/amplitude decrease. However, the reduction in beta amplitude is less observed in PD patients compared to healthy individuals, resulting in exaggerated beta synchronisation (Jenkinson and Brown, 2011). Beta synchronisation facilitates stopping behaviour via hyperdirect and indirect pathway (Chambers et al., 2009; Jahfari et al., 2011) and may hence explain PD symptoms of bradykinesia and rigidity (Engel and Fries, 2010; Jenkinson and Brown, 2011). Animal models and electrophysiological recordings in PD patients have linked dopamine depletion to excessive synchronisation in the beta band within the basal ganglia and thalamocortical circuits (Engel and Fries, 2010; Little and Brown, 2014). In nonhuman primates (Wichmann and DeLong, 2003) and mice (Jackson-Lewis and Przedborski, 2007), dopamine depletion and Parkinson like symptoms can be induced via MPTP (1-methyl-4-phenyl-1,2,3,6-tetrahydropyridine) injections. Excessive beta synchrony after MPTP injection has been repeatedly reported for the motor cortex and parts of the basal ganglia including the STN and pallidum (Bergman et al., 1994; Moran et al., 2012). Non-invasive brain stimulation studies in healthy participants have provided further causal evidence for the relationship between increased beta oscillations and slowing of movements (Pogosyan et al. 2009; Joundi et al. 2012; Moisa et al. 2016). Altogether, data from lesion and entrainment studies suggest that dopamine depletion results in increased beta synchronisation, which is associated with slowed movements as one key symptom observed in PD patients.

Moreover, clinical studies have repeatedly reported that PD patients show higher amplitudes of beta frequencies at rest in particular within the STN, compared to healthy controls (Brown et al., 2001; Kühn et al., 2009; Marsden et al., 2001; Pollok et al., 2012). Higher amplitudes in beta frequency, as well as the spatial extent of exaggerated beta synchronisation, have been related to the severity of symptoms, in particular with bradykinesia and rigidity (Gatev et al., 2006; Hammond et al., 2007; Pogosyan et al., 2010; Zaidel et al., 2009). While these studies remain observational and thus largely correlational, intervention studies in PD patients can provide more direct evidence for the involvement of exaggerated beta oscillations in the manifestation of motor symptoms in PD patients. First, recognized treatment options used to manage motor symptoms including STN DBS and dopamine replacement (L-Dopa) therapy result in a suppression of beta activity that is proportional to observed clinical improvements (Little and Brown, 2014). Specifically, motor improvements for rigidity and bradykinesia correlate with observed decreases in beta amplitudes across DBS treatment techniques (Hammond et al., 2007; Kühn et al., 2009; Santaniello et al., 2015; van Wijk, 2017; Wang et al., 2018). Moreover, treatment response to STN DBS has been predicted by the spatial extent of beta oscillatory activity within the STN (Zaidel et al., 2010). Likewise, treatment with L-Dopa reduces beta oscillations (Brown et al., 2001), and the extent of clinical improvements has been found correlated with the change in beta oscillatory activity (Weinberger et al., 2006). Besides changes in amplitude, also correlations between activity of different brain areas within specific frequency bands (i.e. coherence) have been investigated. For instance, two studies reported that levodopa treatment reduced coherence in beta frequencies between the GPi and STN, and between the STN and SMA during rest and during motor tasks, while increasing coherence in gamma frequencies between these structures (Brown et al., 2001; Cassidy et al., 2002). Hence, changes in coherence may provide new markers to understand the disturbed communication between parts of the involved circuitry (Fig 1.3), and how these relate to clinical changes observed after PD treatment. In light of these findings it has been speculated whether beta oscillations maintain a status quo of the motor system by inhibiting movement (Engel et al., 2005; Engel and Fries, 2010). Attenuating the amplitude in beta rhythms may thus represent a potential target for EEG-NF in PD patients. Of interest, one early single patient case study (Thompson and Thompson, 2002) trained a PD patient in increasing lower beta band oscillations (12-15 Hz) while inhibiting higher beta band activity (25-32Hz) over motor cortices. This training was combined with biofeedback of physiological recordings from the respiratory and heart rate. The authors reported that the patient showed subjective improvements in rigidity and quality of life, however, no standardised measures were reported.

Whereas corticostriatal beta oscillations have been implicated in the suppression of movement, gamma oscillations have been linked to promoting movement. For instance, non-invasive brain stimulation studies reported increased movement speeds after the entrainment of motor cortices in the gamma frequency band (60-90Hz) for healthy participants (Joundi et al., 2012; Moisa et al., 2016). In PD patients, dopaminergic treatment has been found to increase the amplitude of gamma oscillations measured over central electrodes proportional to the degree that movements were facilitated (Jenkinson et al., 2013). Other recent work suggests that the interplay between beta and gamma oscillations may represent a key coupling mechanism that sub-serves maintaining motor programmes and mediates treatment effects. For instance, exaggerated coupling between low frequency beta band (13-22 Hz) phase-and-amplitude of high frequency gamma (300 Hz) band amplitude within the STN correlated with bradykinesia and rigidity (van Wijk et al., 2016). Moreover, STN DBS treatment reduces phase-and-amplitude remarkably, and this measure has been shown to correlate with rigidity scores before and after DBS (De Hemptinne et al., 2015). In order to use gamma oscillations as a target signature for EEG-NF training, participants would need to be able to modulate its amplitude (e.g. via motor imagery). Indeed, similar to beta band frequencies, also gamma band frequencies (55-85Hz) can be modulated using motor imagery as revealed by invasive recordings from STN of DBS treated PD patients (Fischer et al., 2017). Specifically, these data showed increased amplitudes (in contrast to decreased amplitude in the beta band) that patients could modulate via motor imagery tasks of different intensity. Overall, the presented literature suggests that gamma rhythms likely promote movement in PD. However, beyond physiological movements, exaggerated gamma oscillations have been linked to pathological movements such as dyskinesia, which represents one common motor complication following long-term treatment with L-Dopa. Although the neural mechanisms of dyskinesia are still poorly understood, gamma oscillations likely contribute (see for review Esmail and Linden 2014). For instance, gamma oscillations [60-90 Hz] in the motor cortex have been linked to dyskinetic symptoms, as reported for the 6-OHDA PD model in rodents (Halje et al., 2012), and more recently also in PD patients (Swann et al., 2016). Therefore, although EEG-NF training of gamma frequencies may promote movement, it may also promote dyskinetic brain states and thus be contraindicated.

Another target for EEG-NF motor rehabilitation in PD may be reducing power in the alpha frequency band (8-12 Hz). Event Related De-synchronisation (ERD) in the alpha bands over

central (C3, C4, Cz, Fz) electrodes precedes voluntary self-paced movements for a short period (~2 seconds) (Pfurtscheller and Aranibar, 1979). The onset of ERD during voluntary movement is delayed in PD patients compared to healthy controls (Defebvre et al., 1994), which may result from a later initiation of motor programming due to sustained movement inhibition in consequence to elevated beta rhythms. L-Dopa replacement therapy reduces delayed ERD in PD patients, as reported for different motor tasks including self-paced button press (Magnani et al., 2002) and wrist flexion movements (Devos et al., 2003). Further, studies have reported correlations between ERD delay and PD severity (Devos et al., 2003), as well as between differences in ERD delays "ON" vs. "OFF" medication and changes in movement speed "ON" vs. "OFF", including self-paced elbow flexions and hand squeezes (Wang et al., 1999). Further evidence for the relationship between successful treatment for PD symptoms and increases in alpha ERD latencies stems from DBS studies. For instance, STN DBS in combination with L-Dopa was associated with an increase in alpha ERD latency during movement preparation over the contralateral central region, as well as an increase in alpha ERD during movement execution over central electrodes (Devos and Defebvre, 2006). The same study also reported that more advanced PD patients showed a larger decrease in alpha ERD latency. Of interest, the same study could not find such relationships for pallidal DBS. This dissociation may indicate that ERD latency modulation during movement preparation (and motor imagery) is mainly influenced by the hyperdirect pathway (Nambu et al., 2002). A more recent STN DBS study extended the focus on lower frequencies, investigating amplitude changes within the alpha band (mu rhythm) and lower beta band frequencies over sensorimotor cortices. The authors reported similar reductions in amplitude for both frequency bands after unilateral STN DBS (Abbasi et al., 2018). Noteworthy, high synchronisations between motor cortices in the lower and higher alpha band have previously been reported for early-stage (mostly HY stage I-II) PD patients compared to controls (Stoffers et al., 2008). Further, decreases in alpha amplitude during STN DBS were correlated with improvements of patient's UPDRS rigidity scores (Airaksinen et al., 2012). Taken together, these observational and interventional studies suggest that the ERD degenerates with disease progression, and it may indicate that PD treatment enhances movement speed by normalising pre-movement alpha ERD. Further, these findings are in line with the notion that the basal ganglia promote alpha ERD by releasing the motor regions from an idling state, thereby permitting effective motor response selection and programming (Brinkman et al., 2016). With regards to motor imagery-based EEG-NF training in PD patients, alpha rhythms may thus provide a promising target to facilitate voluntary movement initiation. Importantly, previous EEG studies suggest alpha rhythms decrease during

motor imagery in healthy participants (Broetz et al., 2010; Llanos et al., 2013). Further, BCI and EEG-NF studies in healthy participants have shown that reduction of premovement alpha desynchronisation (11-13 Hz) is trainable (Mcfarland et al., 2015; Ring et al., 2015). Ring and colleagues demonstrated in a sham (yoked) controlled EEG-NF study that healthy participants could learn to decrease amplitudes in central (Fz) alpha (together with theta amplitude) over three 1-hour long training sessions with recreational golfers (Ring et al., 2015). The study found that participants assigned to the neurofeedback group reduced premovement high-alpha power more compared to the control group that received yoked sham feedback. Pre-post comparisons of behavioural performance (number of golf putts) showed no group difference, however, indicating that although EEG-NF training resulted in a specific modulation of alpha power, it was not associated with behavioural benefits. Yet, it should be noted that the study featured a relatively small sample size (N = 12) and contained only three sessions of EEG-NF training. Moreover, the performance measure focused on kinematic precision. Although this measure represents an important aspect of movement preparation, it may not have been sufficiently sensitive to detect subtle behavioural changes that have resulted from the EEG-NF training. Other transfer tasks that relate to PD symptoms that have been linked to delayed ERD latencies (e.g. slowing of movement) may provide more suitable alternatives. These limitations could be addressed in future trial designs that are modified to test for improvement of motor symptoms and performance in PD patients following alpha ERD EEG-NF training. Of interest, data from a recent preliminary single patient BCI case study reported that a PD patient was able to control their alpha rhythms using simple motor imagery strategies "ON" as well as "OFF" medication (Kasahara et al., 2018). Altogether, besides beta oscillations, alpha ERD in particular remains a promising target for EEG-NF training early-stage PD patients.

8.1.2 Conclusion and trial design

To summarise, electrophysiological correlates reported for PD motor symptoms and treatment effects highlight three EEG markers that may provide suitable targets for EEG-NF training protocol in PD patients: 1) reinforcing beta-band ERD (13-30Hz) to counteract exaggerated beta synchronisation, 2) increasing the amplitude of pro-kinetic gamma-band (30-80Hz) frequencies, and 3) reinforcing pre-movement alpha ERD (8-12Hz) to improve movement voluntary initiation. Of interest, these three frequency bands have also been identified as being most informative with regards to variations in the BOLD signal (Magri et al., 2012). For the case that one of these training frequencies should be successfully piloted in PoC studies, future combined EEG-fMRI studies that aim to develop more effective fMRI-informed EEG markers

(Keynan et al., 2016) of subcortical structures (e.g. the STN) may benefit from this relationship. Entrainment of a frequency band that shares information with the BOLD signal will likely increase the signal to noise ratio within informative frequency bands (i.e. the EEG-NF target) and increase mutual information present in fMRI and EEG time courses. This translational chapter set out to review the evidence for potential EEG markers that may promote movement initiation in medicated PD patients. Based on Ring et al., 2015 and the considerations summarised in this chapter, a PoC study protocol has been developed and registered (https://doi.org/10.1186/ISRCTN16783092). Briefly, the PoC study aims to train 30 people with PD to reduce pre-movement alpha power over three EEG-NF sessions. Patients will not be provided with explicit mental imagery instructions. A within-subjects design with six sessions will be employed (one pre-test "ON" medication at session 1, a pre-test "OFF" medication at session 2, three one-hour EEG-NF sessions during sessions 3 to 5, and one post-test "OFF" medication during session 6). to train reduction of pre-movement EEG alpha power. Such design will allow comparing the effects of PD medication (contrast pre-tests 1 and 2) and the effects of EEG-NF training (contrast pre-test 2 and post-test). The study's primary outcome measures will be motor function (measured with a grip force accuracy task, a reaction time task, and the UDPRS-III motor examination). Changes in cortical activity and self-reported measures of acceptability of the EEG-NF intervention, motor function, and quality of life are used as secondary outcome measures.

9 GENERAL DISCUSSION

The research presented in this thesis has identified and addressed challenges and research gaps in the development of clinical fMRI-NF training. It has demonstrated that fMRI-NF training can yield substantial clinical effects under controlled conditions in a patient population for which standard treatment options did not yield remission. Clinical improvement was related to changes in a validated measure of self-efficacy, which may represent a promising outcome measure for future RCTs. Graded feedback was introduced for motor imagery-based fMRI-NF training, providing a measure to test for self-regulation success. First piloted with healthy participants, self-regulation ability between two motor areas was compared. This protocol was subsequently translated to a preregistered PoC study in stroke survivors. Challenges related to recruitment and quality assurance were encountered, but not anticipated for the preregistration protocol. They thus required post hoc adjustments in the preregistered study design and analysis, highlighting the need for more flexible RR formats that can accommodate the needs of translational research. Besides fMRI-NF, EEG-NF remains a viable alternative, in particular for patients with limited mobility and where fMRI image quality can be compromised due to brain lesions and/or excessive head motion. Moreover, recent PoC studies have explored the potential of electrocorticography based neurofeedback training of beta oscillations in DBS transplanted PD patients, showing general feasibility and simulation voltage while therapeutic efficacy was maintained (Khanna et al., 2017). This work underscores the relevance and broader interested of the research community in identifying suitable electrophysiological targets for neurofeedback interventions in PD. Potential EEG markers for future translation to EEG-NF interventions in Parkinson's disease patients were identified and have informed the design of an ongoing registered trial (https://doi.org/10.1186/ISRCTN16783092).

Treatment resistant depression represents a major challenge for psychiatric care. The development of circuit-based interventions such as neurofeedback training represents a promising avenue to address treatment needs in patients who do not respond to standard care. Based on clinical and neuroimaging data collected in one of the first fMRI-NF RCTs conducted in long-term depressed and medicated patients, this thesis has provided evidence for the clinical efficacy of this technique (Chapter 4). Both groups were matched for clinical factors such as depression severity and medication status. Further, both groups engaged in veridical neurofeedback training in paradigms that have been successfully implemented in earlier studies. Both groups were thus matched for the rewarding experience of successful self-regulation training as a major motivational factor. Specifically, training areas were localised for each patient individually and at each session. This approach allowed to target voxels that

were most responsive within networks of interest (i.e. within the limbic system of the NFE and the scene processing network in the NFS group). Self-regulation based on mental imagery of positive emotions was hypothesised to yield larger clinical improvement, however, no group mean difference was found on the primary outcome measure. In contrast, both groups showed relatively large clinical improvements to a similar extent with over 40% reduction on the HDRS-17 and about 40% remission rate. A detailed survey of reported clinical findings for treatment and placebo groups in pharmacological studies and interventions that involve high-technology revealed that the overall clinical findings were within the range, if not superior to treatment responses that can be expected with accepted augmentation therapies. Moreover, the clinical improvement exceeded even large placebo responses that have been reported in previous literature. Hence, these findings suggest that the experience of successful self-regulation training in combination with effective mental imagery strategies (imagery of positive affect and relaxing scenes, respectively) may engage cognitive processes in depressed individuals that yield therapeutic effects. Furthermore, immediate improvements in depressive symptoms and mood in general were documented with the POMS. Exploratory analyses suggested that POMS changes were partly related the role of feedback provision. Specifically, a relationship between pre and post changes was found for the third training session during which no neurofeedback was provided (i.e. the transfer session). Notably, during this session a relatively large improvement on the POMS $_{\text{Depression}}$ subscale was achieved. These findings are limited because no direct measures of depression severity were measured during the third session to assess clinical effects directly and because these results remain largely descriptive. Yet, they may suggest that intermittent feedback presentation could improve clinical effects. Within-session intermittent feedback has been studied with the aim to increase self-regulation learning rates (Emmert et al., 2017; Hellrung et al., 2018; Johnson et al., 2012) and are a promising feature to improve training designs. Regarding clinical effects, these may occur at a slower cycle and hence may benefit from "between-session" intermittent feedback. This notion remains speculative, however, and requires further testing in larger samples with more repeated measurement that allow studying the time course of clinical effects (Rance et al., 2018). As a measure that can be applied immediately before and after an intervention, POMS data further provided insight into the time course of mood states over sessions. In contrast to clinical scales such as the HDRS-17, which are usually validated for re-assessments after a one or two-week period, implicit mood measures such as the POMS hence allow detecting potential effects at a higher temporal rate and inform the design of future trials with regards to when and how often clinical assessments should be conducted during an intervention.

This thesis has contributed to the exploration of self-efficacy measures and their relationship to clinical outcome and self-regulation success. In particular the reported correlation between residualised changes in depression and changes in self-efficacy represents a promising exploratory finding. Changes in self-efficacy have been recently reported for robot-assisted motor rehabilitation with stroke survivors (Rowe et al., 2017) as well as a physical self-activity programme in individuals with Huntington's disease (HD; Busse et al., 2017). In the context of depression, longitudinal data from a population-based study has shown that self-efficacy mediates the effects of stressful life events and the development of depressive symptoms (Maciejewski et al., 2000). Hence, besides relatively fast occurring clinical effects, entrainment of self-efficacy may yield long-term improvements in patients. Indeed, improvement of depression scores remained, if not slightly continued at follow-up in the presented trial. It is thus possible that successful self-regulation training may yield lasting cognitive changes that promote individuals' self-beliefs. Of interest, a recent review documented the temporal structure of clinical effects reported for two sham-controlled fMRI-NF trials (obsessive compulsive disorder and Tourette Syndrome). The authors reported gradual clinical improvement over time for the treatment, but not the placebo controlled sham-groups for up to 5 months post intervention completion (Rance et al., 2018). This finding seems surprising, given that the included trials trained different target regions in different clinical conditions. Unfortunately, these trials did not report measures of self-efficacy, however, successful self-regulation of target areas represents one main common denominator between their experimental groups as well as both groups of the RCT presented in this thesis. Therefore, it seems advisable for future trials to include longer follow-up periods that allow detecting delayed treatment effects (Rance et al., 2018). Moreover, exploring changes in self-efficacy may yield a better understanding of psychophysiological processes involved in successful rehabilitation. However, neural correlates of changes in self-efficacy remain to be identified.

Besides measures of self-efficacy, personality traits likely mediate treatment responses and constitute predisposing factors to develop depression (Mulder, 2002). For instance, one study investigated treatment effects associated with pharmacotherapy and CBT, and how these may interact with domains of the five-factor model of personality (neuroticism, extraversion, openness, agreeableness and conscientiousness) and their respective facts (e.g. trust as a facet of agreeableness) at baseline for moderately depressed patients (Bagby et al., 2008). The study reported a main effect on both treatment modalities such that the personality trait openness

predicted larger clinical improvements. The authors further identified an interaction between the personality trait of neuroticism and treatment response such that highly neurotic individuals seemed to benefit in particular from psychotherapy compared to pharmacotherapy. In contrast, individuals with lower trust scores seemed to benefit in particular from pharmacotherapy. For fMRI-NF training, it is possible that openness to the treatment facilitates clinical effects. The studies presented in this thesis implemented fMRI-NF in form of supervised mental imagery training and thus represent an interaction of several co-occurring processes including individual mental strategies (thoughts), a biologically informed reinforcement signal and a self-regulation task. As such, other psychometric facets that relate to self-regulation ability will likely influence participant's success and thus their experience during training. Of interest, one recent neurofeedback study reported higher neurofeedback performance for individuals with regular spiritual practice compared to a control group (Kober et al., 2017). Further, high performers showed differences in grey matter of areas that have been implicated to self-regulation training, including the insula and inferior frontal gyrus (Emmert et al., 2016). Another study that investigated whether participants could learn to regulate bi-directional fMRI-NF of the amygdala without any prior instructions reported a negative relationship between participants who showed higher scores for the personality traits agreeableness, and their self-regulation ability in the absence of feedback presentation during a transfer run (Marxen et al., 2016). Taken together, these findings motivate further exploration of the role of personality traits and measures of self-efficacy. Apart from psychiatric conditions, they may also play an important role in motor rehabilitation, e.g. with regards to the use of the affected limb or patients' engagement (Busse et al., 2017; Rowe et al., 2017).

fMRI-NF training represents a new clinical treatment and, besides EEG-NF, the first therapeutic use of neuroimaging data. Assuring sufficient fMRI data quality, which may be compromised by various sources of artefacts and noise, forms a prerequisite to provide valid feedback (Thibault et al., 2017). One main source of confounding lies at the fMRI data level and includes confounding variables such as physiological sources of structured noise related to breathing (and heart rate) that may inflate statistical estimates. According to a recent review that included 99 published fMRI-NF studies, only one third explicitly reported to account for respiration (Heunis et al., 2018; Thibault et al., 2017). The concern of physiological noise was addressed in the analysis of imaging data presented in this thesis (Chapters 5 and 6). However, additional online correction for physiological confounding remains desirable. In particular (long) block designs during which participants engage in challenging tasks may yield to

substantial fluctuations in physiological responses (Brookings et al., 1996; Pattyn et al., 2008) that may confound the BOLD signal used for online feedback presentation. Iterative online correction of physiological fluctuations requires specialised equipment and extensive computing power and has to-date been only realised in one published methodological study (Misaki et al., 2015), but not in fMRI-NF experiments (Thibault et al., 2017).

Another source of confounding is subject head motion. Given the complications encountered in the translational PoC stroke study that resulted from patients' excessive head motion, a detailed discussion about potential solutions seems warranted. All presented online analyses were intra-session aligned (Chapter 3.3.1) to assure that feedback was provided from the same set of voxels. Further, offline analyses included head motion nuisance regressors to correct for fluctuations in the BOLD signal due to head motion. Online feedback in the studies reported in this thesis presented feedback as a ratio (i.e. form of percent signal change) between task and rest periods (Chapter 7.2.2.2). As such, signals were corrected for sporadic head motion that occurred during rest and task period, but not for potential task correlated head motion, which would have required model-based feedback presentation such as an iterative GLM. Similar to offline GLM analyses, online iterative GLM analysis allows including head motion parameters as nuisance regressors, and thereby corrects model estimates that are being used for feedback presentation. An alternative approach to control for global effects on imaging data, including head motion, are reference regions, where their BOLD signal is subtracted from the BOLD signal of the target region, yielding a more robust feedback signal (Sitaram et al., 2012). However, this approach was not considered during the planning phase and hence not preregistered. One question is where reference voxels are best positioned to capture global signal variance, including head motion, but also physiological noise (Murphy et al., 2013). Based on fMRI studies that have characterised head motion and its effects on spurious activity, voxels close to ventricles may be suitable (Power et al., 2014, 2012). A similar approach was used by a previous fMRI-NF study in stroke patients, although measures of effectiveness were not reported (Sitaram et al., 2012). Anecdotal evidence from real-time recordings, however, suggests that voxels adjacent to ventricles may not always sufficiently capture signal fluctuations that result from head motion (Appendix C Figure C2). Future studies may benefit from a systematic evaluation of the effectiveness of this approach on online feedback presentation. Taken together, real-time analysis pipelines can partly correct for head motion either in a model-based approach, or by subtracting global signal variations. Besides approaches that correct for spurious signal activity in the data, another set of techniques allows

correcting for head motion before data is being acquired. Prospective motion correction (PMC) uses optical or mechanical (e.g. bite bar) markers that register head motion. This information is used to adjust data recordings in real-time by updating the position of the acquisition box before the next radiofrequency pulse (RF) is sent (Maclaren et al., 2013). Thereby, PMC can prevent distortions in structural data as well as spurious correlations in functional data. A PMC setup was not available during the data acquisition period. Different implementations of PMC have been tested in the last decade and the trade-off between patient convenience and high measurement precision represents on main challenge. Another real-time technique that can help reducing the effective proportion of motion corrupted volumes is the Framewise Integrated Real-time MRI Monitoring (FIRMM) software (Dosenbach et al., 2017). The software uses real-time motion parameters to estimate at acquisition the expected loss for a given data exclusion criteria. In a second step, scan sequences are prolonged to achieve a desired proportion of data that contains head motion that is still tolerable. Thereby, FIRMM can largely maintain statistical power for a scan session. Another advantage is that run lengths can be flexibly prolonged, for instance by repeated affected blocks and conditions. Data within a run for which head motion does not exceed a threshold can still be used for analysis. This real-time technology was not available for the present study, however, future patient studies that anticipate excessive head motion may benefit from using these approaches. Besides high-technology real-time approaches, one low-technology solution that aims to reduce head motion at the source is surgical tape attached to the forehead and receiver coil. This approach has been explored in previous fMRI work where head motion had posed a particular challenge, for instance in studies of speech production (Gracco et al., 2005; Shuster and Lemieux, 2005). Movement of the head results in friction and resistance at the forehead where the tape is attached, which can serve as a feedback for the participant to reduce head motion. No systematic evaluation has been published to-date to quantify its effectiveness in reducing head motion, however, its effectiveness has been tested in an unpublished study by Krause and colleagues. The authors used a single-blind study design and tested the effects of surgical tape on head motion for passive viewing, motor imagery and speeded responses. Their results found reduced translation and rotation short-term (volume-to-volume) as well as long-term (drift) head motion to a similar extent for all tasks. Further, these effects scaled with severity of head motion when no tape was applied, such that participants with large head motion without the tape showed the largest reduction when the tape was applied (Dr. Florian Krause, Donders Institute for Brain, Cognition and Behaviour, personal communication, 14/04/2018). To summarise, different online and offline methods exist to control for head motion or reduce the

proportion of data that is affected by head motion. Compared to data discarding techniques that were used in Chapter 7, methods that control online for movement can largely preserve statistical power. While none of these techniques will solve the problem entirely, a combination of techniques – for instance of surgical tape, reference voxels, and if available PMC – will likely maximise their positive effect on data quality, feedback presentation and statistical inference. Moreover, online motion correction methods likely prevent a sampling bias compared to data discarding techniques. For instance, patients with more severe motor symptoms tend to show more head motion and if this data is being discarded, the generalisability of results is compromised (Wylie et al., 2014). Furthermore, assuring sufficient control of head motion becomes especially important for advanced neurofeedback metrics that exploit information about neural variability, including DecNef and connectivity based neurofeedback training (Koush et al., 2017; Liew et al., 2015; Watanabe et al., 2017).

One major concern among neurofeedback sceptics refers to the lack of adequate control conditions that allow disentangling unspecific (e.g. placebo) from specific therapeutic effects (Thibault et al., 2017; Thibault and Raz, 2016). To demonstrate specific relationships between self-regulation training and behavioural changes well controlled measures of self-regulation success are desirable. Co-occurring cognitive tasks that are processed by regions that feedback is being provided from are one potential confound, in particular when self-regulation success is inferred from mere target region activation. For instance, previous motor imagery based fMRI-NF studies conducted in individuals suffering from PD or HD have trained participants based on feedback from activity of the SMA (Buyukturkoglu et al., 2013; Papoutsi et al., 2018; Subramanian et al., 2016, 2011). However, besides motor imagery, participant's intention-to-control feedback may already yield activation of the SMA, as well as other potentially relevant training areas including the basal ganglia (Ninaus et al., 2013). Graded fMRI-NF (Krause et al., 2017; Sorger et al., 2016) as employed in Chapter 5 aims to addresses this confound. By calibrating subject specific target levels, it represents a within-subject control that allows testing to which extent the response of a brain area results from true self-regulation, opposed to mere activation due to general co-occurring cognitive processes. A comparison between up-regulation training of M1 and SMA revealed that M1 up-regulation training was associated with M1 deactivation irrespective of the target level, indicating that motor imagery based M1 up-regulation training is likely impractical. Further, SMA activation showed some graded responses, however, this effect was not larger when feedback was provided from the SMA compared to when feedback was provided from M1. Hence, a single session of graded fMRI-

NF training of the SMA may be insufficient to demonstrate true self-regulation effects. Although it remains speculative, this finding may also apply to other areas, for instance the basal ganglia, which are associated with the intention to control or related cognitive aspects co-occurring during fMRI-NF training (Emmert et al., 2016).

Indeed, neurofeedback interventions are complex interventions with many potential confounds, rendering the design of efficient and well controlled RCT challenging (Thibault et al., 2017; Thibault et al., 2016). Moreover, given that active patient engagement (i.e. learning to self-regulate brain activity) forms part of the intervention, the choice of control conditions limits the level of blinding that can be employed (single vs. double-blinded). The RCT presented in this thesis employed an active control group which engaged in veridical neurofeedback training from areas of a distinct brain network and thus did not allow to blind patients to their assigned condition. Specifically, patients received feedback from higher visual areas whilst imaging relaxing scenes. Both groups were thereby matched for a similar experience of successful self-regulation, as well as for a potentially therapeutic form of mental imagery that is also being used in CBT. Such active control condition is relatively conservative compared to other types of control groups that have been suggested for fMRI-NF trials (Thibault et al., 2016). Another type of control aims to match groups for motivation levels and the belief of self-control by providing contingent sham feedback from an area that is not related to pathophysiological model of interest. The trial design by Young et al. 2017 used such an approach in a double-blinded RCT. However, one limitation of contingent sham feedback based on a different brain area is that patients likely don't achieve comparable levels of self-regulation, and groups are thus not matched for the rewarding experience. Indeed, the results reported by Young and colleagues indicated that the experimental group achieved up-regulation of the target region (the left amygdala), while patients in the control group did not attain self-control over their target area (the intraparietal sulcus). Both groups were thus not matched for the rewarding experience associated with successful self-regulation training. This approach may be suboptimal in patients suffering from moderate to severe depression, who will likely show negative cognitive biases (Disner et al., 2011; Geddes et al., 2012). Another approach to control for unspecific clinical effects involves matching the experimental and control group for received feedback (yoked feedback). Specifically, participants in the control condition are provided with non-contingent sham feedback based on recorded brain activity from participants of the experimental group. Yoked feedback conditions hence usually build on blinding of participants and thereby allow double-blind trials. Of interest, this form of control can

theoretically also be conducted in a sham scanner environment that simulates the experience of an MRI machine, but at lower cost. However, yoked feedback conditions bear two risks and thus require careful piloting: 1) patients may notice that no veridical feedback is provided when the presented feedback correlates poorly with their subjective mental imagery experience, in particular when combined with mental imagery tasks where participants already have insight to effective strategies (Marchesotti et al., 2016). Yoked feedback is thus in particular suitable for study designs in which participants are not given a starting strategy (e.g. explicit motor imagery). 2) When yoked feedback closely matches the time course of an area or network of interest, it provides a weaker, albeit potentially still effective reinforcing signal with similar clinical effects. Hence, future trials that employ yoked feedback should report correlations between yoked and veridical feedback time courses as a manipulation check. An alternative way to provide evidence for the specificity of neurofeedback training is inverse sham feedback. In this setup, participants engage in neurofeedback training for which opposite effects are expected, for instance up-regulation and down-regulation motor imagery-based fMRI-NF training of the motor area. This form of bi-directional training can test for specific relationships between self-regulation parameters and behavioural changes. Specifically, the technique builds up on the assumption that directional changes of self-regulation (e.g. up-regulation or down-regulation of a motor area) are associated with behavioural changes in the corresponding direction (e.g. increased and decreased reaction times during a motor transfer task). Successful implementation of bi-directional feedback training ideally controls for a similar success rate between different self-regulation conditions to balance motivation levels and learning success between conditions. Further, bi-directional self-regulation training allows modelling statistical dependencies of brain networks despite the presence of latent confounders (Grosse-Wentrup et al., 2016). It therefore provides an elegant (blinded) design to demonstrate specificity in behavioural changes as well as to identify robust statistical (causal) interactions of involved brain networks. However, given that pathophysiological models usually desire uni-directional changes in behaviour (e.g. an increase in reaction times or a decrease of depression severity), bi-directional fMRI-NF likely does not provide a viable option for most clinical settings. Taken together, bi-directional fMRI-NF training remains a promising paradigm to identify specific neural target signatures in healthy participants. Subsequent translation into a clinical setting, however, likely requires switching to a different control form, which may be informed by neural models generated from bi-directional fMRI-NF training.

While these control conditions include fMRI acquisition or exposure to an fMRI environment other control types exclude this element. For instance, mental imagery training only control groups engage in comparable mental imagery training programmes but do not receive fMRI-NF. However, mental imagery only groups do not control sufficiently for unspecific placebo effects that may result from engaging with an fMRI research setting or fMRI-NF training environment. Placebo effects are subject to participant's beliefs and thus exposure to high-technology interventions can result in relatively large placebo effects, which have been reported for psychiatric (Janssen et al., 2016) and neurological patients (Chou et al., 2015). Another, control condition more commonly used in CBT research involves a waiting list or treatment as usual control (Johnsen and Friborg, 2015; Watts et al., 2015). However, also these controls groups do not capture potential placebo effects and may thus overestimate group differences. Moreover, the lack of neural data in such control groups prohibits testing for neural mechanisms that can be compared between groups. Although controlling for multiple possible confounding factors is desirable, implementing sufficiently powered multi arm RCT will likely have to be balanced against practical constraints such as limited resources and access to patients. In particular early phase fMRI-NF trials with limited sample sizes may thus benefit from focusing on one or two control groups, aiming to provide evidence for clinical effects that are within the range or beyond those reported for established therapies. To yield reliable estimates of long-term remission, relapse rates and hospitalisations, longer follow-up periods are recommended. This information will further allow performing cost-effectiveness calculations and thus additional measures to compare the clinical efficacy of fMRI-NF against other established treatment techniques for specific patient groups, for instance patients who are more likely to relapse, or treatment resistant depression (Berto et al., 2000; Scott et al., 2003).

Another major concern of neurofeedback sceptics is the transferability of learned tasks (Thibault et al., 2017; Thibault et al., 2018b; Thibault and Raz, 2016). A similar critique has been raised with regards to studies that use computer-assisted training of cognitive capacities ("brain training"), and where training effects either did not transfer to other related, non-trained tasks, or where such transfer was not superior to placebo training groups (Kable et al., 2017; van de Ven et al., 2017). Indeed, the goal of neurofeedback interventions should be to support improving ecologically relevant behavioural outcomes that allow patients to cope better with everyday activities and challenges. In the case of psychiatric conditions such as depression, standardised assessments such as the reported HDRS-17 allow evaluating global treatment effects, as well as domain specific effect using item-level based analyses (Chapter 4.3.6.5).

Future designs of neurofeedback trials in depression could additionally include cognitive tasks that allow assessing whether training effects (e.g. reinforced supervised mental imagery training of autobiographic memory) is associated with transferrable capacities (e.g. improved memory performance), and how these relate to clinical changes and measures of self-regulation success (Young et al., 2017). With regards to motor rehabilitation, previous studies included reaction time and balance tasks in trials working with PD patients (Subramanian et al., 2016, 2011), or grip force tasks for studies that aim to reduce motor impairment in stroke survivors (Blefari et al., 2015). Future PoC studies of motor rehabilitation should thus incorporate such tasks in addition to standardised clinical motor assessments (Chapter 8.1.2).

Besides behavioural makers of treatment success, exploring (neuro) physiological markers may be indicative of treatment response, underlying mechanisms, and could thus ultimately help optimising clinical neurofeedback protocols. One impressive example has recently been published in the context of TMS treatment in depressed patients (Drysdale et al., 2016), where the authors used a multivariate classification algorithm to identify neurophysiological subtypes of depression based on resting resting-state fMRI data to predict response rates to TMS. In a next step, such knowledge can be used to stratify patients prospectively based on their neurophysiological subtype to assign them to TMS protocols that seem most suitable. Notably, however, a recent replication attempt in a more heterogenous (but also smaller) sample of depressed patients failed to corroborate the reported biotypes (Dinga et al., 2018). With regards to optimising fMRI-NF training protocols based on resting-state fMRI data, a promising framework has been recently introduced that allows fine-tuning parameters offline (Ramot and Gonzalez-castillo, 2018). Another marker that has received interest to predict treatment responses in depression is brain glucose metabolism, which can be measured with positron emission tomography (PET). PET contrasts in certain brain areas (e.g. the insula) have been found indicative of subsequent treatment response to CBT or pharmacotherapy (McGrath et al., 2013). Notably, the authors reported that in particular response in the right anterior insula could dissociate between responders and non-responders for both treatments. This structure has also been implicated with self-regulation and is a common neural correlate across fMRI-NF paradigms (Emmert et al., 2016). Hence, similar markers may be used to stratify patients or to optimise training paradigm. An exploratory analysis in this thesis did not find a significant relation between treatment response and insula activity, but further exploration is needed. Besides functional markers, structural markers remain promising. Recent DBS treatment protocols for instance integrate information from tractography to identify optimal placement

of stimulation electrodes (Riva-Posse et al., 2017). A similar approach may be promising for motor rehabilitation protocols in stroke patients, given the relevance of the CST in predicting motor impairment (Rondina et al., 2017). Indeed, algorithms for TMS stimulation in stroke survivors have been developed that are informed by the remaining integrity of the CST (Stinear, 2010). For motor imagery-based training, such information may be of particular value because the ability to mentally simulate movements is likely modulated by afferent signals of limb positions. Indeed, performance on a motor imagery based BCI has been related to measures of structural integrity and myelination of structures such as the corpus callosum and cingulum (Halder et al., 2013). Neural markers also provide valuable markers to detect neuroplastic changes during skill learning (Dayan and Cohen, 2011; Zatorre et al., 2012) such as neurofeedback training. TMS measurements for instance could detect changes in intracortical inhibition and excitability in motor imagery fMRI-NF paradigm (Sitaram et al., 2012). Besides electrophysiological markers of neuroplasticity, changes in white and grey matter may result from neurofeedback training, especially in the context of motor rehabilitation (Sampaio-baptista et al., 2018; Sampaio-Baptista et al., 2014). Structural plasticity is maintained to a substantial degree in adulthood and may serve to stratify patients and assess rehabilitation outcome in neurological and psychiatric rehabilitation. For motor imagery-based paradigms, preliminary evidence in neurological patients suggests that fMRI-NF training can result in structural changes. One PoC study conducted in individuals suffering from HD suggested that clinical improvement after several sessions of fMRI-NF training of the SMA was related to increased grey matter volume of the pre-SMA (Papoutsi et al., 2018). Hence, plastic changes may occur in connected areas of an extended network that is being trained during a specific task (e.g. motor imagery). Further, a case study with three patients that completed home-based motor imagery guided EEG-NF training reported white matter changes in the CST for the patient that showed the largest training effect (Zich et al., 2017). Although no reports have been documented yet, it is possible that similar structural plasticity can also be detected for neurofeedback interventions conducted in depression. For instance, reduced myelination on whole brain level has been reported for depressed patients compared to (age, gender and handedness matched) healthy control subjects (Sacchet and Gotlib, 2017). Moreover, the same study found that decreased myelination was related to the number of depressive episodes patients had experienced. Further, reduced myelination was reported for patients suffering from treatment resistant depression relative to patients without treatment resistant depression (Jia et al., 2017). Integrating structural imaging in long-term follow-up assessments of fMRI-NF protocols may allow to detect remyelination after remission. Apart from markers of brain

plasticity, measures of neurovascular reactivity may provide additional insight into mechanisms of action and recovery potential of fMRI-NF training. Such measures may of be of particular interest for applications of BOLD based neurofeedback training in stroke, where the health of brain vessels is compromised. For instance, breath-hold BOLD imaging can assess vascular reactivity in stroke patients (Geranmayeh et al., 2015; Murphy et al., 2011) and would allow testing for neurovascular coupling entrainment. Furthermore, for patients where no or very attenuated BOLD responses are observed, alternative fMRI measures such as cerebral blood flow weighted (CBFw) and cerebral blood volume weighted (CBVw) provide additional measures that can help understanding null findings related to BOLD (Blicher et al., 2012). Taken together, (neuro) physiological makers may be used to predict treatment response, helping to stratify patients and optimise treatment protocols. Moreover, they may support behavioural and clinical effects, thereby providing additional independent measures of evidence to quantify therapeutic effects. Some markers, e.g. electrophysiological signatures, may further allow transferring from relatively resource intensive fMRI-NF to more affordable and portable modalities such as EEG-NF.

Keynan and colleagues achieved the first successful translation from fMRI-NF to EEG-NF for relaxation based neurofeedback training for post-traumatic stress disorder (Keynan et al., 2016). In an initial study, the authors recorded simultaneous EEG-fMRI while participants engaged in relaxation biofeedback training (alpha/theta EEG-NF) based on which an EEG regression model was created that reliably predicted the time course amygdala (Meir-Hasson et al., 2014). In subsequent transfer experiments, this fMRI-informed EEG fingerprint marker (EFP) was successfully used to provide feedback to patients. Comparing EFG-NF vs. sham EFP-NF, the authors demonstrated that EFP-NF allowed participants to down-regulate the amygdala. Further, this group showed a significant decrease in reaction times in an emotion adaptation test, indicating improved implicit emotion regulation (Keynan et al., 2016). A similar approach could be applied to depression, where EEG-NF markers are currently limited mainly to training based on frontal lateralization differences in the alpha frequency domain. Currently available evidence for therapeutic effects of EEG-NF training in depression is limited to a small number of studies (Linden, 2014). For instance, one randomized single-blind EEG-NF observed improvements similar to those as reported here (7.25 points improvement on the HDRS) after five weeks EEG-NF training (Choi et al., 2010), although it should be noted that patients in this trial were less severely depressed and partly already remitted at baseline. An even larger improvement (from 21.38 to 6.23 on the HDRS-17) was reported in a more recent

8-week open-label (i.e. unblinded) study of EEG-NF (Cheon et al., 2016). However, both studies lacked active self-regulation control groups and thus do not allow estimating the expected placebo effect of EEG-NF in depression. By combining EEG and fMRI, new EEG-NF markers for brain regions of the limbic system (e.g. the amygdala) that are currently only accessible with fMRI may be developed. Also, with regards to motor rehabilitation, EEG-based training remains a viable option. For motor rehabilitation in stroke, a recent literature review of EEG-based BCI training suggested that the technique is feasible and shows moderate clinical effects compared to control conditions (Cervera et al., 2017). In particular stroke patients with very limited mobility and potentially excessive head motion, EEG-NF setups that address patient comfort are worth exploring. Alternatively, fNIRS-NF may represent a promising option, for which proof-of-concept has been shown with regards to motor rehabilitation in stroke patients (Mihara et al., 2013). One main advantage of neurofeedback setups that are relatively portable (and more affordable) is that they allow earlier training, potentially already on specialised stroke wards, as well as more training sessions. The factors time after stroke and training insensitivity are both crucial for the restoration of impaired motor function after stroke and should thus be considered for future developments. With regards to motor rehabilitation in PD, this thesis has provided a detailed review of potential EEG-NF target signatures that may reduce motor symptoms in patients (Chapter 8). This work has contributed to developing a PoC EEG-NF training study that is currently being conducted in PD patients. Taken together, transfer technologies address aspects of treatment cost, as well as scalability of an intervention. They hence increase the feasibility of larger clinical trials.

Patient recruitment remains one main challenge in the development of large clinical trials. The recruitment data presented in this thesis suggested differences between conditions with regards to eligibility screening. Whereas nearly 50% of screened depression patients were in principle eligible to enrol in the study, about 70% of screened stroke patients did not meet all inclusion criteria. Among patients that were eligible, nearly 50% in the depression trial were randomised and 5/12 stroke survivors (42%) completed neurofeedback training. Among depressed patients, the attrition rate was about 25% among randomised patients, which is largely comparable to attrition rates for other interventions for depression that involved high-technology (Janssen et al., 2016). None of the stroke patients interrupted the intervention, however, a substantial among of data (30%) had to be discarded due to head motion. Comparisons of recruitment flows between the stroke and depression trial, however, are limited given the differences in sample sizes (e.g. about four times as many patients were screened for the depression compared

to the stroke study) and number of sessions patients were required to attend (five vs. two sessions). Recruitment data from trials of established therapies for stroke motor rehabilitation indicate, however, that substantial screening is required to reach acceptable sample sizes. For instance, one study that compared the effectiveness of two different forms of physiotherapy randomised 288 out of more than 5,000 individuals (about 5%) and reported 15% attrition (Hunter et al., 2018). Taken together, recruitment data provides an indication for the proportion of patients that may be eligible for fMRI-NF training. Future trials are advised to comply with CONSORT reporting standards and document their recruitment in all detail to allow evaluating the scalability of neurofeedback training (Chapter 3.4.4).

Besides scalability and compliance rate of an intervention, recruitment rates also inform adequate power calculations. Early phase clinical trials (Chapter 3.4.3) often face a statistical power dilemma such that they are often limited to a relatively small sample, resulting in relatively low precision estimates (Fig 1.1). This circumstance tends to be overlooked when subsequent larger trials are powered based on estimates from small pilot studies. For depression research, meta-analyses have revealed that superiority of medication increases with depression severity at baseline (Fournier et al., 2010; Kirsch et al., 2008; Sugarman et al., 2014). However, depression severity may be inversely related to compliance. Hence, although the expected margin of therapeutic effects compared to placebo can be increased by including more severely depressed individuals, attrition likely increases, too (Delgadillo et al., 2014; Issakidis and Andrews, 2004). Interactions between baseline severity and attrition may cause sampling biases and should thus be invested using intention-to-treat analyses (Chapters 3.4.6 and 4.3.3). Future trials will likely benefit from more frequent clinical assessments, which does not only allow detecting clinical changes with a higher temporal precision, but also providing more informative intention-to-treat analyses.

Multi-centre clinical trials can accelerate recruitment and may provide sufficient statistical power to study multi-arm designs that pivot various control conditions against each other. Economical sampling plans such as cross-over designs can increase the statistical power of clinical trials to detect smaller effects. However, cross-over designs for fMRI-NF RCTs may be confounded by delayed treatment effects that have been documented in this thesis (Chapter 4.3.4) as well as for fMRI-NF RCTs in other conditions (Rance et al., 2018). Research presented in this thesis introduced an alternative sampling plan: Bayesian Sequential Sampling (BSS). BSS allows stopping to allocate patients to a particular treatment arms once a

prespecified level of evidence for or against an effect has been reached, rendering BSS particularly attractive for multi-arm trial designs. Moreover, if power calculations are based on SESOIs, the absence of an effect can be more likely demonstrated using frequentist equivalence tests (Lakens et al., 2018). SESOIs for clinical trials are ideally informed by minimal clinically important differences (MCIDs) that allow assessing whether a detected change is likely of clinical relevance (Algermissen and Mehler, 2018). However, given the complex nature of fMRI-NF experiments, standardisation of intervention protocols between participating centres will be required. This process involves standardising neuroimaging protocols to ensure similar levels of signal to noise and thus comparable self-regulation conditions (Stoeckel et al., 2014). It will also require standardising decisions related to the online analysis, including the localisation procedure used to target voxels and areas. The TIDierR (Template for Intervention and Replication) checklist may provide a suitable frame work to guide this process. A first modified version of the TIDierR checklist has been recently published and can be used as a tool to assess and promote convergence of research practices among independent groups to increase replicability of clinical fMRI-NF protocols (Randell et al., 2018). As a next step, the translation of neurofeedback training from an idea to clinical practice will likely benefit from the introduction of a framework that guides the stages of innovation, development, exploration, assessment and long-term studies (IDEAL). Originally introduced for complex surgical interventions (McCulloch et al., 2009) and recently adapted for the purpose of physiotherapy (Beard et al., 2017), the development of an IDEAL for neurofeedback training (and brain stimulation more generally) in combination with a public repository where protocols are stored in a standardised way will likely help establishing quality standards for different phases in the translation and increase comparability between approaches. Moreover, it will allow researchers to plan new studies in a way to confirms with ongoing adaptations that may not be evident yet in the published literature. These factors seem crucial in the successful translation of neurofeedback as a complex non-invasive intervention into clinical practice in specialised health care settings.

The research presented in this thesis has been conducted in line with recent recommendations for replicable and reproducible research (Munafò et al., 2017). For instance, the RCT presented in Chapters 4 and 5 was registered and exploratory analyses were explicitly declared as such throughout the thesis. Further, the lack of a group effect was followed up by statistical methods that allow testing for the absence of an effect (Dienes, 2014; Lakens et al., 2018). Chapter 6 is based on a preregistration protocol, which has been logged as the first preregistered fMRI-NF

study on the Open Science Framework (Foster and Deardorff, 2017). As such, the preregistration of a translational neuroimaging technique and its associated challenges and unknowns constitutes an experiment in its own right, documenting unanticipated challenges of a translational PoC study. As documented in detail in Chapter 7, the translation to stroke patients required deviations from the preregistered analysis plans. Importantly, however, these aspects remain documented and publicly available, following open science principles. In contrast to non-preregistered (i.e. closed) science, where this information may have remained unreported, this documentation can thus inform future attempts of fMRI-NF in stroke survivors. Furthermore, the trial protocol introduced a BSS plan that was adapted to neurofeedback training, employing prior distributions that were informed by data of participant's localiser scans. Thereby, feedback presentation, offline analysis and the sampling plan were calibrated based on individual brain responses. This approach is a novel use and integration of patient specific information obtained during the localiser, which may allow optimising sampling plans, potentially yielding more economic trial designs. Future neurofeedback trials may benefit from such approach and modify the preregistered BSS to their needs.

With regards to challenges that were not sufficiently addressed in the preregistration protocol, in particular subject head motion affected data quality. It may serve as an example for the advantage of RRs over preregistration. While this aspect was addressed during internal peer review by formulating data exclusion criteria, more effective strategies may have resulted from external stage 1 peer review (Figure 3.6). Hence, if the project time plan allows, RRs are likely favourable. Although stage 1 peer review may not result in an acceptance in principle, the external peer review may still provide valuable feedback that can be incorporated in a preregistration protocol, improving the quality of the study's design. However, in particular for translation work but also other complex experiments, it is possible that challenges remain unknown to researchers and reviewers during stage 1. Hence, to make RRs more accessible to potentially risky translational work, some modifications of the publishing format may be desirable. For instance, RRs for translational science may offer to provide additional flexibility for researchers in the form of fast ad-hoc review. This additional review would allow incorporating changes that are necessitated by the clinical setting. To promote preregistration and RRs among researchers, current incentive structures need to reflect the increased costs and efforts to implement open science practices. The current research system still largely rewards positive results, biasing the literature (Munafò et al., 2017; Ramsey and Scoggins, 2008). To

meet conservative statistical power requirements, pooling resources and expertise in form of large-scale collaborative science becomes increasingly important. In imaging genetics, for instance, large consortia have formed to achieve this goal (Guglielmi, 2018). However, with ever larger groups of researchers contributing to projects, the dilution of authorship may be a concern, in particular for early career researchers (ECRs). Hence, the successful promotion of more robust translational work will likely require new authorship formats to acknowledge research that is increasingly carried out by large teams, rather than individuals (Fontanarosa et al., 2017). Taken together, preregistration and RRs allow documenting publicly the development of translational research projects at a level of details that exceeds conventional research practices. Besides insights to graded fMRI-NF in MCA stroke survivors, the work documented in Chapter 7 has informed suggestions for future adaptions of the format.

As outlined in the introduction (Chapter 1.2.2), transparent reporting practices are a key element of robust science and it seems that the fMRI-NF field can learn from mistakes of earlier translational attempts in other neuroscience areas. For instance, the credibility of published literature of EEG-NF training in ADHD has been questioned due to small sample sizes of trials, insufficient trial registration, opaque reporting practices and poor trial designs that largely lack adequate blinding and control conditions (Cortese et al., 2016; Sonuga-Barke et al., 2013; Thibault et al., 2018a)Moreover, financial interests may partly conflict with trial reporting practices of EEG-NF interventions (Thibault et al., 2018a). Although these issues are certainly not representative for the whole field of EEG-NF (Arns et al., 2014; Arns and Strehl, 2013), they indicate the importance of introducing counter measures early in the development. Real-time imaging pre-specifies by default substantial aspects of data pre-processing and analysis as well as main hypotheses (Chapter 1.2.2). Hence, amending these protocols to comprehensive analysis plans for submission as preregistration protocols, RRs, or trial protocol pre-publication (Cox et al., 2016) likely require less additional effort compared to other complex experiments (Chapter 1.2.1). Further, although analysis plans are not required yet for trial registrations, current open science developments in clinical medicine indicate that they become required in the near future (Gamble et al., 2017; Kiley et al., 2017). Taken together, real-time analysis neuroimaging trials seem particularly well suited to lead translational research into this new era of science, aiming to increase valid inferences and maximise chances for successful translation (Lorenz et al., 2017; Poldrack et al., 2017; Steward and Balice-Gordon, 2014).

In conclusion, neurofeedback is one of only a few clinical neuroimaging applications that are being developed into protocols for RCTs. The presented research suggests that clinical fMRI-NF training shows promise as an augmentation therapy in depression. Moreover, this thesis has incorporated recent meta-research developments and recommendations (Bosch, 2018; Nosek et al., 2018; Picciotto, 2018) of good scientific practice and thus contributed to their translation in fMRI-NF. To establish fMRI-NF as an augmentation treatment, replications in larger RCTs with additional control conditions are needed. Multi-arm RCTs may benefit from innovative sampling techniques such as BSS that are informed by MCIDs, as well as from longer follow-up periods to assess delayed treatment responses and potential relapses. A detailed review of different types of fMRI-NF control conditions provided an overview for future trial designs. The rationale behind individual feedback training is largely in line with the goals of precision medicine which aims to exploit biological markers to tailor treatments towards the individual (Insel and Cuthbert, 2015). Multimodal profiling of patients including psychometric measures and the discovery of (neuro-) physiological markers will help to evaluate training success, optimise paradigms and unfold further potential of clinical neurofeedback. However, real-time imaging data control remains a challenge, in particular for more advanced metrics that exploit signal variability for feedback presentation and are thus in particular susceptible to artefacts induced by physiological noise and head motion. From a public health care perspective, the cost-effectiveness of fMRI-NF training for individual conditions and patient groups remains to be assessed to allow evaluating its scalability. For certain conditions such as stroke, where many training sessions are required, and patients' mobility is often impaired, more affordable and mobile alternatives such EEG-NF may provide alternatives. Adopting transparent reporting practices early will increase the robustness of reported findings and the credibility of the field. Sitting at the forefront of scientific progress, the neurofeedback field draws upon principles from cognitive neuroscience, bioengineering and translational medicine to optimise their designs. Neurofeedback research thereby also drives methodological innovation that will contribute translating brain imaging from the "bench to the bedside".

10 APPENDIX A-C

10.1 Appendix A

CONSORT 2010 checklist of information to include when reporting a randomised trial*

Section/Topic	Item No	Checklist item	Reported on page No
Title and abstract			
	1a	Identification as a randomised trial in the title	1
	1b	Structured summary of trial design, methods, results, and conclusions (for specific guidance see CONSORT for abstracts)	3
Introduction			
Background and objectives	2a	Scientific background and explanation of rationale	4-5
	2b	Specific objectives or hypotheses	5
Methods			
Trial design	3a	Description of trial design (such as parallel, factorial) including allocation ratio	6-7
	3b	Important changes to methods after trial commencement (such as eligibility criteria), with reasons	N/A
Participants	4a	Eligibility criteria for participants	5
	4b	Settings and locations where the data were collected	5, 7
Interventions	5	The interventions for each group with sufficient details to allow replication, including how and when they were actually administered	6-10; in SM: 1-3
Outcomes	6a	Completely defined pre-specified primary and secondary outcome measures, including how and when they were assessed	7-8
	6b	Any changes to trial outcomes after the trial commenced, with reasons	N/A
Sample size	7a	How sample size was determined	8
	7b	When applicable, explanation of any interim analyses and stopping guidelines	N/A
Randomisation:			
Sequence generation	8a	Method used to generate the random allocation sequence	6
	8b	Type of randomisation; details of any restriction (such as blocking and block size)	6
Allocation concealment mechanism	9	Mechanism used to implement the random allocation sequence (such as sequentially numbered containers), describing any steps taken to conceal the sequence until interventions were assigned	6
Implementation	10	Who generated the random allocation sequence, who enrolled participants, and who assigned participants to interventions	6
Blinding	11a	If done, who was blinded after assignment to interventions (for example, participants, care providers, those assessing outcomes) and how	7
	11b	If relevant, description of the similarity of interventions	7; 8-10; in SM 2-3
Statistical methods	12a	Statistical methods used to compare groups for primary and secondary outcomes	10-11
	12b	Methods for additional analyses, such as subgroup analyses and adjusted analyses	10-11 in SM 3-4
Results			
Participant flow (a diagram is strongly recommended)	13a	For each group, the numbers of participants who were randomly assigned, received intended treatment, and were analysed for the primary outcome	12
	13b	For each group, losses and exclusions after randomisation, together with reasons	12; in SM: 5
Recruitment	14a	Dates defining the periods of recruitment and follow-up	7
	14b	Why the trial ended or was stopped	12
Baseline data	15	A table showing baseline demographic and clinical characteristics for each group	25
Numbers analysed	16	For each group, number of participants (denominator) included in each analysis and whether the analysis was by original assigned groups	12
Outcomes and estimation	17a	For each primary and secondary outcome, results for each group, and the estimated effect size and its precision (such as 95% confidence interval)	12-13; SM 7-10
	17b	For binary outcomes, presentation of both absolute and relative effect sizes is recommended	12-13
Ancillary analyses	18	Results of any other analyses performed, including subgroup analyses and adjusted analyses, distinguishing pre-specified from exploratory	13 and 14, in SM: 6 and 11-14
Harms	19	All important harms or unintended effects in each group (for specific guidance see CONSORT for harms)	In SM: 5
Discussion			
Limitations	20	Trial limitations, addressing sources of potential bias, imprecision, and, if relevant, multiplicity of analyses	15-17
Generalisability	21	Generalisability (external validity, applicability) of the trial findings	17-18
Interpretation	22	Interpretation consistent with results, balancing benefits and harms, and considering other relevant evidence	15-18
Other information			
Registration	23	Registration number and name of trial registry	3
Protocol	24	Where the full trial protocol can be accessed, if available	3
Funding	25	Sources of funding and other support (such as supply of drugs), role of funders	19

Figure A1 CONSORT checklist of publication submission for (Mehler et al., 2018), which was based on excerpts of Chapters 4 and 5.

Outcome Measures	**Change from baseline to follow-up (18 weeks) for NFE / NFS**				
	M	95% CI	% -age	Hedge's g	95% CI
HDRS-17	-9.65/ -11.53	-6.25 to -13.06/ -8.61 to -14.45	-48/-59	-1.57 / -2.06	-2.14 to -1.04/ -3.72 to -0.96
HADS (A)	-3.54/ -5.07	-0.82 to -6.25/ -3.19 to -6.94	-30/ -39	-0.72 / -1.40	-1.19 to -0.03/ -1.84 to -1.03
HADS (D)	-4.38/ -4.33	-1.41 to -7.36/ -2.21 to -6.46	-35/-34	-0.82 / -1.06	-1.35 to -0.37/ -1.43 to -0.71
BIS	1.62/ 1.33	-0.82 to 4.05/ -0.43 to 3.1	12/9	0.37 / 0.38	-0.17 to 0.82/ -0.11 to 0.88
BAS	-2.00/ 2.67	-5.65 to 1.65/ -2.27 to 7.61	-6/7	-0.30 / 0.27	-1.19 to 0.33/ -0.25 to 0.67
QoL	13.31/ 17.47	7.47 to 19.15/ 10.48 to 24.45	24/32	1.26 / 1.30	0.68 to 2.00/ 0.68 to 2.00
TCQ	2.77/ -0.20	-3.70 to 9.24/ -5.38 to 4.98	5/0	0.24 / -0.02	-0.39 to 0.70/ -0.56 to 0.50
TCAQ	9.15/ 10.73	1.52 to 16.79/ 4.1 to 17.37	16/20	0.66 / 0.84	0.11 to 1.15/ 0.34 to 1.35
SES (G)	4.15/ 8.93	-1 to 9.31/ 4.27 to 13.60	9/18	0.45 / 1.00	-0.12 to 1.06/ 0.42 to 1.46
SES (S)	0.69/ 4.93	-1.74 to 3.12/ 2.39 to 7.48	4/36	0.16 / 1.01	-0.46 to 0.63/ 0.55 to 1.56
EQ-5D-5L	-1.73/ -1.85	-0.57 to -2.88/ -0.89 to -2.81	-15/-18	-0.88 / -1.06	-1.44 to -0.50/ -1.83 to -0.49

Table A1 Changes over time in raw mean scores and standardised effect sizes both including 95% Confidence interval estimates at FU assessment. Change in percentage (%-age). Adapted from (Mehler et al., 2018).

10.2 Appendix B

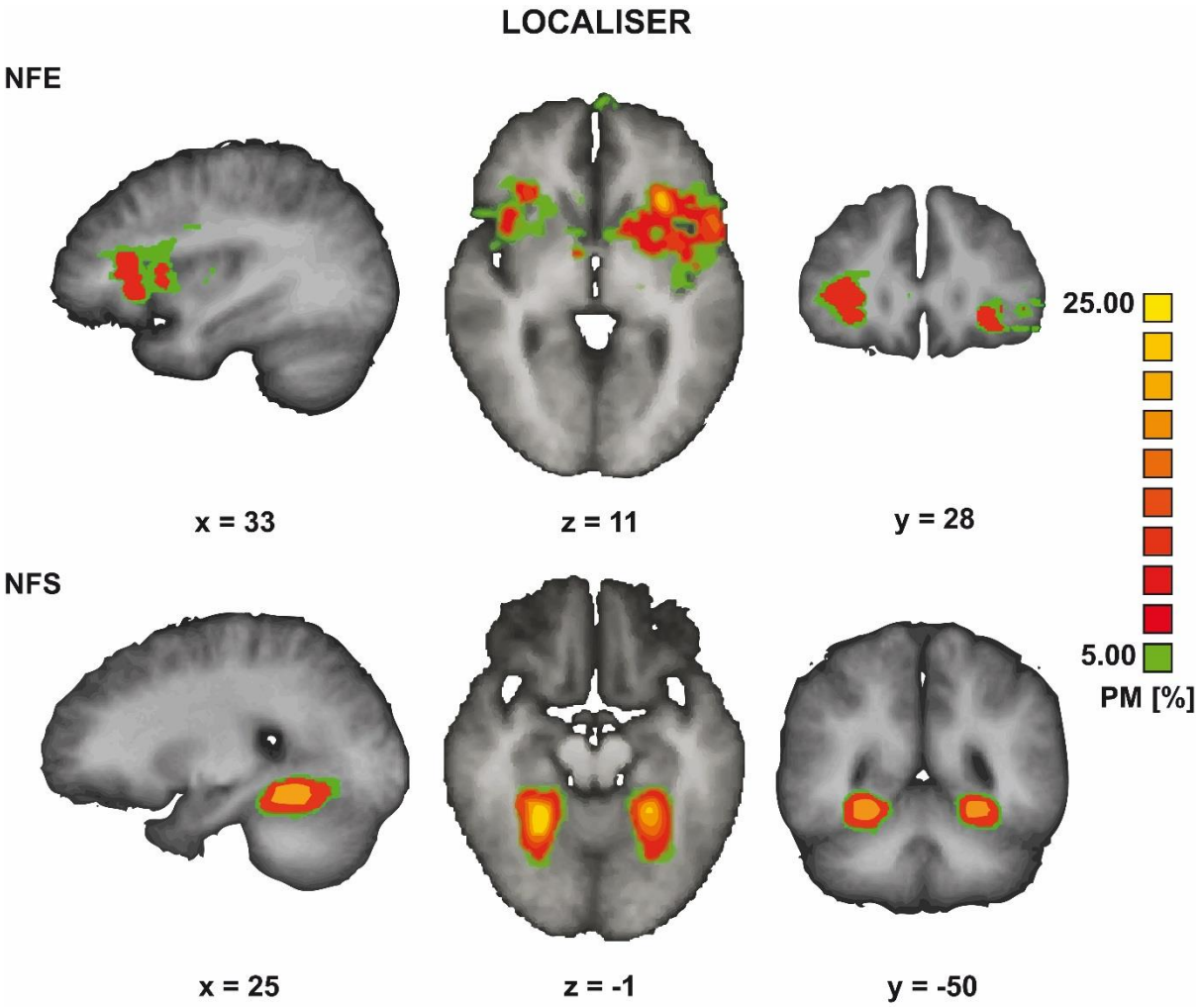

Figure B1 Whole brain probability map of target region selection for both treatment groups. Adapted from (Mehler et al., 2018).

Anatomical label	Mean ± SD coordinates [X/Y/Z]	Mean t-value	Mean p-value	Cluster size [mm^{3}]
NFE Activation				
Cerebellum, Uvula	17±15/-68±5/-23±3	4.666	0.0004	2251
Caudate nucleus (body)	15±3/25±2/11±3	5.189	0.0002	630
Insula/ IFG	-29±14/20±6/7±9	5.209	0.0002	7584
Thalamus (VAN)	13±5/-5±5/-5±5	4.387	0.0006	1008
Brainstem/midbrain	2±7/-22±4/-11±2	4.588	0.0004	1834
Posterior Cingulate/ Retrosplenial Cortex	-8±3/-54±9/7±5	4.707	0.0004	3765
SMA	-5±3/6±6/51±8	4.684	0.0004	2137
Thalamus (VLN)	-18±3/-9±7/11±4	4.825	0.0003	2194
Hippocampus	-25±4/-12±7/-13±4	4.615	0.0004	2004
Entorhinal cortex	-26±4/8±4/-10±3	4.570	0.0005	726
Extrastriate cortex	-35±2/-57±3/12±5	4.516	0.0005	1177
NFE Deactivation				
IPL	49±6/-52±10/39±5	-4.422	0.0006	2165
DLPFC	44±3/31±5/26±5	-4.483	0.0005	1365
Superior frontal gyrus	30±5/18±4/50±	-4.510	0.0005	1891
NFS Activation				
PPA	29±4/-39±5/-14±5	4.592	0.0005	2518
Posterior Cingulate	15±4/-45±3/10±3	4.393	0.0006	783

Lingual Gyrus	13±3/-73±5/-10±4	4.456	0.0005	1613
Posterior Cingulate	-3±5/-49±4/5±6	4.580	0.0005	2519
PPA	-29±4/-35±5/-16±5	4.518	0.0005	3219
Insula/ IFG	-37±3/22±3/-1±2	4.373	0.0006	493
Premotor Cortex	-41±3/2±1/46±4	4.356	0.0006	529
NFS Deactivation				
Temporoparietal Junction	54±2/-49±3/28±4	-4.371	0.0006	672

Table B1 Whole brain analysis. Brain activity during neurofeedback blocks compared to rest for NFE and NFS group (positive x-coordinates denote right side of the brain); Monte Carlo simulation cluster threshold corrected at p < 0.001. Adapted from (Mehler et al., 2018).

Anatomical label	Mean ± SD coordinates [X/Y/Z]	Mean t-value	Mean p-value	Cluster size [mm^3]
NFS > NFE				
DLPFC	51±2/31±2/12±2	-3.545	0.0006	362
MTG	43±5/-71±4/25±4	-4.239	0.0002	2254
Premotor Cortex	27±9/0±6/49±4	-3.871	0.0003	4180
SPL	21±8/-61±6/49±6	-3.739	0.0004	4757
Precentral Gyrus	36±2/4±4/22±2	-3.820	0.0003	629
PPA, right	26±3/-36±2/-14±3	-3.957	0.0003	731
Lingual Gyrus	13±3/-73±5/-7±4	-3.789	0.0004	1224
Cingulate	15±3/-52±4/10±5	-3.683	0.0004	1191
Precuneus	-16±7/-64±4/47±7	-3.833	0.0004	3427

MTG	-40±9/-75±6/24±7	-3.832	0.0003	3069
PPA, left	-26±3/-39±4/-13±4	-3.762	0.0004	1340
MFG (BA46), extending into DLPFC	-44±4/34±5/16±5	-4.141	0.0002	3500
MFG (BA9)	-52±3/14±4/30±4	-3.975	0.0003	607
IFG	-61±3/38±3/-12±2	-3.803	0.0003	412
NFE > NFS				
MTG	60±3/-14±2/-10±3	3.988	0.0003	703
Insula	-29±3/-40±6/20±7	3.613	0.0005	1160
ITG	-52±3/-5±2/-28±3	3.371	0.0004	412

Table B2 Whole brain analysis. Activity during neurofeedback blocks for group contrast (group difference NFE>NFS; positive x-coordinates denote right side of the brain); Monte Carlo simulation cluster threshold corrected at p < 0.001. Adapted from (Mehler et al., 2018).

Anatomical label	Mean ± SD coordinates [X/Y/Z]	Mean t-value	Mean p-value	Cluster size [mm^{3}]
IFG, extending into DLPFC	47±6/4±5/33±8	3.011	0.0044	4997
Premotor Cortex	35±3/3±4/57±2	3.056	0.0039	661
Putamen	25±4/-10±4/11±3	2.966	0.0048	942
SMA	-2±3/-16±4/57±5	3.051	0.0038	1493
Caudate Body	-21±3/-7±8/18±3	2.865	0.0055	811

Table B3 Whole brain analysis. Within Factor (session); Monte Carlo simulation cluster threshold corrected at p < 0.01.

Anatomical label abbreviations in order of appearance. IFG: Inferior Frontal Gyrus; VAN: Ventral Anterior Nucleus of the Thalamus; SMA: Supplementary Motor Area; IPL: Inferior Parietal Lobe; VLN: Ventral Lateral Nucleus of the Thalamus; DLPFC: Dorsolateral Prefrontal Cortex; SFG: Superior Frontal Gyrus; MTG: Middle Temporal Gyrus; SPL: Superior Parietal Lobe; MFG: Middle Frontal Gyrus; ITG: Inferior Temporal Gyrus; BA: Brodmann area.

10.3 Appendix C

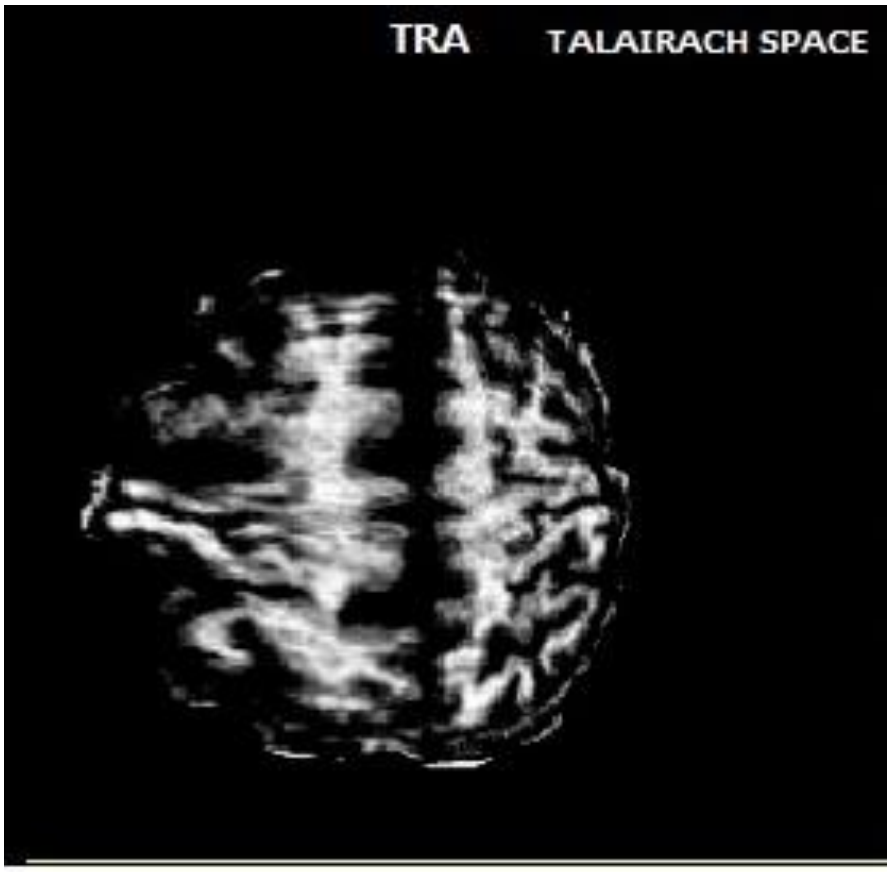

Figure C1 Example anatomical (T1) image in Talairach space. Artefact due to head motion, deformation due to lesion (left).

Figure C2 View of control interface of Turbo-BrainVoyager (TBV) (Version 3.0, Brain innovation, Maastricht, The Netherlands). Red rectangular on the left indicates target region selection, respective time course shown in the top right. Green background indicates task period, bar graph (green) percent signal change. Underneath display of six rigid body motion parameters. Example of excessive head motion that resulted in spurious activation within target area, but not in post-hoc chosen control area (blue circle) close to ventricles.

11 BIBLIOGRAPHY

Abbasi, O., Hirschmann, J., Storzer, L., Özkurt, T.E., Elben, S., Vesper, J., Wojtecki, L., Schmitz, G., Schnitzler, A., Butz, M., 2018. Unilateral deep brain stimulation suppresses alpha and beta oscillations in sensorimotor cortices. Neuroimage 174, 201–207. https://doi.org/10.1016/j.neuroimage.2018.03.026

Abraham, A., Pedregosa, F., Eickenberg, M., Gervais, P., Mueller, A., Kossaifi, J., Gramfort, A., Thirion, B., Varoquaux, G., 2014. Machine learning for neuroimaging with scikit-learn. Front. Neuroinform. 8, 14. https://doi.org/10.3389/fninf.2014.00014

Ahlskog, J., Muenter, M., 2001. Frequency of levodopa related dyskinesia and motor fluctuations as estimated from the cumulative literature. Mov Disord 16, 448–58.

Airaksinen, K., Butorina, A., Pekkonen, E., Nurminen, J., Taulu, S., Ahonen, A., Schnitzler, A., Mäkelä, J.P., 2012. Somatomotor mu rhythm amplitude correlates with rigidity during deep brain stimulation in Parkinsonian patients. Clin. Neurophysiol. 123, 2010–2017. https://doi.org/10.1016/j.clinph.2012.03.004

Ajimsha, M.S., Majeed, N.A., Chinnavan, E., Thulasyammal, R.P., 2014. Effectiveness of autogenic training in improving motor performances in parkinson's disease. Complement. Ther. Med. 22, 419–425. https://doi.org/10.1016/j.ctim.2014.03.013

Akram, H., Georgiev, D., Mahlknecht, P., Hyam, J., Foltynie, T., Limousin, P., Jahanshahi, M., Hariz, M., Zrinzo, L., 2017. Subthalamic deep brain stimulation sweet spots and hyperdirect cortical connectivity in Parkinson's disease. Neuroimage 158, 332–345. https://doi.org/10.1016/j.neuroimage.2017.07.012

Algermissen, J., Mehler, D.M.A., 2018. May the power be with you: are there highly powered studies in neuroscience, and how can we get more of them? J. Neurophysiol. 119, 2114–2117. https://doi.org/10.1152/jn.00765.2017

Alper, B.S., Malone-Moses, M., McLellan, J.S., Prasad, K., Manheimer, E., 2015. Thrombolysis in acute ischaemic stroke: time for a rethink? BMJ 350, h1075–h1075. https://doi.org/10.1136/bmj.h1075

Altman, D.G., Doré, C.J., 1990. Randomisation and baseline comparisons in clinical trials. Lancet (London, England) 335, 149–53.

Aminoff, E.M., Kveraga, K., Bar, M., 2013. The role of the parahippocampal cortex in

cognition. Trends Cogn. Sci. 17, 379–390. https://doi.org/10.1016/j.tics.2013.06.009

Anderson, I.M., Nutt, D.J., Deakin, J.F., 2000. Evidence-based guidelines for treating depressive disorders with antidepressants: a revision of the 1993 British Association for Psychopharmacology guidelines.British Association for Psychopharmacology. J Psychopharmacol. 14, 3–20. https://doi.org/10.1177/026988110001400101

Andrew, E., 1994. A Proposal for Structured Reporting of Randomized Controlled Trials. JAMA J. Am. Med. Assoc. 272, 1926. https://doi.org/10.1001/jama.1994.03520240054041

Andrews, G., 2001. Should depression be managed as a chronic disease? Bmj 322, 419. https://doi.org/10.1136/bmj.322.7283.419

Antonini, A., Tolosa, E., Mizuno, Y., Yamamoto, M., Poewe, W.H., 2009. A reassessment of risks and benefits of dopamine agonists in Parkinson's disease. Lancet Neurol. 8, 929–937. https://doi.org/10.1016/S1474-4422(09)70225-X

Aouad, P., Hughes, A., Valecha, N., Gawarikar, Y., Ahmad, K., O'Neil, R., Guduguntla, M., Lueck, C., 2013. Prevalence, Comorbid Associations and Prognostic Value of the Hyperdense Middle Cerebral Artery Sign. ISRN Stroke 2013, 1–6. https://doi.org/10.1155/2013/954825

Arnone, D., McIntosh, A.M., Ebmeier, K.P., Munafò, M.R., Anderson, I.M., 2012. Magnetic resonance imaging studies in unipolar depression: Systematic review and meta-regression analyses. Eur. Neuropsychopharmacol. 22, 1–16. https://doi.org/10.1016/j.euroneuro.2011.05.003

Arns, M., Batail, J.-M., Bioulac, S., Congedo, M., Daudet, C., Drapier, D., Fovet, T., Jardri, R., Le-Van-Quyen, M., Lotte, F., Mehler, D., Micoulaud-Franchi, J.-A., Purper-Ouakil, D., Vialatte, F., 2017. Neurofeedback: One of today's techniques in psychiatry? Encephale. 43, 135–145. https://doi.org/10.1016/j.encep.2016.11.003

Arns, M., Heinrich, H., Strehl, U., 2014. Evaluation of neurofeedback in ADHD: The long and winding road. Biol. Psychol. 95, 108–115. https://doi.org/10.1016/j.biopsycho.2013.11.013

Arns, M., Strehl, U., 2013. Evidence for efficacy of neurofeedback in ADHD? Am. J. Psychiatry 170, 799–800. https://doi.org/10.1176/appi.ajp.2013.13020208

Asendorpf, J.B., Conner, M., De Fruyt, F., De Houwer, J., Denissen, J.J.A., Fiedler, K., Fiedler, S., Funder, D.C., Kliegl, R., Nosek, B.A., Perugini, M., Roberts, B.W., Schmitt, M., van Aken, M.A.G., Weber, H., Wicherts, J.M., 2013. Recommendations for Increasing Replicability in Psychology. Eur. J. Pers. 27, 108–119. https://doi.org/10.1002/per.1919

Ashby, D., 2009. Growth rates in epidemic models: application to a model for HIV/AIDS progression. Stat. Med. 28, 221–239. https://doi.org/10.1002/sim

Ashby, G., 2011. Statistical Analysis of fMRI Data, 1st ed. MIT Press.

Auer, T., Dewiputri, W.I., Frahm, J., Schweizer, R., 2016. Higher-order brain areas associated with real-time functional MRI neurofeedback training of the somato-motor cortex. Neuroscience. https://doi.org/10.1016/j.neuroscience.2016.04.034

Ault, A., 2003. Climbing a medical everest. Science (80-.). 300, 2024–2025. https://doi.org/10.1126/science.300.5628.2024

Bagby, R.M., Quilty, L.C., Segal, Z. V, McBride, C.C., Kennedy, S.H., Costa, P.T., 2008. Personality and differential treatment response in major depression: a randomized controlled trial comparing cognitive-behavioural therapy and pharmacotherapy. Can. J. Psychiatry. 53, 361–70. https://doi.org/10.1037/a0013262.Open

Bagby, R.M., Taylor, G.J., Parkers, J.D. a, 1994. The twenty-item Toronto Alexithymia Scale-II. Convergent , Discriminant, and Concurrent Validity. J. Psychosom. Res. 38, 33–40.

Baguley, T., 2010. Book Review. Br. J. Math. Stat. Psychol. 63, 695–698. https://doi.org/10.1348/000711009X481027

Bajaj, S., Butler, A.J., Drake, D., Dhamala, M., 2015. Brain effective connectivity during motor-imagery and execution following stroke and rehabilitation. NeuroImage Clin. 8, 572–582. https://doi.org/10.1016/j.nicl.2015.06.006

Bandura, A., 1982. Self-efficacy mechanism in human agency. Dev. Psychol. 37, 122–147. https://doi.org/10.1037/0003-066X.37.2.122

Bar, M., 2009. A cognitive neuroscience hypothesis of mood and depression. Trends Cogn. Sci. 13, 456–463. https://doi.org/10.1016/j.tics.2009.08.009

Barclay-Goddard, R., Stevenson, T., Thalman, L., Poluha, W., 2011. Mental practice for treating upper extremity deficits in individuals with hemiparesis after stroke. Stroke 42, 574–576. https://doi.org/10.1161/STROKEAHA.111.627414

Barry, G., Galna, B., Rochester, L., 2014. The role of exergaming in Parkinson's disease rehabilitation: A systematic review of the evidence. J. Neuroeng. Rehabil. 11.

Bastos, A.M., Vezoli, J., Bosman, C.A., Schoffelen, J.M., Oostenveld, R., Dowdall, J.R., DeWeerd, P., Kennedy, H., Fries, P., 2015. Visual areas exert feedforward and feedback influences through distinct frequency channels. Neuron 85, 390–401. https://doi.org/10.1016/j.neuron.2014.12.018

Baune, B.T., Stuart, M., Gilmour, A., Wersching, H., Heindel, W., Arolt, V., Berger, K., 2012. The relationship between subtypes of depression and cardiovascular disease: A systematic

review of biological models. Transl. Psychiatry 2, e92-19. https://doi.org/10.1038/tp.2012.18

Beard, D., Hamilton, D.F., Davies, L., Cook, J.A., Beard, D., Depart-, N., 2017. Evidence-Based Evaluation of Practice and Innovation in Rehabilitation Using the IDEAL-Evidence-Based Evaluation of Practice and Innovation in Physical Therapy. Phys. Ther. 98, 108–121. https://doi.org/10.1016/j.apmr.2017.08.206

Beck, A., 2008. The evolution of the cognitive model of depression and its neurobiological correlates. Am. J. Psychiatry 165, 969–977. https://doi.org/10.1176/appi.ajp.2008.08050721

Beckmann, C.F., Smith, S.M., 2004. Probabilistic Independent Component Analysis for Functional Magnetic Resonance Imaging. IEEE Trans. Med. Imaging 23, 137–152. https://doi.org/10.1109/TMI.2003.822821

Begg, C., Cho, M., Eastwood, S., Horton, R., Moher, D., Olkin, I., Pitkin, R., Rennie, D., Schulz, K.F., Simel, D., Stroup, D.F., 1996. Improving the quality of reporting of randomized controlled trials. The CONSORT statement. JAMA 276, 637–9.

Begley, C.G., Ellis, L.M., 2012. Drug development: Raise standards for preclinical cancer research. Nature 483, 531–533. https://doi.org/10.1038/483531a

Behrens, T.E.J., Jenkinson, M., Robson, M.D., Smith, S.M., Johansen-Berg, H., 2006. A consistent relationship between local white matter architecture and functional specialisation in medial frontal cortex. Neuroimage 30, 220–227. https://doi.org/10.1016/j.neuroimage.2005.09.036

Benabid, A.L., Chabardes, S., Mitrofanis, J., Pollak, P., 2009. Deep brain stimulation of the subthalamic nucleus for the treatment of Parkinson's disease. Lancet Neurol. 8, 67–81. https://doi.org/10.1016/S1474-4422(08)70291-6

Benjamini, Y., Yekutieli, D., 2001a. The Control of the False Discovery Rate in Multiple Testing under Dependency. Ann. Stat. https://doi.org/10.2307/2674075

Benjamini, Y., Yekutieli, D., 2001b. The Control of the False Discovery Rate in Multiple Testing under Dependency. Ann. Stat. https://doi.org/10.2307/2674075

Berger, J.O., Sellke, T., 1987. Testing a point null hypothesis: The irreconcilability of P values and evidence. J. Am. Stat. Assoc. 82, 112–122. https://doi.org/10.1080/01621459.1987.10478397

Bergman, H., Wichmann, T., Karmon, B., DeLong, M.R., 1994. The primate subthalamic nucleus. II. Neuronal activity in the MPTP model of parkinsonism. J. Neurophysiol. 72, 507–520. https://doi.org/10.1152/jn.1994.72.2.507

Berlim, M.T., Van Den Eynde, F., Jeff Daskalakis, Z., 2013. Clinically meaningful efficacy and acceptability of low-frequency repetitive transcranial magnetic stimulation (rTMS) for treating primary major depression: A meta-analysis of randomized, double-blind and sham-controlled trials. Neuropsychopharmacology 38, 543–551. https://doi.org/10.1038/npp.2012.237

Berlim, M.T., Van Den Eynde, F., Tovar-Perdomo, S., Daskalakis, Z.J., 2014. Response, remission and drop-out rates following high-frequency repetitive transcranial magnetic stimulation (rTMS) for treating major depression: A systematic review and meta-analysis of randomized, double-blind and sham-controlled trials. Psychol. Med. 44, 225–239. https://doi.org/10.1017/S0033291713000512

Berman, B.D., Horovitz, S.G., Venkataraman, G., Hallett, M., 2012. Self-modulation of primary motor cortex activity with motor and motor imagery tasks using real-time fMRI-based neurofeedback. Neuroimage 59, 917–925. https://doi.org/10.1016/j.neuroimage.2011.07.035

Berto, P., D'ilario, D., Ruffo, P., Di Virgilio, R., Rizzo, F., 2000. Depression: cost-of-illness studies in the international literature, a review. J Ment Heal. Policy Econ 3, 3–10.

Berton, O., Nestler, E.J., 2006. New approaches to antidepressant drug discovery: Beyond monoamines. Nat. Rev. Neurosci. 7, 137–151. https://doi.org/10.1038/nrn1846

Bestmann, S., Swayne, O., Blankenburg, F., Ruff, C.C., Teo, J., Weiskopf, N., Driver, J., Rothwell, J.C., Ward, N.S., 2010. The Role of Contralesional Dorsal Premotor Cortex after Stroke as Studied with Concurrent TMS-fMRI. J. Neurosci. 30, 11926–11937. https://doi.org/10.1523/JNEUROSCI.5642-09.2010

Birbaumer, N., Ramos Murguialday, A., Weber, C., Montoya, P., 2009. Chapter 8 Neurofeedback and Brain-Computer Interface. Clinical Applications. Int. Rev. Neurobiol. 86, 107–117. https://doi.org/10.1016/S0074-7742(09)86008-X

Birbaumer, N., Ruiz, S., Sitaram, R., 2013. Learned regulation of brain metabolism. Trends Cogn. Sci. 17, 295–302. https://doi.org/10.1016/j.tics.2013.04.009

Biundo, R., Weis, L., Antonini, A., 2016. Cognitive decline in Parkinson's disease: the complex picture. npj Park. Dis. 2, 16018. https://doi.org/10.1038/npjparkd.2016.18

Blefari, M.L., Sulzer, J., Hepp-Reymond, M.-C., Kollias, S., Gassert, R., 2015. Improvement in precision grip force control with self-modulation of primary motor cortex during motor imagery. Front. Behav. Neurosci. 9, 18. https://doi.org/10.3389/fnbeh.2015.00018

Blicher, J.U., Stagg, C.J., O'Shea, J., Østergaard, L., MacIntosh, B.J., Johansen-Berg, H., Jezzard, P., Donahue, M.J., 2012. Visualization of altered neurovascular coupling in

chronic stroke patients using multimodal functional MRI. J. Cereb. Blood Flow Metab. 32, 2044–2054. https://doi.org/10.1038/jcbfm.2012.105

Bloom, T., Ganley, E., Winker, M., 2014. Data Access for the Open Access Literature: PLOS's Data Policy. PLoS Biol. 12, 1–3. https://doi.org/10.1371/journal.pbio.1001797

Bogousslavsky, J., Regli, F., 1990. Anterior cerebral artery territory infarction in the Lausanne Stroke Registry. Arch. Neurol. 47, 144–150.

Bonita, R., Beaglehole, R., 1988. Recovery of motor function after stroke. Stroke 19, 1497–500.

Bosch, G., 2018. Train PhD students to be thinkers not just specialists. Nature 554, 277. https://doi.org/10.1038/d41586-018-01853-1

Bothwell, L.E., Greene, J.A., Podolsky, S.H., Jones, D.S., 2016. Assessing the Gold Standard — Lessons from the History of RCTs. N. Engl. J. Med. 374, 2175–2181.

Bowers, W.A., 1990. Treatment of depressed in-patients. Cognitive therapy plus medication, relaxation plus medication and medication alone. Br. J. Psychiatry 156, 73–78. https://doi.org/10.1192/bjp.156.1.73

Bright, M.G., Murphy, K., 2017. Cleaning up the fMRI time series: Mitigating noise with advanced acquisition and correction strategies. Neuroimage 154, 1–3. https://doi.org/10.1016/j.neuroimage.2017.03.056

Brinkman, L., Stolk, A., Marshall, T.R., Esterer, S., Sharp, P., Dijkerman, H.C., de Lange, F.P., Toni, I., 2016. Independent Causal Contributions of Alpha- and Beta-Band Oscillations during Movement Selection. J. Neurosci. 36.

Broetz, D., Braun, C., Weber, C., Soekadar, S.R., Caria, A., Birbaumer, N., 2010. Combination of brain-computer interface training and goal-directed physical therapy in chronic stroke: A case report. Neurorehabil. Neural Repair 24, 674–679. https://doi.org/10.1177/1545968310368683

Brookings, J.B., Wilson, G.F., Swain, C.R., 1996. Psychophysiological responses to changes in workload during simulated air traffic control. Biol. Psychol. 42, 361–377. https://doi.org/10.1016/0301-0511(95)05167-8

Brown, P., Oliviero, A., Mazzone, P., Insola, A., Tonali, P., Di Lazzaro, V., 2001. Dopamine dependency of oscillations between subthalamic nucleus and pallidum in Parkinson's disease. J. Neurosci. 21, 1033–1038. https://doi.org/21/3/1033 [pii]

Brugger, F., Galovic, M., Weder, B.J., Kägi, G., 2015. Supplementary Motor Complex and Disturbed Motor Control - a Retrospective Clinical and Lesion Analysis of Patients after Anterior Cerebral Artery Stroke. Front. Neurol. 6, 209.

https://doi.org/10.3389/fneur.2015.00209

Brühl, A.B., Scherpiet, S., Sulzer, J., Stämpfli, P., Seifritz, E., Herwig, U., 2014. Real-time Neurofeedback Using Functional MRI Could Improve Down-Regulation of Amygdala Activity During Emotional Stimulation : A Proof-of-Concept Study 138–148. https://doi.org/10.1007/s10548-013-0331-9

Brunoni, A.R., Chaimani, A., Moffa, A.H., Razza, L.B., Gattaz, W.F., Daskalakis, Z.J., Carvalho, A.F., 2017. Repetitive transcranial magnetic stimulation for the acute treatment of major depressive episodes a systematic review with network meta-analysis. JAMA Psychiatry 74, 143–152. https://doi.org/10.1001/jamapsychiatry.2016.3644

Buch, E.R., Modir Shanechi, A., Fourkas, A.D., Weber, C., Birbaumer, N., Cohen, L.G., 2012. Parietofrontal integrity determines neural modulation associated with grasping imagery after stroke. Brain 135, 596–614. https://doi.org/10.1093/brain/awr331

Buhle, J.T., Silvers, J.A., Wager, T.D., Lopez, R., Onyemekwu, C., Kober, H., Weber, J., Ochsner, K.N., 2014. Cognitive Reappraisal of Emotion: A Meta-Analysis of Human Neuroimaging Studies. Cereb. Cortex 24, 2981–2990. https://doi.org/10.1093/cercor/bht154

Bullmore, E., Brammer, M., Williams, S.C.R., Rabe-Hesketh, S., Janot, N., David, A., Mellers, J., Howard, R., Sham, P., 1996. Statistical methods of estimation and inference for functional MR image analysis. Mag. Res. Med. 35, 261–277.

Busse, M., Quinn, L., Drew, C., Kelson, M., Trubey, R., McEwan, K., Jones, C., Townson, J., Dawes, H., Tudor-Edwards, R., Rosser, A., Hood, K., 2017. Physical Activity Self-Management and Coaching Compared to Social Interaction in Huntington Disease: Results From the ENGAGE-HD Randomized, Controlled Pilot Feasibility Trial. Phys. Ther. 97, 625–639. https://doi.org/10.4088/JCP.14m09298

Button, K.S., Ioannidis, J.P.A., Mokrysz, C., Nosek, B.A., Flint, J., Robinson, E.S.J., Munafò, M.R., 2013. Power failure: why small sample size undermines the reliability of neuroscience. Nat. Rev. Neurosci. 14, 365–376. https://doi.org/10.1038/nrn3475

Buyukturkoglu, K., Rana, M., Ruiz, S., Hackley, S.A., Soekadar, S.R., Birbaumer, N., Sitaram, R., 2013. Volitional Regulation of the Supplementary Motor Area with fMRI - BCI neurofeedback in Parkinson ' s Disease : A Pilot Study 6–8.

Canty, A., Ripley, B., 2017. boot: Boostrap R (S-Plus) Functions.

Carp, J., 2012a. On the Plurality of (Methodological) Worlds: Estimating the Analytic Flexibility of fMRI Experiments. Front. Neurosci. 6, 149. https://doi.org/10.3389/fnins.2012.00149

Carp, J., 2012b. The secret lives of experiments: Methods reporting in the fMRI literature. Neuroimage 63, 289–300. https://doi.org/10.1016/j.neuroimage.2012.07.004

Cartoni, E., Balleine, B., Baldassarre, G., 2016. Appetitive Pavlovian-instrumental Transfer: A review. Neurosci. Biobehav. Rev. 71, 829–848. https://doi.org/10.1016/j.neubiorev.2016.09.020

Cassidy, M., Mazzone, P., Oliviero, A., Insola, A., Tonali, P., Lazzaro, V. Di, Brown, P., 2002. Movement- related changes in synchronization in the human basal ganglia. Brain 125, 1235–1246. https://doi.org/10.1093/brain/awf135

Cervera, M.A., Soekadar, S.R., Ushiba, J., Millan, J. del R., Liu, M., Birbaumer, N., Garipelli, G., 2017. Brain-Computer Interfaces for Post-Stroke Motor Rehabilitation: A Meta-Analysis. bioRxiv 224618. https://doi.org/10.1101/224618

Chalela, J.A., Kidwell, C.S., Nentwich, L.M., Luby, M., Butman, J.A., Demchuk, A.M., Hill, M.D., Patronas, N., Latour, L., Warach, S., 2007. Magnetic resonance imaging and computed tomography in emergency assessment of patients with suspected acute stroke: a prospective comparison. Lancet 369, 293–298.

Chambers, C., 2017. The seven deadly sins of psychology: A manifesto for reforming the culture of scientific practice. Princeton university Press.

Chambers, C.D., 2015. Ten reasons why journals must review manuscripts before results are known. Addiction 110, 10–11. https://doi.org/10.1111/add.12728

Chambers, C.D., 2013. Registered Reports: A new publishing initiative at Cortex. Cortex 49, 609–610. https://doi.org/10.1016/j.cortex.2012.12.016

Chambers, C.D., Garavan, H., Bellgrove, M.A., 2009. Insights into the neural basis of response inhibition from cognitive and clinical neuroscience. Neurosci. Biobehav. Rev. 33, 631–646. https://doi.org/10.1016/j.neubiorev.2008.08.016

Chan, A.W., Altman, D.G., 2005. Epidemiology and reporting of randomised trials published in PubMed journals. Lancet 365, 1159–1162. https://doi.org/10.1016/S0140-6736(05)71879-1

Chang, L.J., Yarkoni, T., Khaw, M.W., Sanfey, A.G., 2013. Decoding the role of the insula in human cognition: Functional parcellation and large-scale reverse inference. Cereb. Cortex 23, 739–749. https://doi.org/10.1093/cercor/bhs065

Chapin, J.K., Moxon, K.A., Markowitz, R.S., Nicolelis, M.A., 1999. Real-time control of a robot arm using simultaneously recorded neurons in the motor cortex. Nat. Neurosci. 2, 664–670. https://doi.org/10.1038/10223

Chaudhary, U., Birbaumer, N., Ramos-Murguialday, A., 2016. Brain-computer interfaces for

communication and rehabilitation. Nat. Rev. Neurol. 12, 513–525. https://doi.org/10.1038/nrneurol.2016.113

Chaudhary, U., Xia, B., Silvoni, S., Cohen, L.G., Birbaumer, N., 2017. Brain–Computer Interface–Based Communication in the Completely Locked-In State. PLoS Biol. 15, 1–25. https://doi.org/10.1371/journal.pbio.1002593

Chen, G., Taylor, P.A., Cox, R.W., 2017. Is the statistic value all we should care about in neuroimaging? Neuroimage 147, 952–959. https://doi.org/10.1016/j.neuroimage.2016.09.066

Chen, L.S., Eaton, W.W., Gallo, J.J., Nestadt, G., 2000. Understanding the heterogeneity of depression through the triad of symptoms, course and risk factors: A longitudinal, population-based study. J. Affect. Disord. 59, 1–11. https://doi.org/10.1016/S0165-0327(99)00132-9

Cheon, E.J., Koo, B.H., Choi, J.H., 2016. The Efficacy of Neurofeedback in Patients with Major Depressive Disorder: An Open Labeled Prospective Study. Appl. Psychophysiol. Biofeedback 41, 103–110. https://doi.org/10.1007/s10484-015-9315-8

Chiew, M., LaConte, S.M., Graham, S.J., 2012. Investigation of fMRI neurofeedback of differential primary motor cortex activity using kinesthetic motor imagery. Neuroimage 61, 21–31. https://doi.org/10.1016/j.neuroimage.2012.02.053

Chiken, S., Nambu, A., 2015. Mechanism of Deep Brain Stimulation: Inhibition, Excitation, or Disruption? Neuroscientist 22, 313–322. https://doi.org/10.1177/1073858415581986

Choi, S.W., Chi, S.E., Chung, S.Y., Kim, J.W., Ahn, C.Y., Kim, H.T., 2010. Is alpha wave neurofeedback effective with randomized clinical trials in depression? A pilot study. Neuropsychobiology 63, 43–51. https://doi.org/10.1159/000322290

Chou, Y., Hickey, P.T., Sundman, M., Song, A.W., Chen, N., 2015. Effects of Repetitive Transcranial Magnetic Stimulation on Motor Symptoms in Parkinson Disease. JAMA Neurol. 72, 432. https://doi.org/10.1001/jamaneurol.2014.4380

Chu, H.Y., McIver, E.L., Kovaleski, R.F., Atherton, J.F., Bevan, M.D., 2017. Loss of Hyperdirect Pathway Cortico-Subthalamic Inputs Following Degeneration of Midbrain Dopamine Neurons. Neuron 95, 1306–1318.e5. https://doi.org/10.1016/j.neuron.2017.08.038

Cipriani, A., Furukawa, T.A., Salanti, G., Chaimani, A., Atkinson, L.Z., Ogawa, Y., Leucht, S., Ruhe, H.G., Turner, E.H., Higgins, J.P.T., Egger, M., Takeshima, N., Hayasaka, Y., Imai, H., Shinohara, K., Tajika, A., Ioannidis, J.P.A., Geddes, J.R., 2018. Comparative efficacy and acceptability of 21 antidepressant drugs for the acute treatment of adults with

major depressive disorder: a systematic review and network meta-analysis. Lancet 391, 1357–1366. https://doi.org/10.1016/S0140-6736(17)32802-7

Cirstea, M.C., Levin, M.F., 2000. Compensatory strategies for reaching in stroke. Brain 123 (Pt 5, 940–953. https://doi.org/10.1093/brain/awg266

Cirstea, M.C., Mitnitski, a. B., Feldman, a. G., Levin, M.F., 2003. Interjoint coordination dynamics during reaching in stroke. Exp. Brain Res. 151, 289–300. https://doi.org/10.1007/s00221-003-1438-0

Clarke, C.E., Patel, S., Ives, N., Rick, C.E., Dowling, F., Woolley, R., Wheatley, K., Walker, M.F., Sackley, C.M., 2016. Physiotherapy and occupational Therapy vs No Therapy in mild to moderate Parkinson disease. JAMA Neurol. 73, 291–299. https://doi.org/10.1001/jamaneurol.2015.4452

Confalonieri, L., Pagnoni, G., Barsalou, L.W., Rajendra, J., Eickhoff, S.B., Butler, A.J., 2012. Brain Activation in Primary Motor and Somatosensory Cortices during Motor Imagery Correlates with Motor Imagery Ability in Stroke Patients. ISRN Neurol. 2012, 1–17. https://doi.org/10.5402/2012/613595

Connolly, B.S., Lang, A.E., 2014. Pharmacological treatment of Parkinson disease: A review. JAMA - J. Am. Med. Assoc. 311, 1670–1683. https://doi.org/10.1001/jama.2014.3654

Corbetta, D., Sirtori, V., Castellini, G., Moja, L., Gatti, R., 2015. Constraint-induced movement therapy for upper extremities in people with stroke (Review) SUMMARY OF FINDINGS FOR THE MAIN COMPARISON. Cochrane Database Syst. Rev. https://doi.org/10.1002/14651858.CD004433.pub3.www.cochranelibrary.com

Cordes, J.S., Mathiak, K.A., Dyck, M., Alawi, E.M., Gaber, T.J., Zepf, F.D., Klasen, M., Zvyagintsev, M., 2015. Cognitive and neural strategies during control of the anterior cingulate cortex by fMRI neurofeedback in patients with schizophrenia 9, 1–10. https://doi.org/10.3389/fnbeh.2015.00169

Coronado-Montoya, S., Levis, A.W., Kwakkenbos, L., Steele, R.J., Turner, E.H., Thombs, B.D., 2016. Reporting of positive results in randomized controlled trials of mindfulness-based mental health interventions. PLoS One 11, 1–18. https://doi.org/10.1371/journal.pone.0153220

Cortese, S., Ferrin, M., Brandeis, D., Holtmann, M., Aggensteiner, P., Daley, D., Santosh, P., Simonoff, E., Stevenson, J., Stringaris, A., Sonuga-Barke, E.J.S., Asherson, P., Banaschewski, T., Brandeis, D., Buitelaar, J., Coghill, D., Cortese, S., Daley, D., Danckaerts, M., Dittmann, R.W., Döpfner, M., Ferrin, M., Hollis, C., Holtmann, M., Konofal, E., Lecendreux, M., Rothenberger, A., Santosh, P., Sergeant, J.A., Simonoff, E.,

Sonuga-Barke, E.J., Soutullo, C., Steinhausen, H.C., Stevenson, J., Stringaris, A., Taylor, E., van der Oord, S., Wong, I., Zuddas, A., 2016. Neurofeedback for Attention-Deficit/Hyperactivity Disorder: Meta-Analysis of Clinical and Neuropsychological Outcomes From Randomized Controlled Trials. J. Am. Acad. Child Adolesc. Psychiatry 55, 444–455. https://doi.org/10.1016/j.jaac.2016.03.007

Coull, J.T., 2004. Functional Anatomy of the Attentional Modulation of Time Estimation. Science (80-.). 303, 1506–1508. https://doi.org/10.1126/science.1091573

Cox, W.M., Subramanian, L., Linden, D.E.J., Lührs, M., McNamara, R., Playle, R., Hood, K., Watson, G., Whittaker, J.R., Sakhuja, R., Ihssen, N., 2016. Neurofeedback training for alcohol dependence versus treatment as usual: study protocol for a randomized controlled trial. Trials 17, 480. https://doi.org/10.1186/s13063-016-1607-7

Cramer, S.C., 2018. Treatments to Promote Neural Repair after Stroke. J. Stroke 20, 57–70. https://doi.org/10.5853/jos.2017.02796

Cremers, H.R., Wager, T.D., Yarkoni, T., 2017. The relation between statistical power and inference in fMRI. PLoS One 12, 1–20. https://doi.org/10.1371/journal.pone.0184923

Cristea, I.A., Ioannidis, J.P.A., 2018. P values in display items are ubiquitous and almost invariably significant : A survey of top science journals. PLoS One 1–15. https://doi.org/10.1371/journal.pone.0197440

Cuijpers, P., Berking, M., Andersson, G., Quigley, L., Kleiboer, A., Dobson, K.S., 2013a. A meta-analysis of cognitive-behavioural therapy for adult depression, alone and in comparison with other treatments. Can. J. Psychiatry 58, 376–385. https://doi.org/10.1177/070674371305800702

Cuijpers, P., Sijbrandij, M., Koole, S.L., Andersson, G., Beekman, A.T., Reynolds, C.F., 2014. Adding psychotherapy to antidepressant medication in depression and anxiety disorders: a meta-analysis. World Psychiatry 13, 56–67. https://doi.org/10.1002/wps.20089

Cuijpers, P., Sijbrandij, M., Koole, S.L., Andersson, G., Beekman, A.T., Reynolds, C.F., 2013b. The efficacy of psychotherapy and pharmacotherapy in treating depressive and anxiety disorders: A meta-analysis of direct comparisons. World Psychiatry 12, 137–148. https://doi.org/10.1002/wps.20038

Cuijpers, P., Smit, F., Bohlmeijer, E., Hollon, S.D., Andersson, G., 2010. Efficacy of cognitive-behavioural therapy and other psychological treatments for adult depression: Meta-analytic study of publication bias. Br. J. Psychiatry 196, 173–178. https://doi.org/10.1192/bjp.bp.109.066001

Cuijpers, P., van Straten, A., Smit, F., Mihalopoulos, C., Beekman, A., 2008. Preventing the

onset of depressive disorders: a meta-analytic review of psychological interventions. Am. J. Psychiatry 165, 1272–1280. https://doi.org/10.1176/appi.ajp.2008.07091422

Cuijpers, P., Ven Straten, A., Warmerdam, L., Andersson, G., 2009. Psychotherapy versus the combination of psychotherapy and pharmacotherapy in the treatment of depression: A meta-analysis. Depress. Anxiety 26, 279–288. https://doi.org/10.1002/da.20519

Cunnington, R., Windischberger, C., Deecke, L., Moser, E., 2002. The Preparation and Execution of Self-Initiated and Externally-Triggered Movement: A Study of Event-Related fMRI. Neuroimage 15, 373–385. https://doi.org/10.1006/nimg.2001.0976

Dayan, E., Cohen, L.G., 2011. Neuroplasticity subserving motor skill learning. Neuron 72, 443–454. https://doi.org/10.1016/j.neuron.2011.10.008

De Angelis, C., Drazen, J.M., Frizelle, F.A., Haug, C., Hoey, J., Horton, R., Kotzin, S., Laine, C., Marusic, A., Overbeke, A.J.P.M., Schroeder, T. V., Sox, H.C., Weyden, M.B. Van Der, 2004. Clinical Trial Registration: A Statement from the International Committee of Medical Journal Editors. N. Engl. J. Med. 351, 1250–1251. https://doi.org/10.1056/NEJMe048225

De Borst, A.W., Sack, A.T., Jansma, B.M., Esposito, F., De Martino, F., Valente, G., Roebroeck, A., Di Salle, F., Goebel, R., Formisano, E., 2012. Integration of "what" and "where" in frontal cortex during visual imagery of scenes. Neuroimage 60, 47–58. https://doi.org/10.1016/j.neuroimage.2011.12.005

De Hemptinne, C., Swann, N.C., Ostrem, J.L., Ryapolova-Webb, E.S., San Luciano, M., Galifianakis, N.B., Starr, P.A., 2015. Therapeutic deep brain stimulation reduces cortical phase-amplitude coupling in Parkinson's disease. Nat. Neurosci. 18, 779–786. https://doi.org/10.1038/nn.3997

de Lange, F.P., Helmich, R.C., Toni, I., 2006. Posture influences motor imagery: An fMRI study. Neuroimage 33, 609–617. https://doi.org/10.1016/j.neuroimage.2006.07.017

Defebvre, L., Bourriez, J., Dujardin, K., Derambure, P., 1994. Spatiotemporal study of Bereitschaftspotential and event-related desynchronization during voluntary movement in Parkinson's disease. Brain Topogr.

DeFelice, C., Tassi, R., De Capua, B., Jaubert, F., Gentile, M., Quartulli, L., Tonni, G., Costantini, D., Strambi, M., Latini, G., 2007. A New Phenotypical Variant of Intrauterine Growth Restriction? Pediatrics 119, e983–e990. https://doi.org/10.1542/peds.2006-2523

Delacre, M., Lakens, D., Leys, C., 2017. Why Psychologists Should by Default Use Welch's t-test Instead of Student's t-test. Int. Rev. Soc. Psychol. 30, 92. https://doi.org/10.5334/irsp.82

Delaloye, S., Holtzheimer, P.E., 2014. Deep brain stimulation in the treatment of depression. Dialogues Clin. Neurosci. 16, 83–91.

Delgadillo, J., McMillan, D., Lucock, M., Leach, C., Ali, S., Gilbody, S., 2014. Early changes, attrition, and dose-response in low intensity psychological interventions. Br. J. Clin. Psychol. 53, 114–130. https://doi.org/10.1111/bjc.12031

Demyttenaere, K., 1997. Compliance during treatment with antidepressants. J. Affect. Disord. 43, 27–39. https://doi.org/10.1016/S0165-0327(96)00095-X

Desmond, J.E., Glover, G.H., 2002. Estimating sample size in functional MRI (fMRI) neuroimaging studies: Statistical power analyses. J. Neurosci. Methods 118, 115–128. https://doi.org/10.1016/S0165-0270(02)00121-8

Dettmers, C., Nedelko, V., Schoenfeld, M.A., 2015. Impact of left versus right hemisphere subcortical stroke on the neural processing of action observation and imagery. Restor. Neurol. Neurosci. 33, 701–712. https://doi.org/10.3233/RNN-140487

Devos, D., Defebvre, L., 2006. Effect of deep brain stimulation and l-Dopa on electrocortical rhythms related to movement in Parkinson's disease. Prog. Brain Res. 159, 331–349. https://doi.org/10.1016/S0079-6123(06)59022-3

Devos, D., Labyt, E., Derambure, P., Bourriez, J.L., Cassim, F., Guieu, J.D., Destée, A., Defebvre, L., 2003. Effect of L-Dopa on the pattern of movement-related (de)synchronisation in advanced Parkinson's disease. Neurophysiol. Clin. 33, 203–212. https://doi.org/10.1016/j.neucli.2003.10.001

Dienes, Z., 2014. Using Bayes to get the most out of non-significant results. Front. Psychol. 5, 781. https://doi.org/10.3389/fpsyg.2014.00781

Dierckx, B., Heijnen, W.T., van den Broek, W.W., Birkenhäger, T.K., 2012. Efficacy of electroconvulsive therapy in bipolar versus unipolar major depression: A meta-analysis. Bipolar Disord. 14, 146–150. https://doi.org/10.1111/j.1399-5618.2012.00997.x

Dimyan, M.A., Cohen, L.G., 2011. Neuroplasticity in the context of motor rehabilitation after stroke. Nat. Rev. Neurol. 7, 76–85. https://doi.org/10.1038/nrneurol.2010.200

Dinga, R., Schmaal, L., Penninx, B., Tol, M.J. Van, Veltman, D.J., Velzen, L. Van, Wee, N. Van Der, Marquand, A., 2018. Evaluating the evidence for biotypes of depression: attempted replication of Drysdale et.al. 2017. bioRxiv.

Disner, S.G., Beevers, C.G., Haigh, E.A.P., Beck, A.T., 2011. Neural mechanisms of the cognitive model of depression. Nat. Rev. Neurosci. 12, 467–477. https://doi.org/10.1038/nrn3027

Dosenbach, N.U.F., Koller, J.M., Earl, E.A., Miranda-Dominguez, O., Klein, R.L., Van, A.N.,

Snyder, A.Z., Nagel, B.J., Nigg, J.T., Nguyen, A.L., Wesevich, V., Greene, D.J., Fair, D.A., 2017. Real-time motion analytics during brain MRI improve data quality and reduce costs. Neuroimage 161, 80–93. https://doi.org/10.1016/j.neuroimage.2017.08.025

Drysdale, A.T., Grosenick, L., Downar, J., Dunlop, K., Mansouri, F., Meng, Y., Fetcho, R.N., Zebley, B., Oathes, D.J., Etkin, A., Schatzberg, A.F., Sudheimer, K., Keller, J., Mayberg, H.S., Gunning, F.M., Alexopoulos, G.S., Fox, M.D., Pascual-Leone, A., Voss, H.U., Casey, B., Dubin, M.J., Liston, C., 2016. Resting-state connectivity biomarkers define neurophysiological subtypes of depression. Nat. Med. 23. https://doi.org/10.1038/nm.4246

Dum, R.P., Strick, P.L., 2002. Motor areas in the frontal lobe of the primate. Physiol. Behav. 77, 677–682.

Dumas-Mallet, E., Button, K.S., Boraud, T., Gonon, F., Munafò, M.R., 2017. Low statistical power in biomedical science: a review of three human research domains. R. Soc. Open Sci. 4.

Dunlop, B.W., Mayberg, H.S., 2014. Neuroimaging-based biomarkers for treatment selection in major depressive disorder. Dialogues Clin. Neurosci. 16, 479–490.

Easterbrook, P.., Gopalan, R., Berlin, J.., Matthews, D.., 1991. Publication bias in clinical research. Lancet 337, 867–872. https://doi.org/10.1016/0140-6736(91)90201-Y

Ebbesen, C.L., Brecht, M., 2017. Motor cortex — to act or not to act? Nat. Rev. Neurosci. 18, 694–705. https://doi.org/10.1038/nrn.2017.119

Editorial, 2017. Promoting reproducibility with registered reports. Nat. Hum. Behav. 1, 0034. https://doi.org/10.1038/s41562-016-0034

Eglen, S.J., Marwick, B., Halchenko, Y.O., Hanke, M., Sufi, S., Gleeson, P., Silver, R.A., Davison, A.P., Lanyon, L., Abrams, M., Wachtler, T., Willshaw, D.J., Pouzat, C., Poline, J.-B., 2017. Toward standard practices for sharing computer code and programs in neuroscience. Nat. Neurosci. 20, 770–773. https://doi.org/10.1038/nn.4550

Eickhoff, S.B., Stephan, K.E., Mohlberg, H., Grefkes, C., Fink, G.R., Amunts, K., Zilles, K., 2005. A new SPM toolbox for combining probabilistic cytoarchitectonic maps and functional imaging data. Neuroimage 25, 1325–1335. https://doi.org/10.1016/j.neuroimage.2004.12.034

Ejaz, N., Hamada, M., Diedrichsen, J., 2015. Hand use predicts the structure of representations in sensorimotor cortex. Nat. Neurosci. 18, 1034–1040. https://doi.org/10.1038/nn.4038

Eklund, A., Nichols, T.E., Knutsson, H., 2016. Cluster failure : Why fMRI inferences for spatial extent have inflated false-positive rates. https://doi.org/10.1073/pnas.1602413113

Emmert, K., Kopel, R., Koush, Y., Maire, R., Senn, P., Van De Ville, D., Haller, S., 2017. Continuous vs. intermittent neurofeedback to regulate auditory cortex activity of tinnitus patients using real-time fMRI - A pilot study. NeuroImage Clin. 14, 97–104. https://doi.org/10.1016/j.nicl.2016.12.023

Emmert, K., Kopel, R., Sulzer, J., Brühl, A., Berman, B., 2016. Meta-analysis of real-time fMRI neurofeedback studies using individual participant data: How is brain regulation mediated? Neuroimage.

Emmert, K., Kopel, R., Sulzer, J., Brühl, A.B., Berman, B.D., Linden, D.E.J., Horovitz, S.G., Breimhorst, M., Caria, A., Frank, S., Johnston, S., Long, Z., Paret, C., Robineau, F., Veit, R., Bartsch, A., Beckmann, C.F., Van De Ville, D., Haller, S., 2016. Meta-analysis of real-time fMRI neurofeedback studies using individual participant data: How is brain regulation mediated? Neuroimage 124, 806–812. https://doi.org/10.1016/j.neuroimage.2015.09.042

Engel, A.K., Fries, P., 2010. Beta-band oscillations—signalling the status quo? Curr. Opin. Neurobiol. 20, 156–165. https://doi.org/10.1016/j.conb.2010.02.015

Engel, A.K., Moll, C.K.E., Fried, I., Ojemann, G.A., 2005. Invasive recordings from the human brain: Clinical insights and beyond. Nat. Rev. Neurosci. 6, 35–47. https://doi.org/10.1038/nrn1585

Epstein, R., Graham, K.S., Downing, P.E., 2003. Viewpoint-specific scene representations in human parahippocampal cortex. Neuron 37, 865–876. https://doi.org/10.1016/S0896-6273(03)00117-X

Epstein, R., Kanwisher, N., 1998. A cortical representation of the local visual environment. Nature 598–601.

Esmail, S., Linden, D., 2014. Neural Networks and Neurofeedback in Parkinson's Disease. NeuroRegulation 1, 240–272. https://doi.org/10.15540/nr.1.3-4.240

Evans, A.C., Marrett, S., Neelin, P., Collins, L., Worsley, K., Dai, W., Milot, S., Meyer, E., Bub, D., 1992. Anatomical mapping of functional activation in stereotactic coordinate space. Neuroimage 1, 43–53. https://doi.org/10.1016/1053-8119(92)90006-9

Fanelli, D., 2012. Negative results are disappearing from most disciplines and countries. Scientometrics 90, 891–904. https://doi.org/10.1007/s11192-011-0494-7

Faul, F., Erdfelder, E., Lang, A.-G., Axel Buchner, 2007. G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. Behav. Res. Methods 39, 175–191. https://doi.org/10.3758/BF03193146

Favre, I., Zeffiro, T.A., Detante, O., Krainik, A., Hommel, M., Jaillard, A., 2014. Upper limb

recovery after stroke is associated with ipsilesional primary motor cortical activity: A meta-analysis. Stroke 45, 1077–1083. https://doi.org/10.1161/STROKEAHA.113.003168

Feigin, V.L., Forouzanfar, M.H., Krishnamurthi, R., Mensah, G.A., Connor, M., Bennett, D.A., Moran, A.E., Sacco, R.L., Anderson, L., Truelsen, T., O'Donnell, M., Venketasubramanian, N., Barker-Collo, S., Lawes, C.M.M., Wang, W., Shinohara, Y., Witt, E., Ezzati, M., Naghavi, M., 2014. Global and regional burden of stroke during 1990-2010: Findings from the Global Burden of Disease Study 2010. Lancet 383, 245–255. https://doi.org/10.1016/S0140-6736(13)61953-4

Feigin, V.L., Lawes, C.M.M., Bennett, D.A., Anderson, C.S., 2003. Stroke epidemiology: A review of population-based studies of incidence, prevalence, and case-fatality in the late 20th century. Lancet Neurol. 2, 43–53. https://doi.org/10.1016/S1474-4422(03)00266-7

Feyerabend, P., 1993. Against method, 3rd ed, Pool-108-46-235-15.Nycmny.Fios. ….

Fischer, P., Pogosyan, A., Cheeran, B., Green, A.L., Aziz, T.Z., Hyam, J., Little, S., Foltynie, T., Limousin, P., Zrinzo, L., Hariz, M., Samuel, M., Ashkan, K., Brown, P., Tan, H., 2017. Subthalamic nucleus beta and gamma activity is modulated depending on the level of imagined grip force. Exp. Neurol. 293, 53–61. https://doi.org/10.1016/j.expneurol.2017.03.015

Flanagan, J.C., 1982. Measurement of quality of life: current state of the art. Arch. Phys. Med. Rehabil. 63, 56–9.

Fontanarosa, P., Bauchner, H., Flanagin, A., 2017. Authorship and Team Science. JAMA 318, 2433. https://doi.org/10.1001/jama.2017.19341

Foster, E.D., Deardorff, A., 2017. Open Science Framework (OSF). J. Med. Libr. Assoc. 105, 203–206. https://doi.org/10.5195/JMLA.2017.88

Fournier, J.C., DeRubeis, R.J., Hollon, S.D., Dimidjian, S., Amsterdam, J.D., Shelton, R.C., Fawcett, J., 2010. Antidepressant Drug Effects and Depression Severity. JAMA 303, 47. https://doi.org/10.1001/jama.2009.1943

Franco, A., Malhotra, N., Simonovits, G., 2014. Social science. Publication bias in the social sciences: unlocking the file drawer. Science 345, 1502–5. https://doi.org/10.1126/science.1255484

Franklin, G., Carson, A.J., Welch, K.A., 2016. Cognitive behavioural therapy for depression: Systematic review of imaging studies. Acta Neuropsychiatr. 28, 61–74. https://doi.org/10.1017/neu.2015.41

Friedman, L.M., Furberg, C.D., DeMets, D.L., 2010. Fundamentals of Clinical Trials, 4th ed.

Springer Science + Business Media, LLC. https://doi.org/10.1007/978-1-4419-1586-3

Friston, K.J., Holmes, A.P., Worsley, K.J., Poline, J., Frith, C.D., Frackowiak, R.S.J., 1994. Statistical parametric maps in functional imaging: A general linear approach. Hum. Brain Mapp. 2.

Gage, H., Storey, L., 2004. Rehabilitation for Parkinson's disease: A systematic review of available evidence. Clin. Rehabil. 18, 463–482. https://doi.org/10.1191/0269215504cr764oa

Gamble, C., Krishan, A., Stocken, D., Lewis, S., Juszczak, E., Doré, C., Williamson, P.R., Altman, D.G., Montgomery, A., Lim, P., Berlin, J., Senn, S., Day, S., Barbachano, Y., Loder, E., 2017. Guidelines for the Content of Statistical Analysis Plans in Clinical Trials. JAMA 318, 2337. https://doi.org/10.1001/jama.2017.18556

García Carrasco, D., Aboitiz Cantalapiedra, J., 2016. Effectiveness of motor imagery or mental practice in functional recovery after stroke: a systematic review. Neurol. (English Ed. 31, 43–52. https://doi.org/10.1016/j.nrleng.2013.02.008

Gatev, P., Darbin, O., Wichmann, T., 2006. Oscillations in the basal ganglia under normal conditions and in movement disorders. Mov. Disord. 21, 1566–1577. https://doi.org/10.1002/mds.21033

Geddes, J., Price, J., McKnight, R., 2012. Mood Disorders, in: Psychiatry. Oxford University Press, pp. 221–245.

Gelman, A., Stern, H., 2006. The difference between "significant" and "not significant" is not itself statistically significant. Am. Stat. 60, 328–331. https://doi.org/10.1198/000313006X152649

Geranmayeh, F., Wise, R.J.S., Leech, R., Murphy, K., 2015. Measuring vascular reactivity with breath-holds after stroke: A method to aid interpretation of group-level BOLD signal changes in longitudinal fMRI studies. Hum. Brain Mapp. 36, 1755–1771. https://doi.org/10.1002/hbm.22735

Gerin, M.I., Fichtenholtz, H., Roy, A., Walsh, C.J., Krystal, J.H., Southwick, S., Hampson, M., 2016. Real-Time fMRI Neurofeedback with War Veterans with Chronic PTSD: A Feasibility Study 7, 1–11. https://doi.org/10.3389/fpsyt.2016.00111

Goetz, C.G., Tilley, B.C., Shaftman, S.R., Stebbins, G.T., Fahn, S., Martinez-Martin, P., Poewe, W., Sampaio, C., Stern, M.B., Dodel, R., Dubois, B., Holloway, R., Jankovic, J., Kulisevsky, J., Lang, A.E., Lees, A., Leurgans, S., LeWitt, P.A., Nyenhuis, D., Olanow, C.W., Rascol, O., Schrag, A., Teresi, J.A., van Hilten, J.J., LaPelle, N., 2008. Movement Disorder Society-sponsored revision of the Unified Parkinson's Disease Rating Scale

(MDS-UPDRS): Scale presentation and clinimetric testing results. Mov. Disord. 23, 2129–2170. https://doi.org/10.1002/mds.22340

Golding, K., Fife-Schaw, C., Kneebone, I., 2017. Twelve month follow-up on a randomised controlled trial of relaxation training for post-stroke anxiety. Clin. Rehabil. 31, 1164–1167. https://doi.org/10.1177/0269215516682820

Gotlib, I.H., Krasnoperova, E., Yue, D.N., Joormann, J., 2004. Attentional Biases for Negative Interpersonal Stimuli in Clinical Depression. J. Abnorm. Psychol. 113, 127–135. https://doi.org/10.1037/0021-843X.113.1.121

Gracco, V.L., Tremblay, P., Pike, B., 2005. Imaging speech production using fMRI. Neuroimage 26, 294–301. https://doi.org/10.1016/j.neuroimage.2005.01.033

Gratwicke, J., Jahanshahi, M., Foltynie, T., 2015. Parkinson's disease dementia: A neural networks perspective. Brain 138, 1454–1476. https://doi.org/10.1093/brain/awv104

Greenland, S., Senn, S.J., Rothman, K.J., Carlin, J.B., Poole, C., Goodman, S.N., Altman, D.G., 2016. Statistical tests, P values, confidence intervals, and power: a guide to misinterpretations. Eur. J. Epidemiol. 31, 337–350. https://doi.org/10.1007/s10654-016-0149-3

Grefkes, C., Ward, N.S., 2014. Cortical reorganization after stroke: How much and how functional? Neuroscientist 20, 56–70. https://doi.org/10.1177/1073858413491147

Grosse-Wentrup, M., Janzing, D., Siegel, M., Schölkopf, B., 2016. Identification of causal relations in neuroimaging data with latent confounders: An instrumental variable approach. Neuroimage 125, 825–833. https://doi.org/10.1016/j.neuroimage.2015.10.062

Grunhaus, L., Dannon, P.N., Schreiber, S., Dolberg, O.H., Amiaz, R., Ziv, R., Lefkifker, E., 2000. Repetitive transcranial magnetic stimulation is as effective as electroconvulsive therapy in the treatment of nondelusional major depressive disorder: An open study. Biol. Psychiatry 47, 314–324. https://doi.org/10.1016/S0006-3223(99)00254-1

Grunhaus, L., Schreiber, S., Dolberg, O.T., Polak, D., Dannon, P.N., 2003. A Randomized Controlled Comparison of Electroconvulsive Therapy and Repetitive Transcranial Magnetic Stimulation in Severe and Resistant Nonpsychotic Major Depression 3223.

Guglielmi, G., 2018. The world's largest set of brain scans are helping reveal the workings of the mind and how diseases ravage the brain. Science (80-.). https://doi.org/10.1126/science.aat0994

Guillot, A., Collet, C., Nguyen, V.A., Malouin, F., Richards, C., Doyon, J., 2009. Brain activity during visual versus kinesthetic imagery: An fMRI study. Hum. Brain Mapp. 30, 2157–2172. https://doi.org/10.1002/hbm.20658

Gupta, S.K., 2011. Intention-to-treat concept: A review. Perspect. Clin. Res. 2, 109–12. https://doi.org/10.4103/2229-3485.83221

Haar, S., Dinstein, I., Shelef, I., Donchin, O., 2017a. Effector-Invariant Movement Encoding in the Human Motor System. J. Neurosci. 37, 1663–17. https://doi.org/10.1523/JNEUROSCI.1663-17.2017

Haar, S., Donchin, O., Dinstein, I., 2017b. Individual Movement Variability Magnitudes Are Explained by Cortical Neural Variability. J. Neurosci. 37, 9076–9085. https://doi.org/10.1523/JNEUROSCI.1650-17.2017

Habes, I., Rushton, S., Johnston, S.J., Sokunbi, M.O., Barawi, K., Brosnan, M., Daly, T., Ihssen, N., Linden, D.E.J., 2016. fMRI neurofeedback of higher visual areas and perceptual biases. Neuropsychologia 85, 208–215. https://doi.org/10.1016/j.neuropsychologia.2016.03.031

Halder, S., Varkuti, B., Bogdan, M., Kübler, A., Rosenstiel, W., Sitaram, R., Birbaumer, N., 2013. Prediction of brain-computer interface aptitude from individual brain structure. Front. Hum. Neurosci. 7, 1–9. https://doi.org/10.3389/fnhum.2013.00105

Halje, P., Tamte, M., Richter, U., Mohammed, M., Cenci, M.A., Petersson, P., 2012. Levodopa-Induced Dyskinesia Is Strongly Associated with Resonant Cortical Oscillations. J. Neurosci. 32, 16541–16551. https://doi.org/10.1523/JNEUROSCI.3047-12.2012

Hamilton, J.P., Glover, G.H., Bagarinao, E., Chang, C., Mackey, S., Sacchet, M.D., Gotlib, I.H., 2016. Effects of salience-network-node neurofeedback training on affective biases in major depressive disorder. Psychiatry Res. 249, 91–96. https://doi.org/10.1016/j.pscychresns.2016.01.016

Hamilton, M., 1980. Rating depressive patients. J. Clin. Psychiatry.

Hamilton, M., 1960. Scale for depression. J Neurol Neurosurg Psychiatry 23, 56–63. https://doi.org/10.1136/jnnp.23.1.56

Hammond, C., Bergman, H., Brown, P., 2007. Pathological synchronization in Parkinson's disease: networks, models and treatments. Trends Neurosci. 30, 357–364. https://doi.org/10.1016/j.tins.2007.05.004

Hampson, M., Stoica, T., Saksa, J., Scheinost, D., Qiu, M., Bhawnani, J., Pittenger, C., 2012. Real-time fMRI Biofeedback Targeting the Orbitofrontal Cortex for Contamination Anxiety 1–10. https://doi.org/10.3791/3535

Hanakawa, T., Dimyan, M.A., Hallett, M., 2008. Motor Planning, Imagery, and Execution in the Distributed Motor Network: A Time-Course Study with Functional MRI. Cereb.

Cortex 18, 2775–2788. https://doi.org/10.1093/cercor/bhn036

Harris, J.E., Hebert, A., 2015. Utilization of motor imagery in upper limb rehabilitation: a systematic scoping review. Clin. Rehabil. 29, 1092–107. https://doi.org/10.1177/0269215514566248

Hartwell, K.J., Hanlon, C.A., Li, X., Borckardt, J.J., Canterberry, M., Prisciandaro, J.J., Maria, M.M.M., Lematty, T., George, M.S., Brady, K.T., 2016. Individualized real-time fMRI neurofeedback to attenuate craving in nicotine-dependent smokers 41, 48–55. https://doi.org/10.1503/jpn.140200

Hedna, V.S., Bodhit, A.N., Ansari, S., Falchook, A.D., Stead, L., Heilman, K.M., Waters, M.F., 2013. Hemispheric differences in ischemic stroke: Is left-hemisphere stroke more common? J. Clin. Neurol. 9, 97–102. https://doi.org/10.3988/jcn.2013.9.2.97

Heijnen, W.T., Birkenhäger, T.K., Wierdsma, A.I., Van Den Broek, W.W., 2010. Antidepressant pharmacotherapy failure and response to subsequent electroconvulsive therapy: A meta-analysis. J. Clin. Psychopharmacol. 30, 616–619. https://doi.org/10.1097/JCP.0b013e3181ee0f5f

Heinsius, T., Bogousslavsky, J., Van Melle, G., 1998. Large infarcts in the middle cerebral artery territory. Etiology and outcome patterns.[Erratum appears in Neurology 1998 Jun;50(6):1940-3]. Neurology 50, 341–350.

Hellrung, L., Dietrich, A., Hollmann, M., Pleger, B., Kalberlah, C., Roggenhofer, E., Villringer, A., Horstmann, A., 2018. Intermittent compared to continuous real-time fMRI neurofeedback boosts control over amygdala activation. Neuroimage 166, 198–208. https://doi.org/10.1016/J.NEUROIMAGE.2017.10.031

Hendriks, M.H.A., Daniels, N., Pegado, F., de Beeck, H.P.O., 2017. The effect of spatial smoothing on representational similarity in a simple motor paradigm. Front. Neurol. 8, 1–11. https://doi.org/10.3389/fneur.2017.00222

Héroux, M.E., Loo, C.K., Taylor, J.L., Gandevia, S.C., 2017. Questionable science and reproducibility in electrical brain stimulation research. PLoS One 12, e0175635. https://doi.org/10.1371/journal.pone.0175635

Hétu, S., Grégoire, M., Saimpont, A., Coll, M.-P., Eugène, F., Michon, P.-E., Jackson, P.L., 2013. The neural network of motor imagery: An ALE meta-analysis. Neurosci. Biobehav. Rev. 37, 930–949. https://doi.org/10.1016/j.neubiorev.2013.03.017

Heunis, S., Lamerichs, R., Zinger, S., Aldenkamp, B., Breeuwer, M., 2018. Quality and denoising in real-time fMRI neurofeedback : a methods review. Open Sci. Framew.

Hieronymus, F., Emilsson, J.F., Nilsson, S., Eriksson, E., 2016. Consistent superiority of

selective serotonin reuptake inhibitors over placebo in reducing depressed mood in patients with major depression. Mol. Psychiatry 21, 523–530. https://doi.org/10.1038/mp.2015.53

Higginson, A.D., Munafò, M.R., 2016. Current Incentives for Scientists Lead to Underpowered Studies with Erroneous Conclusions. PLOS Biol. 14, e2000995. https://doi.org/10.1371/journal.pbio.2000995

Hinds, O., Thompson, T.W., Ghosh, S., Yoo, J.J., Whitfield-Gabrieli, S., Triantafyllou, C., Gabrieli, J.D.E., 2013. Roles of default-mode network and supplementary motor area in human vigilance performance: evidence from real-time fMRI. J. Neurophysiol. 109, 1250–1258. https://doi.org/10.1152/jn.00533.2011

Hoenig, J.M., Heisey, D.M., 2001. The abuse of power: The pervasive fallacy of power calculations for data analysis. Am. Stat. 55, 19–24. https://doi.org/10.1198/000313001300339897

Hofmann, S.G., Asnaani, A., Vonk, I.J.J., Sawyer, A.T., Fang, A., 2012. The efficacy of cognitive behavioral therapy: A review of meta-analyses. Cognit. Ther. Res. 36, 427–440. https://doi.org/10.1007/s10608-012-9476-1

Hohenfeld, Christian Nellessen, N., Dogan, I., Kuhn, H., Müller, C., Papa, F., Ketteler, S., Goebel, R., Heinecke, A., Shah, J., Schulz, J., Reske, M., Reetz, K., 2017. Cognitive Improvement and Brain Changes after Real-Time Functional MRI Neurofeedback Training in Healthy Elderly and Prodromal Alzheimer's Disease 8, 1–15. https://doi.org/10.3389/fneur.2017.00384

Hollon, S.D., DeRubeis, R.J., Fawcett, J., Amsterdam, J.D., Shelton, R.C., Zajecka, J., Young, P.R., Gallop, R., 2014. Effect of cognitive therapy with antidepressant medications vs antidepressants alone on the rate of recovery in major depressive disorder: a randomized clinical trial. JAMA psychiatry 71, 1157–64. https://doi.org/10.1001/jamapsychiatry.2014.1054

Holmes, E.A., Blackwell, S.E., Burnett Heyes, S., Renner, F., Raes, F., 2016. Mental Imagery in Depression: Phenomenology, Potential Mechanisms, and Treatment Implications. Annu. Rev. Clin. Psychol. 12, 249–80. https://doi.org/10.1146/annurev-clinpsy-021815-092925

Holmes, E.A., Mathews, A., 2010. Mental imagery in emotion and emotional disorders. Clin. Psychol. Rev. 30, 349–362. https://doi.org/10.1016/j.cpr.2010.01.001

Holtzheimer, P.E., Husain, M.M., Lisanby, S.H., Taylor, S.F., Whitworth, L.A., Mcclintock, S., Slavin, K. V, Berman, J., Mckhann, G.M., Patil, P.G., Rittberg, B.R., Abosch, A.,

Pandurangi, A.K., Holloway, K.L., Lam, R.W., Honey, C., Neimat, J.S., Henderson, J.M., Debattista, C., Rothschild, A.J., Pilitsis, J.G., Espinoza, R.T., Petrides, G., Mogilner, A.Y., Matthews, K., Bs, D.P., Gross, R.E., Hamani, C., Lozano, A.M., Mayberg, H.S., 2017. Subcallosal cingulate deep brain stimulation for treatment-resistant depression: a multisite, randomised, sham-controlled trial. The Lancet Psychiatry 0366. https://doi.org/10.1016/S2215-0366(17)30371-1

Horvath, J.C., Carter, O., Forte, J.D., 2014. Transcranial direct current stimulation: five important issues we aren't discussing (but probably should be). Front. Syst. Neurosci. 8, 1–8. https://doi.org/10.3389/fnsys.2014.00002

Horváth, K., Aschermann, Z., Ács, P., Deli, G., Janszky, J., Komoly, S., Balázs, É., Takács, K., Karádi, K., Kovács, N., 2015. Minimal clinically important difference on the Motor Examination part of MDS-UPDRS. Parkinsonism Relat. Disord. 21, 1421–1426. https://doi.org/10.1016/j.parkreldis.2015.10.006

Hosp, J.A., Luft, A.R., 2013. Dopaminergic meso-cortical projections to M1: Role in motor learning and motor cortex plasticity. Front. Neurol. 4 OCT, 1–7. https://doi.org/10.3389/fneur.2013.00145

Hsu, W.Y., Cheng, C.H., Liao, K.K., Lee, I.H., Lin, Y.Y., 2012. Effects of repetitive transcranial magnetic stimulation on motor functions in patients with stroke: A meta-analysis. Stroke 43, 1849–1857. https://doi.org/10.1161/STROKEAHA.111.649756

Hunter, S.M., Johansen-Berg, H., Ward, N., Kennedy, N.C., Chandler, E., Weir, C.J., Rothwell, J., Wing, A.M., Grey, M.J., Barton, G., Leavey, N.M., Havis, C., Lemon, R.N., Burridge, J., Dymond, A., Pomeroy, V.M., 2018. Functional Strength Training and Movement Performance Therapy for Upper Limb Recovery Early Poststroke—Efficacy, Neural Correlates, Predictive Markers, and Cost-Effectiveness: FAST-INdiCATE Trial. Front. Neurol. 8, 1–24. https://doi.org/10.3389/fneur.2017.00733

Hutson, M., 2018. Artificial intelligence faces reproducibility crisis. Science (80-.). 359, 725–726. https://doi.org/10.1126/science.359.6377.725

Huys, Q.J.M., Gölzer, M., Friedel, E., Heinz, A., Cools, R., Dayan, P., Dolan, R.J., 2016. The specificity of Pavlovian regulation is associated with recovery from depression. Psychol. Med. 46, 1027–1035. https://doi.org/10.1017/S0033291715002597

Iadecola, C., 2004. Neurovascular regulation in the normal brain and in Alzheimer's disease. Nat. Rev. Neurosci. 5, 347–360. https://doi.org/10.1038/nrn1387

Ietswaart, M., Johnston, M., Dijkerman, H.C., Joice, S., Scott, C.L., MacWalter, R.S., Hamilton, S.J.C., 2011. Mental practice with motor imagery in stroke recovery:

randomized controlled trial of efficacy. Brain 134, 1373–86. https://doi.org/10.1093/brain/awr077

Insel, T.R., Cuthbert, B.N., 2015. CORTICOSPINAL AND CORTICORUBRAL PROJECTIONS FROM THE SUPPLEMENTARY MOTOR AREA IN THE MONKEY. Science 348, 499–500. https://doi.org/10.1126/science.aab2358

Ioannidis, J.P.A., 2018. The Proposal to Lower *P* Value Thresholds to .005. Jama 94305. https://doi.org/10.1001/jama.2018.1536

Ioannidis, J.P.A., 2012. Why Science Is Not Necessarily Self-Correcting. Perspect. Psychol. Sci. 7, 645–654. https://doi.org/10.1177/1745691612464056

Ioannidis, J.P.A., 2005. Why most published research findings are false. PLoS Med.

Ionta, S., Perruchoud, D., Draganski, B., Blanke, O., Luca, G. De, 2012. Body Context and Posture Affect Mental Imagery of Hands. PLoS One 7, e34382. https://doi.org/10.1371/journal.pone.0034382

Issakidis, C., Andrews, G., 2004. Pretreatment attrition and dropout in an outpatient clinic for anxiety disorders. Acta Psychiatr. Scand. 109, 426–433. https://doi.org/10.1111/j.1600-0047.2004.00264.x

Jackson-Lewis, V., Przedborski, S., 2007. Protocol for the MPTP mouse model of Parkinson's disease. Nat. Protoc. 2, 141–151. https://doi.org/10.1038/nprot.2006.342

Jackson, P.L., Lafleur, M.F., Malouin, F., Richards, C., Doyon, J., 2001. Potential role of mental practice using motor imagery in neurologic rehabilitation. Arch. Phys. Med. Rehabil. 82, 1133–1141. https://doi.org/10.1053/apmr.2001.24286

Jahfari, S., Waldorp, L., van den Wildenberg, W.P.M., Scholte, H.S., Ridderinkhof, K.R., Forstmann, B.U., 2011. Effective Connectivity Reveals Important Roles for Both the Hyperdirect (Fronto-Subthalamic) and the Indirect (Fronto-Striatal-Pallidal) Fronto-Basal Ganglia Pathways during Response Inhibition. J. Neurosci. 31, 6891–6899. https://doi.org/10.1523/JNEUROSCI.5253-10.2011

Jansen, I.G.H., Mulder, M.J.H.L., Goldhoorn, R.-J.B., 2018. Endovascular treatment for acute ischaemic stroke in routine clinical practice: prospective, observational cohort study (MR CLEAN Registry). Bmj k949. https://doi.org/10.1136/bmj.k949

Janssen, C.W., Lowry, C.A., Mehl, M.R., Allen, J.J.B., Kelly, K.L., Gartner, D.E., Medrano, A., Begay, T.K., Rentscher, K., White, J.J., Fridman, A., Roberts, L.J., Robbins, M.L., Hanusch, K., Cole, S.P., Raison, C.L., 2016. Whole-Body Hyperthermia for the Treatment of Major Depressive Disorder. JAMA Psychiatry 5, 1580. https://doi.org/10.1001/jamapsychiatry.2016.1031

Jenkinson, N., Brown, P., 2011. New insights into the relationship between dopamine, beta oscillations and motor function. Trends Neurosci. 34, 611–618. https://doi.org/10.1016/j.tins.2011.09.003

Jenkinson, N., Kühn, A.A., Brown, P., 2013. Gamma oscillations in the human basal ganglia. Exp. Neurol. 245, 72–76. https://doi.org/10.1016/j.expneurol.2012.07.005

Jia, Z., Peng, W., Chen, Z., Sun, H., Zhang, H., Kuang, W., Huang, X., Lui, S., Gong, Q., 2017. Magnetization Transfer Imaging of Treatment-resistant Depression. Radiology 284, 521–529. https://doi.org/10.1148/radiol.2017160820

Johansen-Berg, H., Rushworth, M.F.S., Bogdanovic, M.D., Kischka, U., Wimalaratna, S., Matthews, P.M., 2002. The role of ipsilateral premotor cortex in hand movement after stroke. Proc. Natl. Acad. Sci. 99, 14518–14523. https://doi.org/10.1073/pnas.222536799

Johnsen, T.J., Friborg, O., 2015. The effects of cognitive behavioral therapy as an anti-depressive treatment is falling: A meta-analysis. Psychol. Bull. 141, 747–768.

Johnson, D.A.W., 1981. Compliance with therpeutic regimens. Acta Psychiatr. Scand. 447–453.

Johnson, K.A., Hartwell, K., Lematty, T., Borckardt, J., Morgan, P.S., Govindarajan, K., Brady, K., George, M.S., 2012. Intermittent "Real-time" fMRI Feedback Is Superior to Continuous Presentation for a Motor Imagery Task: A Pilot Study. J. Neuroimaging 22, 58–66. https://doi.org/10.1111/j.1552-6569.2010.00529.x

Johnston, K.C., Li, J.Y., Lyden, P.D., Hanson, S.K., Feasby, T.E., Adams, R.J., Faught, R.E., Haley, E.C., 1998. Medical and Neurological Complications of Ischemic Stroke. Stroke.

Johnston, S., Linden, D., Healy, D., Goebel, R., 2011. Upregulation of emotion areas through neurofeedback with a focus on positive mood. Cogn. Affect.

Jones, C.W., Handler, L., Crowell, K.E., Keil, L.G., Weaver, M.A., Platts-Mills, T.F., 2013. Non-publication of large randomized clinical trials: cross sectional analysis. BMJ 347, f6104. https://doi.org/10.1136/BMJ.F6104

Jongsma, M.L.A., Meulenbroek, R.G.J., Okely, J., Baas, C.M., van der Lubbe, R.H.J., Steenbergen, B., 2013. Effects of Hand Orientation on Motor Imagery - Event Related Potentials Suggest Kinesthetic Motor Imagery to Solve the Hand Laterality Judgment Task. PLoS One 8, e76515. https://doi.org/10.1371/journal.pone.0076515

Jorm, A., Morgan, A., Hetrick, S., 2008. Relaxation for depression (Review). Cochrane Database Syst. Rev. https://doi.org/10.1002/14651858.CD007142.pub2.www.cochranelibrary.com

Joundi, R.A., Jenkinson, N., Brittain, J.-S., Aziz, T.Z., Brown, P., 2012. Driving oscillatory

activity in the human cortex enhances motor performance. Curr. Biol. 22, 403–7. https://doi.org/10.1016/j.cub.2012.01.024

Joyce, P.R., Paykel, E.S., 1989. Predictors of drug response in depression. Arch. Gen. Psychiatry 46, 89–99.

Kable, J.W., Caulfield, M.K., Falcone, M., McConnell, M., Bernardo, L., Parthasarathi, T., Cooper, N., Ashare, R., Audrain-McGovern, J., Hornik, R., Diefenbach, P., Lee, F.J., Lerman, C., 2017. No Effect of Commercial Cognitive Training on Brain Activity, Choice Behavior, or Cognitive Performance. J. Neurosci. 37, 7390–7402. https://doi.org/10.1523/JNEUROSCI.2832-16.2017

Kalu, U.G., Sexton, C.E., Loo, C.K., Ebmeier, K.P., 2012. Transcranial direct current stimulation in the treatment of major depression: A meta-analysis. Psychol. Med. 42, 1791–1800. https://doi.org/10.1017/S0033291711003059

Kaplan, R.M., Irvin, V.L., 2015. Likelihood of Null Effects of Large NHLBI Clinical Trials Has Increased over Time. PLoS One 10, e0132382. https://doi.org/10.1371/journal.pone.0132382

Karch, S., Keeser, D., Hümmer, S., Paolini, M., 2015. Modulation of Craving Related Brain Responses Using Real-Time fMRI in Patients with Alcohol Use Disorder. https://doi.org/10.1371/journal.pone.0133034

Kasahara, K., Hoshino, H., Furusawa, Y., Sayo DaSalla, C., Honda, M., Murata, M., Hanakawa, T., 2018. Initial experience with a sensorimotor rhythm-based brain-computer interface in a Parkinson's disease patient. Brain-Computer Interfaces 1–9. https://doi.org/10.1080/2326263X.2018.1440781

Kasess, C.H., Windischberger, C., Cunnington, R., Lanzenberger, R., Pezawas, L., Moser, E., 2008. The suppressive influence of SMA on M1 in motor imagery revealed by fMRI and dynamic causal modeling. Neuroimage 40, 828–837. https://doi.org/10.1016/j.neuroimage.2007.11.040

Kass, R.E., Raftery, A.E., 1995. Bayes Factors Bayes Factors. J. Am. Stat. Assoc. 90, 773–795.

Keller, M.B., Lavori, P.W., Mueller, T.I., Endicott, J., Coryell, W., Hirschfeld, R.M.A., Shea, T., 1992. Time to Recovery, Chronicity, and Levels of Psychopathology in Major Depression. Arch. Gen. Psychiatry 49, 809. https://doi.org/10.1001/archpsyc.1992.01820100053010

Kemlin, C., Moulton, E., Samson, Y., Rosso, C., 2016. Do Motor Imagery Performances Depend on the Side of the Lesion at the Acute Stage of Stroke? Front. Hum. Neurosci. 10,

321. https://doi.org/10.3389/fnhum.2016.00321

Kerr, N., 1998. HARKing: -hypnothesizing after the results are known. Personal. Soc. Psychol. Rev. 2, 196–217. https://doi.org/10.1207/s15327957pspr0203_4

Kessler, D., Angstadt, M., Sripada, C.S., 2017. Reevaluating "cluster failure" in fMRI using nonparametric control of the false discovery rate. Proc. Natl. Acad. Sci. U. S. A. 114, E3372–E3373. https://doi.org/10.1073/pnas.1614502114

Kessler, R.C., Berglund, P., Demler, O., Jin, R., Koretz, D., Merikangas, K.R., Rush, A.J., Walters, E.E., Wang, P.S., 2003. The Epidemiology of Major. Am. Med. Assoc. 289, 3095–3105.

Keynan, J.N., Meir-Hasson, Y., Gilam, G., Cohen, A., Jackont, G., Kinreich, S., Ikar, L., Or-Borichev, A., Etkin, A., Gyurak, A., Klovatch, I., Intrator, N., Hendler, T., 2016. Limbic Activity Modulation Guided by Functional Magnetic Resonance Imaging–Inspired Electroencephalography Improves Implicit Emotion Regulation. Biol. Psychiatry 80, 490–496. https://doi.org/10.1016/j.biopsych.2015.12.024

Khanna, P., Swann, N.C., de Hemptinne, C., Miocinovic, S., Miller, A., Starr, P.A., Carmena, J.M., 2017. Neurofeedback Control in Parkinsonian Patients Using Electrocorticography Signals Accessed Wirelessly With a Chronic, Fully Implanted Device. IEEE Trans. Neural Syst. Rehabil. Eng. 25, 1715–1724. https://doi.org/10.1109/TNSRE.2016.2597243

Khoury, B., Lecomte, T., Fortin, G., Masse, M., Therien, P., Bouchard, V., Chapleau, M.-A., Paquin, K., Hofmann, S.G., 2013. Mindfulness-based therapy: A comprehensive meta-analysis. Clin. Psychol. Rev. 33, 763–771. https://doi.org/10.1016/j.cpr.2013.05.005

Kiley, R., Peatfield, T., Hansen, J., Reddington, F., 2017. Data Sharing from Clinical Trials — A Research Funder's Perspective. N. Engl. J. Med. 377, 1990–1992. https://doi.org/10.1056/NEJMsb1708278

Kim, S.J., Moon, G.J., Bang, O.Y., 2013. Biomarkers for stroke. J. stroke 15, 27–37. https://doi.org/10.5853/jos.2013.15.1.27

Kirsch, I., Deacon, B.J., Huedo-Medina, T.B., Scoboria, A., Moore, T.J., Johnson, B.T., 2008. Initial Severity and Antidepressant Benefits: A Meta-Analysis of Data Submitted to the Food and Drug Administration. PLoS Med. 5, e45. https://doi.org/10.1371/journal.pmed.0050045

Kirschner, M., Sladky, R., Haugg, A., Stämpfli, P., Jehli, E., Engeli, E., Hösli, S., Baumgartner, M.R., Sulzer, J., Huys, Q.J.M., 2018. Self-regulation of the Dopaminergic Reward Circuit in Cocaine Users with Mental Imagery and Neurofeedback 1–29.

Klamroth, S., Steib, S., Devan, S., Pfeifer, K., 2016. Effects of exercise therapy on postural instability in Parkinson disease: A meta-analysis. J. Neurol. Phys. Ther. 40, 3–14. https://doi.org/10.1097/NPT.0000000000000117

Kober, S.E., Witte, M., Ninaus, M., Koschutnig, K., Wiesen, D., Zaiser, G., Neuper, C., Wood, G., 2017. Ability to Gain Control Over One's Own Brain Activity and its Relation to Spiritual Practice: A Multimodal Imaging Study. Front. Hum. Neurosci. 11, 271. https://doi.org/10.3389/fnhum.2017.00271

Kohn, N., Eickhoff, S.B., Scheller, M., Laird, A.R., Fox, P.T., Habel, U., 2014. Neural network of cognitive emotion regulation - An ALE meta-analysis and MACM analysis. Neuroimage 87, 345–355. https://doi.org/10.1016/j.neuroimage.2013.11.001

Koizumi, A., Amano, K., Cortese, A., Shibata, K., Yoshida, W., Seymour, B., Kawato, M., Lau, H., 2016. Fear reduction without fear through reinforcement of neural activity that bypasses conscious exposure. Nat. Hum. Behav. 1, 0006. https://doi.org/10.1038/s41562-016-0006

Kolshus, E., Jelovac, A., McLoughlin, D.M., 2017. Bitemporal v. high-dose right unilateral electroconvulsive therapy for depression: A systematic review and meta-analysis of randomized controlled trials. Psychol. Med. 47, 518–530. https://doi.org/10.1017/S0033291716002737

Koush, Y., Ashburner, J., Prilepin, E., Sladky, R., Zeidman, P., Bibikov, S., Scharnowski, F., Nikonorov, A., Van De Ville, D., 2017. Real-time fMRI data for testing OpenNFT functionality. Data Br. 14, 344–347. https://doi.org/10.1016/j.dib.2017.07.049

Kraft, E., Schaal, M.C., Lule, D., König, E., Scheidtmann, K., 2015. The functional anatomy of motor imagery after sub-acute stroke. NeuroRehabilitation 36, 329–337. https://doi.org/10.3233/NRE-151221

Krause, F., Benjamins, C., Lührs, M., Eck, J., Noirhomme, Q., Rosenke, M., Brunheim, S., Sorger, B., Goebel, R., 2017. Real-time fMRI-based self-regulation of brain activation across different visual feedback presentations. Brain-Computer Interfaces 1–15. https://doi.org/10.1080/2326263X.2017.1307096

Krause, F., Lindemann, O., 2014. Expyriment: A Python library for cognitive and neuroscientific experiments. Behav. Res. Methods 46, 416–428. https://doi.org/10.3758/s13428-013-0390-6

Kriegeskorte, N., Simmons, W.K., Bellgowan, P.S., Baker, C.I., 2009. Circular analysis in systems neuroscience: The dangers of double dipping. Nat. Neurosci. 12, 535–540. https://doi.org/10.1038/nn.2303

Krishnan, V., Nestler, E.J., 2008. The molecular neurobiology of depression. Nature 455, 894–902. https://doi.org/10.1038/nature07455

Kühberger, A., Fritz, A., Scherndl, T., 2014. Publication bias in psychology: A diagnosis based on the correlation between effect size and sample size. PLoS One 9, 1–8. https://doi.org/10.1371/journal.pone.0105825

Kühn, A.A., Doyle, L., Pogosyan, A., Yarrow, K., Kupsch, A., Schneider, G.H., Hariz, M.I., Trottenberg, T., Brown, P., 2006. Modulation of beta oscillations in the subthalamic area during motor imagery in Parkinson's disease. Brain 129, 695–706. https://doi.org/10.1093/brain/awh715

Kühn, A.A., Tsui, A., Aziz, T., Ray, N., Brücke, C., Kupsch, A., Schneider, G.H., Brown, P., 2009. Pathological synchronisation in the subthalamic nucleus of patients with Parkinson's disease relates to both bradykinesia and rigidity. Exp. Neurol. 215, 380–387. https://doi.org/10.1016/j.expneurol.2008.11.008

Kumral, E., Bayulkem, G., Evyapan, D., Yunten, N., 2002. Spectrum of anterior cerebral artery territory infarction: clinical and MRI findings. Eur. J. Neurol. 9, 615–624. https://doi.org/10.1046/j.1468-1331.2002.00452.x

Kwakkel, G., Kollen, B.J., 2013. Predicting activities after stroke: What is clinically relevant? Int. J. Stroke 8, 25–32. https://doi.org/10.1111/j.1747-4949.2012.00967.x

Lakens, D., 2017. Equivalence Tests - A practical primer for t tests, correlations, and meta-analyses. Soc. Psychol. Personal. Sci. 8, 355–362. https://doi.org/10.1177/1948550617697177

Lakens, D., 2013. Calculating and reporting effect sizes to facilitate cumulative science: a practical primer for t-tests and ANOVAs. Front. Psychol. 4, 863. https://doi.org/10.3389/fpsyg.2013.00863

Lakens, D., Scheel, A.M., Isager, P., 2018. Equivalence testing for psychological research: A tutorial. Adv. Methods Pract. Psychol. Sci. 1–11. https://doi.org/10.17605/OSF.IO/V3ZKT

Lancaster, J.L., Woldorff, M.G., Parsons, L.M., Liotti, M., Freitas, C.S., Rainey, L., Kochunov, P. V., Nickerson, D., Mikiten, S.A., Fox, P.T., 2000. Automated Talairach Atlas labels for functional brain mapping. Hum. Brain Mapp. 10, 120–131. https://doi.org/10.1002/1097-0193(200007)10:3<120::AID-HBM30>3.0.CO;2-8

Langhorne, P., Bernhardt, J., Kwakkel, G., 2011. Stroke rehabilitation. Lancet 377, 1693–1702. https://doi.org/10.1016/S0140-6736(11)60325-5

Langhorne, P., Coupar, F., Pollock, A., 2009. Motor recovery after stroke: a systematic review.

Lancet Neurol. 8, 741–754. https://doi.org/10.1016/S1474-4422(09)70150-4

Langner, R., Leiberg, S., Hoffstaedter, F., Eickhoff, S.B., 2018. Towards a human self-regulation system: Common and distinct neural signatures of emotional and behavioural control. Neurosci. Biobehav. Rev. https://doi.org/10.1016/j.neubiorev.2018.04.022

Lauritzen, M.J., 2005. Reading vascular changes in brain imaging: Is dendritic calcium the key? J. Cereb. Blood Flow Metab. 25, S684–S684. https://doi.org/10.1038/sj.jcbfm.9591524.0684

Lawrence, E.S., Coshall, C., Dundas, R., Stewart, J., Rudd, A.G., Howard, R., Wolfe, C.D., 2001. Estimates of the prevalence of acute stroke impairments and disability in a multiethnic population. Stroke 32, 1279–84.

Lenoski, B., Baxter, L.C., Karam, L.J., Maisog, J., Debbins, J., 2008. On the Performance of Autocorrelation Estimation Algorithms for fMRI Analysis. IEEE J. Sel. Top. Signal Process. 2, 828–838. https://doi.org/10.1109/JSTSP.2008.2007819

Lenze, E.J., Mulsant, B.H., Blumberger, D.M., Karp, J.F., Newcomer, J.W., Anderson, S.J., Dew, M.A., Butters, M.A., Stack, J.A., Begley, A.E., Reynolds, C.F., 2015. Efficacy, safety, and tolerability of augmentation pharmacotherapy with aripiprazole for treatment-resistant depression in late life: A randomised, double-blind, placebo-controlled trial. Lancet 386, 2404–2412. https://doi.org/10.1016/S0140-6736(15)00308-6

Lewis, S.J.G., Caldwell, M.A., Barker, R.A., 2003. Modern therapeutic approaches in Parkinson's disease. Expert Rev. Mol. Med. 5, 1–20. https://doi.org/10.1017/S1462399403006008

Leys, D., Pruvo, J.P., Godefroy, O., Rondepierre, P., Leclerc, X., 1992. Prevalence and significance of hyperdense middle cerebral artery in acute stroke. Stroke 23.

Li, X., Hartwell, K.J., Borckardt, J., Prisciandaro, J.J., Saladin, M.E., Morgan, P.S., Johnson, K.A., Lematty, T., Brady, K.T., George, M.S., 2012. Volitional reduction of anterior cingulate cortex activity produces decreased cue craving in smoking cessation : a preliminary real-time fMRI study 739–748. https://doi.org/10.1111/j.1369-1600.2012.00449.x

Liebeskind, D.S., 2003. Collateral circulation. Stroke 34, 2279–2284. https://doi.org/10.1161/01.STR.0000086465.41263.06

Liew, S.-L., Otr, L., Rana, M., Cornelsen, S., Fortunato, M., Filho, D.B., Birbaumer, N., Sitaram, R., Cohen, L.G., Soekadar, S.R., 2015. Improving Motor Corticothalamic Communication After Stroke Using Real-Time fMRI Connectivity-Based Neurofeedback. Neurorehabil. Neural Repair 30, 671–675.

https://doi.org/10.1177/1545968315619699

Lin, W.H., Hao, Q., Rosengarten, B., Leung, W.H., Wong, K.S., 2011. Impaired neurovascular coupling in ischaemic stroke patients with large or small vessel disease. Eur. J. Neurol. 18, 731–736. https://doi.org/10.1111/j.1468-1331.2010.03262.x

Linden, D., 2014. Neurofeedback and networks of depression. Dialogues Clin. Neurosci. 16, 103–112.

Linden, D., Habes, I., Johnston, S., Linden, S., 2012. Real-time self-regulation of emotion networks in patients with depression. PLoS One 7, e38115.

Linden, D.E.J., 2014. Neurofeedback and networks of depression. Dialogues Clin. Neurosci. 16, 103–112. https://doi.org/10.1016/j.siny.2015.10.004

Linden, D.E.J., 2012. The Challenges and Promise of Neuroimaging in Psychiatry. Neuron 73, 8–22. https://doi.org/10.1016/j.neuron.2011.12.014

Lindquist, M.A., 2008. The Statistical Analysis of fMRI Data. Stat. Sci. 23, 439–464. https://doi.org/10.1214/09-STS282

Little, S., Brown, P., 2014. The functional role of beta oscillations in Parkinson's disease. Park. Relat. Disord. 20, S44–S48. https://doi.org/10.1016/S1353-8020(13)70013-0

Llanos, C., Rodriguez, M., Rodriguez-Sabate, C., Morales, I., Sabate, M., 2013. Mu-rhythm changes during the planning of motor and motor imagery actions. Neuropsychologia 51, 1019–1026. https://doi.org/10.1016/j.neuropsychologia.2013.02.008

Logothetis, N.K., 2008. What we can do and what we cannot do with fMRI. Nature 453, 869–878. https://doi.org/10.1038/nature06976

Logothetis, N.K., 2002. The neural basis of the blood-oxygen-level-dependent functional magnetic resonance imaging signal. Philos. Trans. R. Soc. B Biol. Sci. 357, 1003–1037. https://doi.org/10.1098/rstb.2002.1114

Loken, E., Gelman, A., 2017. Measurement error and the replication crisis. Science (80-.). 355, 584–585. https://doi.org/10.1126/science.aal3618

Lopes da Silva, F., 2010. EEG: Origin and Measurement, in: Mulert, C., Lemieux, L. (Eds.), EEG-FMRI. Springer Berlin Heidelberg, pp. 19–38.

Lorenz, R., 2017. Neuroadaptive Bayesian optimization: Implications for the Cognitive Sciences. Imperial College London.

Lorenz, R., Hampshire, A., Leech, R., 2017. Neuroadaptive Bayesian optimization and hypothesis testing. Trends Cogn. Sci. 21, 155–167. https://doi.org/10.1016/j.tics.2017.01.006

Losseff, N., Brown, M., Joan Grieve, 2009. Stroke and Cerebrovascular Disease, in: Clarke,

C., Howard, R., Rossor, M., Shorvon, S. (Eds.), Neurology: A Queen Square Textbook. Wiley-Blackwell, pp. 109–154.

Lotze, M., Beutling, W., Loibl, M., Domin, M., Platz, T., Schminke, U., Byblow, W.D., 2012. Contralesional motor cortex activation depends on ipsilesional corticospinal tract integrity in well-recovered subcortical stroke patients. Neurorehabil. Neural Repair 26, 594–603. https://doi.org/10.1177/1545968311427706

Lotze, M., Markert, J., Sauseng, P., Hoppe, J., Plewnia, C., Gerloff, C., 2006. The role of multiple contralesional motor areas for complex hand movements after internal capsular lesion. J. Neurosci. 26, 6096–102. https://doi.org/10.1523/JNEUROSCI.4564-05.2006

Lotze, M., Montoya, P., Erb, M., Hülsmann, E., Flor, H., Klose, U., Birbaumer, N., Grodd, W., 1999. Activation of Cortical and Cerebellar Motor Areas during Executed and Imagined Hand Movements: An fMRI Study. J. Cogn. Neurosci. 11, 491–501. https://doi.org/10.1162/089892999563553

Luciano, J. V., Algarabel, S., Tomás, J.M., Martínez, J.L., 2005. Development and validation of the thought control ability questionnaire. Pers. Individ. Dif. 38, 997–1008. https://doi.org/10.1016/j.paid.2004.06.020

Luck, S.J., Gaspelin, N., 2017. How to get statistically significant effects in any ERP experiment (and why you shouldn't). Psychophysiology 54, 146–157. https://doi.org/10.1111/psyp.12639

Luengo-Fernandez, R., Leal, J., Gray, A., 2015. UK research spend in 2008 and 2012: Comparing stroke, cancer, coronary heart disease and dementia. BMJ Open 5, 1–8. https://doi.org/10.1136/bmjopen-2014-006648

Lv, Y., Margulies, D.S., Cameron Craddock, R., Long, X., Winter, B., Gierhake, D., Endres, M., Villringer, K., Fiebach, J., Villringer, A., 2013. Identifying the perfusion deficit in acute stroke with resting-state functional magnetic resonance imaging. Ann. Neurol. 73, 136–139. https://doi.org/10.1002/ana.23763

Macar, F., Coull, J., Vidal, F., 2006. The supplementary motor area in motor and perceptual time processing: fMRI studies. Cogn. Process. 7, 89–94. https://doi.org/10.1007/s10339-005-0025-7

Maciejewski, P.K., Prigerson, H.G., Mazure, C.M., 2000. Self-efficacy as a mediator between stressful life events and depressive symptoms. Br. J. Psychiatry 176, 373–378. https://doi.org/http://dx.doi.org/10.1192/bjp.176.4.373

Maclaren, J., Herbst, M., Speck, O., Zaitsev, M., 2013. Prospective motion correction in brain imaging: A review. Magn. Reson. Med. 69, 621–636. https://doi.org/10.1002/mrm.24314

Magnani, G., Cursi, M., Leocani, L., Volonté, M.A., Comi, G., 2002. Acute effects of L-dopa on event-related desynchronization in Parkinson's disease. Neurol. Sci. 23, 91–97. https://doi.org/10.1007/s100720200033

Magri, C., Schridde, U., Murayama, Y., Panzeri, S., Logothetis, N.K., 2012. The Amplitude and Timing of the BOLD Signal Reflects the Relationship between Local Field Potential Power at Different Frequencies. J. Neurosci. 32, 1395–1407. https://doi.org/10.1523/JNEUROSCI.3985-11.2012

Mak, M.K., Wong-Yu, I.S., Shen, X., Chung, C.L., 2017. Long-term effects of exercise and physical therapy in people with Parkinson disease. Nat. Rev. Neurol. 13, 689–703. https://doi.org/10.1038/nrneurol.2017.128

Maldjian, J.A., Laurienti, P.J., Burdette, J.H., 2004. Precentral gyrus discrepancy in electronic versions of the Talairach atlas. Neuroimage 21, 450–5.

Maldjian, J.A., Laurienti, P.J., Kraft, R.A., Burdette, J.H., 2003. An automated method for neuroanatomic and cytoarchitectonic atlas-based interrogation of fMRI data sets. Neuroimage 19, 1233–9.

Malouin, F., Richards, C.L., Durand, A., 2012. Slowing of motor imagery after a right hemispheric stroke. Stroke Res. Treat. 2012. https://doi.org/10.1155/2012/297217

Malouin, F., Richards, C.L., Durand, A., Doyon, J., 2007. Clinical Assessment of Motor Imagery After Stroke. Neurorehabil. Neural Repair 22, 330–340. https://doi.org/10.1177/1545968307313499

Malouin, F., Richards, C.L., Jackson, P.L., Lafleur, M.F., Durand, A., Doyon, J., 2007. The Kinesthetic and Visual Imagery Questionnaire (KVIQ) for assessing motor imagery in persons with physical disabilities: a reliability and construct validity study. J. Neurol. Phys. Ther. 31, 20–9.

Marchesotti, S., Bassolino, M., Serino, A., Bleuler, H., Blanke, O., 2016. Quantifying the role of motor imagery in brain-machine interfaces. Sci. Rep. 6, 24076. https://doi.org/10.1038/srep24076

Marsden, J.F., Limousin-Dowsey, P., Ashby, P., Pollak, P., Brown, P., 2001. Subthalamic nucleus, sensorimotor cortex and muscle interrelationships in Parkinson's disease. Brain 124, 378–88. https://doi.org/10.1093/brain/124.2.378

Marsili, L., Rizzo, G., Colosimo, C., 2018. Diagnostic criteria for Parkinson's disease: From James Parkinson to the concept of prodromal disease. Front. Neurol. 9, 1–10. https://doi.org/10.3389/fneur.2018.00156

Marx, R., Strauss, C., Williamson, C., 2015. Mindfulness Apprenticeship: A new Model of

NHS-Based MBCT Teacher Training. Mindfulness (N. Y). 6, 253–263. https://doi.org/10.1007/s12671-013-0254-2

Marxen, M., Jacob, M.J., Müller, D.K., Posse, S., Ackley, E., Hellrung, L., Riedel, P., Bender, S., Epple, R., Smolka, M.N., 2016. Amygdala Regulation Following fMRI-Neurofeedback without Instructed Strategies. Front. Hum. Neurosci. 10, 1–14. https://doi.org/10.3389/fnhum.2016.00183

Mayberg, H.S., 2003. Modulating dysfunctional limbic-cortical circuits in depression: Towards development of brain-based algorithms for diagnosis and optimised treatment. Br. Med. Bull. 65, 193–207. https://doi.org/10.1093/bmb/65.1.193

McCulloch, P., Altman, D.G., Campbell, W.B., Flum, D.R., Glasziou, P., Marshall, J.C., Nicholl, J., 2009. No surgical innovation without evaluation: the IDEAL recommendations. Lancet 374, 1105–1112. https://doi.org/10.1016/S0140-6736(09)61116-8

McDonnell, M.N., Stinear, C.M., 2017. TMS measures of motor cortex function after stroke: A meta-analysis. Brain Stimul. 10, 721–734. https://doi.org/10.1016/j.brs.2017.03.008

Mcfarland, D., Mcfarland, D.J., Sarnacki, W.A., Wolpaw, J.R., 2015. Effects of training pre-movement sensorimotor rhythms on behavioral performance Effects of training pre-movement sensorimotor rhythms on behavioral performance 12, 1–21. https://doi.org/10.1088/1741-2560/12/6/066021

Mcglothlin, A.E., Lewis, R.J., 2014. Minimal Clinically Important Difference Defining What Really Matters to Patients. Jama 312, 1342–1343. https://doi.org/10.1001/jama.2014.12660.4

McGrath, C.L., Kelley, M.E., Holtzheimer, P.E., Dunlop, B.W., Craighead, W.E., Franco, A.R., Craddock, R.C., Mayberg, H.S., 2013. Toward a neuroimaging treatment selection biomarker for major depressive disorder. JAMA Psychiatry 70, 821–829. https://doi.org/10.1001/jamapsychiatry.2013.143

McInnes, K., Friesen, C., Boe, S., 2016. Specific brain lesions impair explicit motor imagery ability: A systematic review of the evidence. Arch. Phys. Med. Rehabil. 97, 478–489. https://doi.org/10.1016/j.apmr.2015.07.012

McNair, D.M., 1971. Manual profile of mood states. San Diego: Educational & Industrial testing service.

McNutt, M., 2014. Journals unite for reproducibility. Science 346, 679. https://doi.org/10.1126/SCIENCE.AAA1724

Mehler, D., Williams, A., Krause, F., Luehrs, M., Wise, R., Turner, D., Linden, D., Whittaker,

J., 2017a. Supplementary motor area, but not primary motor cortex - upper limb gradual fMRI neurofeedback, in: Organization of Human Brain Mapping, 23rd Scientific Meeting. Vancouver, British Columbia, Canada.

Mehler, D., Williams, A.N., Whittaker, J.R., Krause, F., Lührs, M., Wise, R., Shetty, H., Turner, D., Linden, D.E.J., 2017b. Gradual real-time fMRI neurofeedback training of motor imagery in middle cerebral artery stroke patientstle. Open Sci. Framew. (preregistration Protoc.

Mehler, D.M.A., Reichenbach, A., Klein, J., Diedrichsen, J., 2017c. Minimizing endpoint variability through reinforcement learning during reaching movements involving shoulder, elbow and wrist. PLoS One 12. https://doi.org/10.1371/journal.pone.0180803

Mehler, D.M.A., Reschechtko, S., 2018. Movement Variability Is Processed Bilaterally by Inferior Parietal Lobule. J. Neurosci. 38, 2413–2415. https://doi.org/10.1523/JNEUROSCI.3224-17.2018

Mehler, D.M.A., Sokunbi, M.O., Habes, I., Barawi, K., Subramanian, L., Range, M., Evans, J., Hood, K., Lührs, M., Keedwell, P., Goebel, R., Linden, D.E.J., 2018. Targeting the affective brain—a randomized controlled trial of real-time fMRI neurofeedback in patients with depression. Neuropsychopharmacology. https://doi.org/10.1038/s41386-018-0126-5

Mehler, D.M.A., Williams, A.N., Krause, F., Lührs, M., Wise, R.G., Turner, D.L., Linden, D.E.J., Whittaker, J.R., 2019. The BOLD response in primary motor cortex and supplementary motor area during kinesthetic motor imagery based graded fMRI neurofeedback. Neuroimage 184, 36–44. https://doi.org/10.1016/j.neuroimage.2018.09.007

Mehrholz, J., Pohl, M., Platz, T., Kugler, J., Elsner, B., 2015. Electromechanical and robot-assisted arm training for improving activities of daily living , arm function , and arm muscle strength after stroke (Review) SUMMARY OF FINDINGS FOR THE MAIN COMPARISON. Cochrane Database Syst. Rev. https://doi.org/10.1002/14651858.CD006876.pub4.www.cochranelibrary.com

Meir-Hasson, Y., Kinreich, S., Podlipsky, I., Hendler, T., Intrator, N., 2014. An EEG Finger-Print of fMRI deep regional activation. Neuroimage 102, 128–141. https://doi.org/10.1016/j.neuroimage.2013.11.004

Meister, R., Jansen, A., Härter, M., Nestoriuc, Y., Kriston, L., 2017. Placebo and nocebo reactions in randomized trials of pharmacological treatments for persistent depressive disorder. A meta-regression analysis. J. Affect. Disord. 215, 288–298.

https://doi.org/10.1016/j.jad.2017.03.024

Mendelsohn, A., Pine, A., Schiller, D., 2014. Between Thoughts and Actions: Motivationally Salient Cues Invigorate Mental Action in the Human Brain. Neuron 81, 207–217. https://doi.org/10.1016/j.neuron.2013.10.019

Mihara, M., Hattori, N., Hatakenaka, M., Yagura, H., Kawano, T., Hino, T., Miyai, I., 2013. Near-infrared spectroscopy-mediated neurofeedback enhances efficacy of motor imagery-based training in poststroke victims: A pilot study. Stroke 44, 1091–1098. https://doi.org/10.1161/STROKEAHA.111.674507

Mikels, J.A., Fredrickson, B.L., Larkin, G.R., Lindberg, C.M., Maglio, S.J., Reuter-Lorenz, P.A., 2005. Emotional category data on images from the international affective picture system. Behav. Res. Methods 37, 626–630. https://doi.org/10.3758/BF03192732

Miller, E.L., Murray, L., Richards, L., Zorowitz, R.D., Bakas, T., Clark, P., Billinger, S.A., 2010. Comprehensive Overview of Nursing and Interdisciplinary Rehabilitation Care of the Stroke Patient. Stroke 41.

Miller, K.J., Schalk, G., Fetz, E.E., den Nijs, M., Ojemann, J.G., Rao, R.P.N., 2010. Cortical activity during motor execution, motor imagery, and imagery-based online feedback. Proc. Natl. Acad. Sci. U. S. A. 107, 4430–5. https://doi.org/10.1073/pnas.0913697107

Minnix, J.A., Blalock, J.A., Marani, S., Prokhorov, A. V., Cinciripini, P.M., 2011. Self-efficacy mediates the effect of depression on smoking susceptibility in adolescents. Nicotine Tob. Res. 13, 699–705. https://doi.org/10.1093/ntr/ntr061

Mirelman, A., Maidan, I., Deutsch, J.E., 2013. Virtual reality and motor imagery: Promising tools for assessment and therapy in Parkinson's disease. Mov. Disord. 28, 1597–1608. https://doi.org/10.1002/mds.25670

Misaki, M., Barzigar, N., Zotev, V., Phillips, R., Cheng, S., Bodurka, J., 2015. Real-time fMRI processing with physiological noise correction – Comparison with off-line analysis. J. Neurosci. Methods 256, 117–121. https://doi.org/10.1016/j.jneumeth.2015.08.033

Miyachi, S., Lu, X., Imanishi, M., Sawada, K., Nambu, A., Takada, M., 2006. Somatotopically arranged inputs from putamen and subthalamic nucleus to primary motor cortex. Neurosci. Res. 56, 300–308. https://doi.org/10.1016/j.neures.2006.07.012

Moher, D., Dulberg, C.S., Wells, G.A., 1994. Statistical power, sample size, and their reporting in randomized controlled trials. J. Am. Med. Assoc. 272, 122–124.

Moher, D., Hopewell, S., Schulz, K.F., Montori, V., Gøtzsche, P.C., Devereaux, P.J., Elbourne, D., Egger, M., Altman, D.G., 2010. CONSORT 2010 explanation and elaboration: updated guidelines for reporting parallel group randomised trials. BMJ 340.

https://doi.org/10.1136/bmj.c869

Moher, D., Schulz, K.F., Altman, D.G., CONSORT, 2001. The CONSORT statement: revised recommendations for improving the quality of reports of parallel group randomized trials. BMC Med. Res. Methodol. 1, 2. https://doi.org/10.1186/1471-2288-1-2

Moisa, M., Polania, R., Grueschow, M., Ruff, C.C., 2016. Brain Network Mechanisms Underlying Motor Enhancement by Transcranial Entrainment of Gamma Oscillations. J. Neurosci. 36, 12053–12065. https://doi.org/10.1523/JNEUROSCI.2044-16.2016

Montgomery, S.A., Asberg, M., 1979. A new depression scale designed to be sensitive to change. Br. J. Psychiatry 134, 382–389. https://doi.org/10.1192/bjp.134.4.382

Moonesinghe, R., Khoury, M.J., Janssens, A.C.J.W., 2007. Most Published Research Findings Are False—But a Little Replication Goes a Long Way. PLoS Med. 4, e28. https://doi.org/10.1371/journal.pmed.0040028

Moran, A., Stein, E., Tischler, H., Bar-Gad, I., 2012. Decoupling neuronal oscillations during subthalamic nucleus stimulation in the parkinsonian primate. Neurobiol. Dis. 45, 583–590. https://doi.org/10.1016/j.nbd.2011.09.016

Morash, V., Bai, O., Furlani, S., Lin, P., Hallett, M., 2008. Classifying EEG Signals Preceding Right Hand, Left Hand, Tonge, and Right Foot Movements and Motor Imageries. Clin. Neurophysiol. 119, 2570–2578. https://doi.org/10.1016/j.clinph.2008.08.013.Classifying

Morey, R.D., Rouder, J.N., 2011. Bayes Factor Approaches for Testing Interval Null Hypotheses. Psychol. Methods 16, 406–419. https://doi.org/10.1037/a0024377

Mosberger, A.C., de Clauser, L., Kasper, H., Schwab, M.E., 2016. Motivational state, reward value, and Pavlovian cues differentially affect skilled forelimb grasping in rats. Learn. Mem. 23, 289–302. https://doi.org/10.1101/lm.039537.115

Mulder, R.T., 2002. Personality Pathology and Treatment Outcome in Major Depression : A Review. Am. J. Psychiatry 159, 359–371.

Müller, V.I., Cieslik, E.C., Serbanescu, I., Laird, A.R., Fox, P.T., Eickhoff, S.B., 2017. Altered brain activity in unipolar depression revisited: Meta-analyses of neuroimaging studies. JAMA Psychiatry 74, 47–55. https://doi.org/10.1001/jamapsychiatry.2016.2783

Mullinger, K.J., Mayhew, S.D., Bagshaw, A.P., Bowtell, R., Francis, S.T., 2014. Evidence that the negative BOLD response is neuronal in origin: A simultaneous EEG-BOLD-CBF study in humans. Neuroimage 94, 263–274. https://doi.org/10.1016/j.neuroimage.2014.02.029

Munafò, M., Neill, J., 2016. Null is beautiful: On the importance of publishing null results. J. Psychopharmacol. 30, 585–585. https://doi.org/10.1177/0269881116638813

Munafò, M., Nosek, B., Bishop, D., Button, K., Chambers, C., Percie du Sert, N., Simonsohn, U., Wagenmakers, E.-J., Ware, J., Ioannidis, J., 2017. A manifesto for reproducible science. Nat. Hum. Behav. 1, 0021. https://doi.org/10.1038/s41562-016-0021

Munzert, J., Lorey, B., Zentgraf, K., 2009. Cognitive motor processes: the role of motor imagery in the study of motor representations. Brain Res. Rev. 60, 306–26. https://doi.org/10.1016/j.brainresrev.2008.12.024

Murphy, K., Birn, R.M., Bandettini, P.A., 2013. Resting-state fMRI confounds and cleanup. Neuroimage 80, 349–59. https://doi.org/10.1016/j.neuroimage.2013.04.001

Murphy, K., Harris, A.D., Wise, R.G., 2011. Robustly measuring vascular reactivity differences with breath-hold: Normalising stimulus-evoked and resting state BOLD fMRI data. Neuroimage 54, 369–379. https://doi.org/10.1016/j.neuroimage.2010.07.059

Muthuraman, M., Deuschl, G., Koirala, N., Riedel, C., Volkmann, J., Groppa, S., 2017. Effects of DBS in parkinsonian patients depend on the structural integrity of frontal cortex. Sci. Rep. 7, 6–11. https://doi.org/10.1038/srep43571

Nambu, A., 2008. Seven problems on the basal ganglia. Curr. Opin. Neurobiol. 18, 595–604. https://doi.org/10.1016/j.conb.2008.11.001

Nambu, A., Tachibana, Y., Chiken, S., 2015. Cause of parkinsonian symptoms: Firing rate, firing pattern or dynamic activity changes? Basal Ganglia 5, 1–6. https://doi.org/10.1016/j.baga.2014.11.001

Nambu, A., Tokuno, H., Takada, M., 2002. Functional significance of the cortico- subthalamo-pallidal ‘ hyperdirect ’ pathway. Neuro Res 43, 111–7.

Nestler, E.J., Barrot, M., Dileone, R.J., Eisch, A.J., Gold, S.J., Monteggia, L.M., 2002. Neurobiology of Depression. Neuron 34, 13–25.

Neyedli, H.F., Sampaio-Baptista, C., Kirkman, M.A., Havard, D., Lührs, M., Ramsden, K., Flitney, D.D., Clare, S., Goebel, R., Johansen-Berg, H., 2017. Increasing lateralized motor activity in younger and older adults using Real-time fMRI during executed movements. Neuroscience 1–10. https://doi.org/10.1016/j.neuroscience.2017.02.010

Nieuwenhuis, S., Forstmann, B.U., Wagenmakers, E.J., 2011. Erroneous analyses of interactions in neuroscience: A problem of significance. Nat. Neurosci. 14, 1105–1107. https://doi.org/10.1038/nn.2886

Ninaus, M., Kober, S.E., Witte, M., Koschutnig, K., Stangl, M., Neuper, C., Wood, G., 2013. Neural substrates of cognitive control under the belief of getting neurofeedback training. Front. Hum. Neurosci. 7, 1–10. https://doi.org/10.3389/fnhum.2013.00914

Nord, C.L., Gray, A., Charpentier, C.J., Robinson, O.J., Roiser, J.P., 2017. Unreliability of

putative fMRI biomarkers during emotional face processing. Neuroimage 156, 119–127. https://doi.org/10.1016/j.neuroimage.2017.05.024

Nord, C.L., Valton, V., Wood, J., Roiser, J.P., 2017. Power-up: a reanalysis of 'power failure' in neuroscience using mixture modelling. J. Neurosci. 37, 3592–16. https://doi.org/10.1523/JNEUROSCI.3592-16.2017

Nosek, B., Spies, J., Motyl, M., 2012. Scientific Utopia. Perspect. Psychol. Sci. 7, 615–631. https://doi.org/10.1177/1745691612459058

Nosek, B.A., Ebersole, C.R., DeHaven, A.C., Mellor, D.T., 2018. The preregistration revolution. Proc. Natl. Acad. Sci. 2017, 201708274. https://doi.org/10.1073/pnas.1708274114

Nosek, B.A., Lakens, D., 2014. Registered reports: A method to increase the credibility of published results. Soc. Psychol. (Gott). 45, 137–141. https://doi.org/10.1027/1864-9335/a000192

Nyenhuis, D.L., Yamamoto, C., Lucheta, T., Terrien, A., Parmentier, A., 1999. Adult and geriatric normative data and validation of the profile of mood states. J. Clin. Psychol. 55, 79–86. https://doi.org/10.1002/(SICI)1097-4679(199901)55:1<79::AID-JCLP8>3.0.CO;2-7

O'Craven, K.M., Kanwisher, N., 2000. Mental Imagery of Faces and Places Activates Corresponding Stimulus-Specific Brain Regions. J. Cogn. Neurosci. 12, 1013–1023. https://doi.org/10.1162/08989290051137549

O'Shea, J., Johansen-Berg, H., Trief, D., Göbel, S., Rushworth, M.F.S., 2007. Functionally Specific Reorganization in Human Premotor Cortex. Neuron 54, 479–490. https://doi.org/10.1016/j.neuron.2007.04.021

Oakes, T.R., Johnstone, T., Ores Walsh, K.S., Greischar, L.L., Alexander, A.L., Fox, A.S., Davidson, R.J., 2005. Comparison of fMRI motion correction software tools. Neuroimage 28, 529–543. https://doi.org/10.1016/j.neuroimage.2005.05.058

Open Science Collaboration, 2015. Estimating the reproducibility of psychological science. Science (80-.). 349, aac4716-aac4716. https://doi.org/10.1126/science.aac4716

Oremus, C., Oremus, M., McNeely, H., Losier, B., Parlar, M., King, M., Hasey, G., Fervaha, G., Graham, A.C., Gregory, C., Hanford, L., Nazarov, A., Restivo, M., Tatham, E., Truong, W., Hall, G.B.C., Lanius, R., McKinnon, M., 2015. Effects of electroconvulsive therapy on cognitive functioning in patients with depression: Protocol for a systematic review and meta-analysis. BMJ Open 5, 1–6. https://doi.org/10.1136/bmjopen-2014-006966

Page, S.J., Levine, P., Leonard, A., Szaflarski, J.P., Kissela, B.M., 2008. Modified Constraint-Induced Therapy in Chronic Stroke: Results of a Single-Blinded Randomized Controlled Trial. Phys. Ther. 88, 333–340. https://doi.org/10.2522/ptj.20060029

Palmer, C., Schmidt, E., McIntosh, J., 1981. CORTICOSPINAL AND CORTICORUBRAL PROJECTIONS FROM THE SUPPLEMENTARY MOTOR AREA IN THE MONKEY. Brain Res. 209, 305–314.

Papoutsi, M., Weiskopf, N., Langbehn, D., Reilmann, R., Rees, G., Tabrizi, S.J., 2018. Stimulating neural plasticity with real-time fMRI neurofeedback in Huntington's disease: A proof of concept study. Hum. Brain Mapp. 39, 1339–1353. https://doi.org/10.1002/hbm.23921

Paret, C., Kluetsch, R., Ruf, M., Demirakca, T., Hoesterey, S., 2014. Down-regulation of amygdala activation with real-time fMRI neurofeedback in a healthy female sample 8, 1–15. https://doi.org/10.3389/fnbeh.2014.00299

Park, C.H., Kou, N., Ward, N.S., 2016. The contribution of lesion location to upper limb deficit after stroke. J. Neurol. Neurosurg. Psychiatry 87, 1283–1286. https://doi.org/10.1136/jnnp-2015-312738

Patil, P., Peng, R.D., Leek, J.T., 2016. Title: A statistical definition for reproducibility and replicability Authors: Prasad Patil, Roger D. Peng, Jeffrey T. Leek. BioRxiV 8–13.

Patten, S.B., 2009. Accumulation of major depressive episodes over time in a prospective study indicates that retrospectively assessed lifetime prevalence estimates are too low. BMC Psychiatry 9, 2–5. https://doi.org/10.1186/1471-244X-9-19

Pattyn, N., Neyt, X., Henderickx, D., Soetens, E., 2008. Psychophysiological investigation of vigilance decrement: Boredom or cognitive fatigue? Physiol. Behav. 93, 369–378. https://doi.org/10.1016/j.physbeh.2007.09.016

Paulus, M.P., Stein, M.B., 2010. Interoception in anxiety and depression. Brain Struct. Funct. 214, 451–463. https://doi.org/10.1007/s00429-010-0258-9

Payne, N., Prudic, J., 2009. Electroconvulsive Therapy Part I: A Perspective on the Evolution and Current Practice of ECT. J Psychiatr Pr. 15, 346–368. https://doi.org/10.1097/01.pra.0000361277.65468.ef.Electroconvulsive

Pearson, J., Naselaris, T., Holmes, E.A., Kosslyn, S.M., 2015. Mental Imagery: Functional Mechanisms and Clinical Applications. Trends Cogn. Sci. 19, 590–602. https://doi.org/10.1016/j.tics.2015.08.003

Peirce, J.W., 2007. PsychoPy-Psychophysics software in Python. J. Neurosci. Methods 162, 8–13. https://doi.org/10.1016/j.jneumeth.2006.11.017

Peng, R.D., 2011. Reproducible Research in Computational Science. Science (80-.). 334, 1226–1227. https://doi.org/10.1126/science.1213847

Perronnet, L., Lécuyer, A., Mano, M., Bannier, E., Lotte, F., Clerc, M., Barillot, C., 2017. Unimodal Versus Bimodal EEG-fMRI Neurofeedback of a Motor Imagery Task. Front. Hum. Neurosci. 11, 193. https://doi.org/10.3389/fnhum.2017.00193

Peters, D.M., Fridriksson, J., Stewart, J.C., Richardson, J.D., Rorden, C., Bonilha, L., Middleton, A., Gleichgerrcht, E., Fritz, S.L., 2018. Cortical disconnection of the ipsilesional primary motor cortex is associated with gait speed and upper extremity motor impairment in chronic left hemispheric stroke. Hum. Brain Mapp. 39, 120–132. https://doi.org/10.1002/hbm.23829

Pettersson, A., Modin, S., Wahlström, R., Af Winklerfelt Hammarberg, S., Krakau, I., 2018. The Mini-International Neuropsychiatric Interview is useful and well accepted as part of the clinical assessment for depression and anxiety in primary care: A mixed-methods study. BMC Fam. Pract. 19, 1–13. https://doi.org/10.1186/s12875-017-0674-5

Pfurtscheller, G., Aranibar, A., 1979. Evaluation of event-related desynchronization (ERD) preceding and following voluntary self-paced movement. Electroencephalogr. Clin. Neurophysiol. 46, 138–46.

Pfurtscheller, G., Neuper, C., 1997. Motor imagery activates primary sensorimotor area in humans. Neurosci. Lett. 239, 65–68. https://doi.org/10.1016/S0304-3940(97)00889-6

Pfurtscheller, G., Neuper, C., Brunner, C., da Silva, F.L., 2005. Beta rebound after different types of motor imagery in man. Neurosci. Lett. 378, 156–9. https://doi.org/10.1016/j.neulet.2004.12.034

Phillips, M.L., Ladouceur, C.D., Drevets, W.C., 2008. A neural model of voluntary and automatic emotion regulation: implications for understanding the pathophysiology and neurodevelopment of bipolar disorder. Mol. Psychiatry 13, 829, 833–57. https://doi.org/10.1038/mp.2008.65

Picard, N., Strick, P.L., 2001. Imaging the premotor areas. Curr. Opin. Neurobiol. 11, 663–72.

Picciotto, M., 2018. Analytical Transparency and Reproducibility in Human Neuroimaging Studies. J. Neurosci. 38, 3375–3376. https://doi.org/10.1523/JNEUROSCI.0424-18.2018

Plow, E.B., Cunningham, D.A., Varnerin, N., Machado, A., 2015. Rethinking stimulation of the brain in stroke rehabilitation: why higher motor areas might be better alternatives for patients with greater impairments. Neuroscientist 21, 225–40. https://doi.org/10.1177/1073858414537381

Pocock, S.J., Hughes, M.D., Lee, R.J., 1987. Statistical Problems in the Reporting of Clinical

Trials. N. Engl. J. Med. 317, 426–432. https://doi.org/10.1056/NEJM198708133170706

Pogosyan, A., Gaynor, L.D., Eusebio, A., Brown, P., 2009. Boosting Cortical Activity at Beta-Band Frequencies Slows Movement in Humans. Curr. Biol. 19, 1637–1641. https://doi.org/10.1016/j.cub.2009.07.074

Pogosyan, A., Yoshida, F., Chen, C.C., Martinez-Torres, I., Foltynie, T., Limousin, P., Zrinzo, L., Hariz, M.I., Brown, P., 2010. Parkinsonian impairment correlates with spatially extensive subthalamic oscillatory synchronization. Neuroscience 171, 245–257. https://doi.org/10.1016/j.neuroscience.2010.08.068

Poldrack, R.A., Baker, C.I., Durnez, J., Gorgolewski, K.J., Matthews, P.M., Munafò, M.R., Nichols, T.E., Poline, J.-B., Vul, E., Yarkoni, T., 2017. Scanning the horizon: towards transparent and reproducible neuroimaging research. Nat. Rev. Neurosci. 18, 115–126. https://doi.org/10.1038/nrn.2016.167

Pollock, A., Baer, G., Campbell, P., Choo, P.L., Forster, A., Morris, J., Pomeroy, V.M., Langhorne, P., 2014. Physical rehabilitation approaches for the recovery of function and mobility following stroke, in: Pollock, A. (Ed.), Cochrane Database of Systematic Reviews. John Wiley & Sons, Ltd, Chichester, UK, p. CD001920. https://doi.org/10.1002/14651858.CD001920.pub3

Pollok, B., Krause, V., Martsch, W., Wach, C., Schnitzler, A., Südmeyer, M., 2012. Motor-cortical oscillations in early stages of Parkinson's disease. J. Physiol. 590, 3203–3212. https://doi.org/10.1113/jphysiol.2012.231316

Popper, K.R., 2005. The Logic of Scientific Discovery, 2nd ed. Routledge. https://doi.org/10.1016/S0016-0032(59)90407-7

Power, J.D., Barnesa, K.A., Snydera, Abraham Z Schlaggara, B.L., Petersena, S.E., 2012. Spurious but systematic correlations in functional connectivity MRI networks arise from subject motion. Neuroimage 59, 2142–2154. https://doi.org/10.1016/S0140-6736(02)11602-3.Association

Power, J.D., Mitra, A., Laumann, T.O., Snyder, A.Z., Schlaggar, B.L., Petersen, S.E., 2014. Methods to detect, characterize, and remove motion artifact in resting state fMRI. Neuroimage 84, 320–41. https://doi.org/10.1016/j.neuroimage.2013.08.048

Price, D.D., Finniss, D.G., Benedetti, F., 2008. A Comprehensive Review of the Placebo Effect: Recent Advances and Current Thought. Annu. Rev. Psychol. 59, 565–590. https://doi.org/10.1146/annurev.psych.59.113006.095941

Pruszynski, J.A., Diedrichsen, J., 2015. Reading the mind to move the body. Science (80-.). 348, 860–861. https://doi.org/10.1126/science.aab3464

Ptak, R., 2012. The frontoparietal attention network of the human brain: Action, saliency, and a priority map of the environment. Neuroscientist 18, 502–515. https://doi.org/10.1177/1073858411409051

Qids, S., Qids-c, C.R., Rush, a J., Trivedi, M.H., Ibrahim, H.M., Carmody, T.J., Arnow, B., Klein, D.N., Markowitz, J.C., Ninan, P.T., Kornstein, S., Manber, R., Thase, M.E., Kocsis, J.H., Keller, M.B., 2003. The 16-Item Quick Inventory of Depressive. Depression 54, 573–583. https://doi.org/10.1016/S0006-3223(03)01866-8

Quintana, D.S., Donald, R.W., 2017. Bayesian alternatives for common null-hypothesis significance tests in psychiatry: A non-technical guide using JASP. Open Sci. Framew. 1–10. https://doi.org/10.17605/OSF.IO/WUN5V

Ramot, A.M., Gonzalez-castillo, J., 2018. A framework for offline evaluation and optimization of real-time algorithms for use in neurofeedback , demonstrated on an instantaneous proxy for correlations. bioRxiv 1–33. https://doi.org/10.1101/351072

Ramot, M., Kimmich, S., Gonzalez-Castillo, J., Roopchansingh, V., Popal, H., White, E., Gotts, S.J., Martin, A., 2017. Direct modulation of aberrant brain network connectivity through real - time NeuroFeedback. Elife 6:e28974, 1–23. https://doi.org/10.1101/139824

Ramsey, S., Scoggins, J., 2008. Practicing on the Tip of an Information Iceberg? Evidence of Underpublication of Registered Clinical Trials in Oncology. Oncologist 13, 925. https://doi.org/10.1634/theoncologist.2008-0133

Rance, M., Walsh, C., Sukhodolsky, D.G., Pittman, B., Qiu, M., Kichuk, S.A., Wasylink, S., Koller, W.N., Bloch, M., Gruner, P., Scheinost, D., Pittenger, C., Hampson, M., 2018. Time course of clinical change following neurofeedback. Neuroimage. https://doi.org/10.1016/j.neuroimage.2018.05.001

Randell, E., McNamara, R., Subramanian, L., Hood, K., Linden, D., 2018. Current practices in clinical neurofeedback with functional MRI-Analysis of a survey using the TIDieR checklist. Eur. Psychiatry 2–7. https://doi.org/10.1016/j.eurpsy.2017.10.011

Rasmussen, K.G., 2009. Sham Electroconvulsive Therapy Studies in Depressive Illness. J. ECT 25, 54–59. https://doi.org/10.1097/YCT.0b013e3181719b23

Read, J., Arnold, C., 2017. Is Electroconvulsive Therapy for Depression More Effective Than Placebo? A Systematic Review of Studies Since 2009. Ethical Hum. Psychol. Psychiatry 19, 5–23.

Rehme, A.K., Eickhoff, S.B., Rottschy, C., Fink, G.R., Grefkes, C., 2012. Activation likelihood estimation meta-analysis of motor-related neural activity after stroke. Neuroimage 59, 2771–2782. https://doi.org/10.1016/j.neuroimage.2011.10.023

Rehme, A.K., Eickhoff, S.B., Wang, L.E., Fink, G.R., Grefkes, C., 2011a. Dynamic causal modeling of cortical activity from the acute to the chronic stage after stroke. Neuroimage 55, 1147–1158. https://doi.org/10.1016/j.neuroimage.2011.01.014

Rehme, A.K., Fink, G.R., Von Cramon, D.Y., Grefkes, C., 2011b. The role of the contralesional motor cortex for motor recovery in the early days after stroke assessed with longitudinal fMRI. Cereb. Cortex 21, 756–768. https://doi.org/10.1093/cercor/bhq140

Rehme, A.K., Grefkes, C., 2013. Cerebral network disorders after stroke: Evidence from imaging-based connectivity analyses of active and resting brain states in humans. J. Physiol. 591, 17–31. https://doi.org/10.1113/jphysiol.2012.243469

Riecker, A., Gröschel, K., Ackermann, H., Schnaudigel, S., Kassubek, J., Kastrup, A., 2010. The role of the unaffected hemisphere in motor recovery after stroke. Hum. Brain Mapp. 31, 1017–1029. https://doi.org/10.1002/hbm.20914

Riley, J.D., Le, V., Der-Yeghiaian, L., See, J., Newton, J.M., Ward, N.S., Cramer, S.C., 2011. Anatomy of stroke injury predicts gains from therapy. Stroke 42, 421–426. https://doi.org/10.1161/STROKEAHA.110.599340

Ring, C., Cooke, A., Kavussanu, M., McIntyre, D., Masters, R., 2015. Investigating the efficacy of neurofeedback training for expediting expertise and excellence in sport. Psychol. Sport Exerc. 16, 118–127. https://doi.org/10.1016/j.psychsport.2014.08.005

Riva-Posse, P., Choi, K.S., Holtzheimer, P.E., Crowell, A.L., Garlow, S.J., Rajendra, J.K., McIntyre, C.C., Gross, R.E., Mayberg, H.S., 2017. A connectomic approach for subcallosal cingulate deep brain stimulation surgery: prospective targeting in treatment-resistant depression. Mol. Psychiatry 1–7. https://doi.org/10.1038/mp.2017.59

Robineau, F., Saj, A., Neveu, R., Ville, D. Van De, Vuilleumier, P., 2017. Using real-time fMRI neurofeedback to restore right occipital cortex activity in patients with left visuo-spatial neglect : proof-of-principle and preliminary results. Neuropsychol. Rehabil. 0, 1–22. https://doi.org/10.1080/09602011.2017.1301262

Rodriguez-Oroz, M.C., Jahanshahi, M., Krack, P., Litvan, I., Macias, R., Bezard, E., Obeso, J.A., 2009. Initial clinical manifestations of Parkinson's disease: features and pathophysiological mechanisms. Lancet Neurol. 8, 1128–1139. https://doi.org/10.1016/S1474-4422(09)70293-5

Rondina, J.M., Park, C.H., Ward, N.S., 2017. Brain regions important for recovery after severe post-stroke upper limb paresis. J. Neurol. Neurosurg. Psychiatry 88, 737–743. https://doi.org/10.1136/jnnp-2016-315030

Rosa, M.A., Gattaz, W.F., Pascual-Leone, A., Fregni, F., Rosa, M.O., Rumi, D.O.,

Myczkowski, M., Silva, M.F., Mansur, C., Rigonatti, S.P., Teixeira, M.J., Marcolin, M.A., 2006. Comparison of repetitive transcranial magnetic stimulation and electroconvulsive therapy in unipolar non-psychotic refractory depression: A randomized, single-blind study. Int. J. Neuropsychopharmacol. 9, 667–676. https://doi.org/10.1017/S1461145706007127

Rosenthal, R., 1979. The file drawer problem and tolerance for null results. Psychol. Bull. 86, 638–641. https://doi.org/10.1037/0033-2909.86.3.638

Rouder, J.N., Morey, R.D., Verhagen, J., Province, J.M., Wagenmakers, E.J., 2016. Is There a Free Lunch in Inference? Top. Cogn. Sci. 8, 520–547. https://doi.org/10.1111/tops.12214

Rouder, J.N., Speckman, P.L., Sun, D., Morey, R.D., Iverson, G., 2009. Bayesian t tests for accepting and rejecting the null hypothesis. Psychon. Bull. Rev. 16, 225–237. https://doi.org/10.3758/PBR.16.2.225

Rousseeuw, P.J., Leroy, A.M., 1987. Robust Regression and Outlier Detection. NY Wiley, New York.

Rowe, J.B., Chan, V., Ingemanson, M.L., Cramer, S.C., Wolbrecht, E.T., Reinkensmeyer, D.J., 2017. Robotic Assistance for Training Finger Movement Using a Hebbian Model: A Randomized Controlled Trial. Neurorehabil. Neural Repair 31, 769–780. https://doi.org/10.1177/1545968317721975

Rüber, T., Schlaug, G., Lindenberg, R., 2012. Compensatory role of the cortico-rubro-spinal tract in motor recovery after stroke. Neurology 79, 515–22. https://doi.org/10.1212/WNL.0b013e31826356e8

Rush, A.J., Trivedi, M.H., Wisniewski, S.R., Nierenberg, A.A., Stewart, J.W., Warden, D., Niederehe, G., Thase, M.E., Lavori, P.W., Lebowitz, B.D., McGrath, P.J., Rosenbaum, J.F., Sackeim, H.A., Kupfer, D.J., Luther, J., Fava, M., 2006. Acute and longer-term outcomes in depressed outpatients requiring one or several treatment steps: a STAR*D report. Am. J. Psychiatry 163, 1905–17. https://doi.org/10.1176/ajp.2006.163.11.1905

Sacchet, M.D., Gotlib, I.H., 2017. Myelination of the brain in major depressive disorder: An in vivo quantitative magnetic resonance imaging study. Sci. Rep. 7, 1–14. https://doi.org/10.1038/s41598-017-02062-y

Sacchet, M.D., Gotlib, I.H., 2016. Neurofeedback training for major depressive disorder: recent developments and future directions. Expert Rev. Neurother. 1–3. https://doi.org/10.1080/14737175.2016.1199959

Sakurada, T., Nakajima, T., Morita, M., Hirai, M., Watanabe, E., 2017. Improved motor performance in patients with acute stroke using the optimal individual attentional strategy.

Sci. Rep. 7, 40592. https://doi.org/10.1038/srep40592

Sampaio-baptista, C., Sanders, Z., Johansen-berg, H., 2018. Annual Review of Neuroscience Structural Plasticity in Adulthood with Motor Learning and Stroke Rehabilitation. Annu. Rev. Neurosci. 25–40.

Sampaio-Baptista, C., Scholz, J., Jenkinson, M., Thomas, A.G., Filippini, N., Smit, G., Douaud, G., Johansen-Berg, H., 2014. Gray matter volume is associated with rate of subsequent skill learning after a long term training intervention. Neuroimage 96, 158–166. https://doi.org/10.1016/j.neuroimage.2014.03.056

Santaniello, S., McCarthy, M.M., Montgomery, E.B., Gale, J.T., Kopell, N., Sarma, S. V., 2015. Therapeutic mechanisms of high-frequency stimulation in Parkinson's disease and neural restoration via loop-based reinforcement. Proc. Natl. Acad. Sci. 112, E586–E595. https://doi.org/10.1073/pnas.1406549111

Sato, J.R., Basilio, R., Paiva, F.F., Garrido, G.J., Bramati, I.E., Bado, P., Tovar-Moll, F., Zahn, R., Moll, J., 2013. Real-time fMRI pattern decoding and neurofeedback using FRIEND: An FSL-integrated BCI toolbox. PLoS One 8, 1–13. https://doi.org/10.1371/journal.pone.0081658

Saver, J.L., Goyal, M., Van Der Lugt, A., Menon, B.K., Majoie, C.B.L.M., Dippel, D.W., Campbell, B.C., Nogueira, R.G., Demchuk, A.M., Tomasello, A., Cardona, P., Devlin, T.G., Frei, D.F., Du Mesnil De Rochemont, R., Berkhemer, O.A., Jovin, T.G., Siddiqui, A.H., Van Zwam, W.H., Davis, S.M., Castaño, C., Sapkota, B.L., Fransen, P.S., Molina, C., Van Oostenbrugge, R.J., Chamorro, Á., Lingsma, H., Silver, F.L., Donnan, G.A., Shuaib, A., Brown, S., Stouch, B., Mitchell, P.J., Davalos, A., Roos, Y.B.W.E.M., Hill, M.D., 2016. Time to treatment with endovascular thrombectomy and outcomes from ischemic stroke: Ameta-analysis. JAMA - J. Am. Med. Assoc. 316, 1279–1288. https://doi.org/10.1001/jama.2016.13647

Schaechter, J.D., 2004. Motor rehabilitation and brain plasticity after hemiparetic stroke. Prog. Neurobiol. 73, 61–72. https://doi.org/10.1016/j.pneurobio.2004.04.001

Schaechter, J.D., Perdue, K.L., Wang, R., 2008. Structural damage to the corticospinal tract correlates with bilateral sensorimotor cortex reorganization in stroke patients. Neuroimage 39, 1370–1382. https://doi.org/10.1016/j.neuroimage.2007.09.071

Scheinost, D., Stoica, T., Saksa, J., Papademetris, X., Constable, R.T., Pittenger, C., Hampson, M., 2013. Orbitofrontal cortex neurofeedback produces lasting changes in contamination anxiety and resting-state connectivity. Transl. Psychiatry 3, e250-6. https://doi.org/10.1038/tp.2013.24

Schenkman, M., Moore, C.G., Kohrt, W.M., Hall, D.A., Delitto, A., Comella, C.L., Josbeno, D.A., Christiansen, C.L., Berman, B.D., Kluger, B.M., Melanson, E.L., Jain, S., Robichaud, J.A., Poon, C., Corcos, D.M., 2017. Effect of High-Intensity Treadmill Exercise on Motor Symptoms in Patients With De Novo Parkinson Disease. JAMA Neurol. 80045, 219–226. https://doi.org/10.1001/jamaneurol.2017.3517

Schnitzler, A., Salenius, S., Salmelin, R., Jousmäki, V., Hari, R., 1997. Involvement of Primary Motor Cortex in Motor Imagery: A Neuromagnetic Study. Neuroimage 6, 201–208. https://doi.org/10.1006/nimg.1997.0286

Schönbrodt, F.D., Wagenmakers, E.-J., 2016. Bayes Factor Design Analysis: Planning for Compelling Evidence. SSRN Electron. J. 1–13. https://doi.org/10.2139/ssrn.2722435

Schönbrodt, F.D., Wagenmakers, E.J., Zehetleitner, M., Perugini, M., 2015. Sequential hypothesis testing with Bayes acfactors: Efficiently testing mean differences. Psychol. Methods 22, 322–339. https://doi.org/10.1037/met0000061

Schotte, C.K.W., Maes, M., Cluydts, R., De Doncker, D., Cosyns, P., 1997. Construct validity of the Beck Depression Inventory in a depressive population. J. Affect. Disord. 46, 115–125. https://doi.org/10.1016/S0165-0327(97)00094-3

Schroeter, M.L., Cutini, S., Wahl, M.M., Scheid, R., Yves von Cramon, D., 2007. Neurovascular coupling is impaired in cerebral microangiopathy-An event-related Stroop study. Neuroimage 34, 26–34. https://doi.org/10.1016/j.neuroimage.2006.09.001

Schuirmann, D., 1987. A comparison of the two-one side test procedures and the power approach for assessing the equivalence of average bioavailability. J Pharmacokinet Biopharm 15, 657–681.

Schulz, K.F., Chalmers, I., Grimes, D.A., Altman, D.G., 1994. Assessing the quality of randomization from reports of controlled trials published in obstetrics and gynecology journals. JAMA 272, 125–8.

Schulz, R., Braass, H., Liuzzi, G., Hoerniss, V., Lechner, P., Gerloff, C., Hummel, F.C., 2015. White matter integrity of premotor-motor connections is associated with motor output in chronic stroke patients. NeuroImage Clin. 7, 82–86. https://doi.org/10.1016/j.nicl.2014.11.006

Schulz, R., Park, C.H., Boudrias, M.H., Gerloff, C., Hummel, F.C., Ward, N.S., 2012. Assessing the integrity of corticospinal pathways from primary and secondary cortical motor areas after stroke. Stroke 43, 2248–2251. https://doi.org/10.1161/STROKEAHA.112.662619

Schulz, R., Park, E., Lee, J., Chang, W.H., Lee, A., Kim, Y.H., Hummel, F.C., 2017.

Interactions Between the Corticospinal Tract and Premotor-Motor Pathways for Residual Motor Output After Stroke. Stroke 48, 2805–2811. https://doi.org/10.1161/STROKEAHA.117.016834

Scott, J., Palmer, S., Paykel, E., Teasdale, J., Hayhurst, H., 2003. Use of cognitive therapy for relapse prevention in chronic depression: Cost-effectiveness study. Br. J. Psychiatry 182, 221–227. https://doi.org/10.1192/bjp.182.3.221

Semkovska, M., McLoughlin, D.M., 2010. Objective cognitive performance associated with electroconvulsive therapy for depression: A systematic review and meta-analysis. Biol. Psychiatry 68, 568–577. https://doi.org/10.1016/j.biopsych.2010.06.009

Sepulveda, P., Sitaram, R., Rana, M., Montalba, C., Tejos, C., Ruiz, S., 2016. How feedback, motor imagery, and reward influence brain self-regulation using real-time fMRI. Hum. Brain Mapp. 37, 3153–3171. https://doi.org/10.1002/hbm.23228

Seto, E., Sela, G., McIlroy, W.E., Black, S.E., Staines, W.R., Bronskill, M.J., McIntosh, A.R., Graham, S.J., 2001. Quantifying head motion associated with motor tasks used in fMRI. Neuroimage 14, 284–297. https://doi.org/10.1006/nimg.2001.0829

Sharma, N., Baron, J.-C., 2013. Does motor imagery share neural networks with executed movement: a multivariate fMRI analysis. Front. Hum. Neurosci. 7, 564. https://doi.org/10.3389/fnhum.2013.00564

Sharma, N., Baron, J.C., Rowe, J.B., 2009a. Motor imagery after stroke: Relating outcome to motor network connectivity. Ann. Neurol. 66, 604–616. https://doi.org/10.1002/ana.21810

Sharma, N., Jones, P.S., Carpenter, T.A., Baron, J.-C., 2008. Mapping the involvement of BA 4a and 4p during Motor Imagery. Neuroimage 41, 92–9. https://doi.org/10.1016/j.neuroimage.2008.02.009

Sharma, N., Pomeroy, V.M., Baron, J.C., 2006. Motor imagery: A backdoor to the motor system after stroke? Stroke 37, 1941–1952. https://doi.org/10.1161/01.STR.0000226902.43357.fc

Sharma, N., Simmons, L.H., Jones, P.S., Day, D.J., Carpenter, T.A., Pomeroy, V.M., Warburton, E.A., Baron, J.C., 2009b. Motor imagery after subcortical stroke: A functional magnetic resonance imaging study. Stroke 40, 1315–1324. https://doi.org/10.1161/STROKEAHA.108.525766

Sharp, K., Hewitt, J., 2014. Dance as an intervention for people with Parkinson's disease: A systematic review and meta-analysis. Neurosci. Biobehav. Rev. 47, 445–456. https://doi.org/10.1016/j.neubiorev.2014.09.009

Shen, X., Wong-Yu, I.S.K., Mak, M.K.Y., 2016. Effects of Exercise on Falls, Balance, and Gait Ability in Parkinson's Disease. Neurorehabil. Neural Repair 30, 512–527. https://doi.org/10.1177/1545968315613447

Sherer, M., Maddux, J.E., Mercandante, B., Prentice-Dunn, S., Jacobs, B., Rogers, R.W., 1982. The Self-Efficacy Scale: Construction and Validation. Psychol. Rep. 51, 663–671. https://doi.org/10.2466/pr0.1982.51.2.663

Shibata, K., Watanabe, T., Sasaki, Y., Kawato, M., 2011. Perceptual Learning Incepted by Decoded fMRI Neurofeedback Without Stimulus Presentation. Science (80-.). 334, 1413–1415. https://doi.org/10.1126/science.1212003

Shmuel, A., 2010. Locally Measured Neuronal Correlates of Functional MRI Signals, in: Mulert, C., Lemieux, L. (Eds.), EEG - FMRI. Physiological Basis, Technique, and Applications. Springer-Verlag, Berlin Heidelberg, pp. 63–82.

Shuster, L.I., Lemieux, S.K., 2005. An fMRI investigation of covertly and overtly produced mono- and multisyllabic words. Brain Lang. 93, 20–31. https://doi.org/10.1016/j.bandl.2004.07.007

Siegel, J.S., Shulman, G.L., Corbetta, M., 2017. Measuring functional connectivity in stroke: Approaches and considerations. J. Cereb. Blood Flow Metab. 37, 2665–2678. https://doi.org/10.1177/0271678X17709198

Simmons, J.P., Nelson, L.D., Simonsohn, U., 2011. False-positive psychology: Undisclosed flexibility in data collection and analysis allwos presenting anything as significant. Psychol. Sci. 22, 1359–1366. https://doi.org/10.1177/0956797611417632

Simmons, L., Sharma, N., Baron, J.C., Pomeroy, V.M., 2008. Motor imagery to enhance recovery after subcortical stroke: Who might benefit, daily dose, and potential effects. Neurorehabil. Neural Repair 22, 458–467. https://doi.org/10.1177/1545968308315597

Sitaram, R., Ros, T., Stoeckel, L., Haller, S., Scharnowski, F., Lewis-Peacock, J., Weiskopf, N., Blefari, M.L., Rana, M., Oblak, E., Birbaumer, N., Sulzer, J., 2017. Closed-loop brain training: The science of neurofeedback. Nat. Rev. Neurosci. 18, 86–100. https://doi.org/10.1038/nrn.2016.164

Sitaram, R., Veit, R., Stevens, B., Caria, A., Gerloff, C., Birbaumer, N., Hummel, F., 2012. Acquired control of ventral premotor cortex activity by feedback training: an exploratory real-time FMRI and TMS study. Neurorehabil. Neural Repair 26, 256–65. https://doi.org/10.1177/1545968311418345

Smith, A.. F.. M.., Spiegelhalter, D.. J.., 1980. Bayes Factors and Choice Criteria for Linear Models. J. R. Stat. Soc. Ser. B 42, 213–220.

Smith, S.M., Nichols, T.E., 2009. Threshold-free cluster enhancement: Addressing problems of smoothing, threshold dependence and localisation in cluster inference. Neuroimage 44, 83–98. https://doi.org/10.1016/j.neuroimage.2008.03.061

Solodkin, A., Hlustik, P., Chen, E.E., Small, S.L., 2004. Fine Modulation in Network Activation during Motor Execution and Motor Imagery. Cereb. Cortex 14, 1246–1255. https://doi.org/10.1093/cercor/bhh086

Sonuga-Barke, E.J.S., Brandeis, D., Cortese, S., Daley, D., Ferrin, M., Holtmann, M., Stevenson, J., Danckaerts, M., Van Der Oord, S., Döpfner, M., Dittmann, R.W., Simonoff, E., Zuddas, A., Banaschewski, T., Buitelaar, J., Coghill, D., Hollis, C., Konofal, E., Lecendreux, M., Wong, I.C.K., Sergeant, J., 2013. Nonpharmacological interventions for ADHD: Systematic review and meta-analyses of randomized controlled trials of dietary and psychological treatments. Am. J. Psychiatry 170, 275–289. https://doi.org/10.1176/appi.ajp.2012.12070991

Sorger, B., Kamp, T., Weiskopf, N., Peters, J.C., Goebel, R., 2016. When the brain takes 'BOLD' steps: Real-time fMRI neurofeedback can further enhance the ability to gradually self-regulate regional brain activation. Neuroscience. https://doi.org/10.1016/j.neuroscience.2016.09.026

Steward, O., Balice-Gordon, R., 2014. Rigor or mortis: Best practices for preclinical research in neuroscience. Neuron 84, 572–581. https://doi.org/10.1016/j.neuron.2014.10.042

Stinear, C., 2010. Prediction of recovery of motor function after stroke. Lancet Neurol. 9, 1228–1232. https://doi.org/10.1016/S1474-4422(10)70247-7

Stinear, C.M., Byblow, W.D., Steyvers, M., Levin, O., Swinnen, S.P., 2006. Kinesthetic, but not visual, motor imagery modulates corticomotor excitability. Exp. Brain Res. 168, 157–164. https://doi.org/10.1007/s00221-005-0078-y

Stinear, C.M., Fleming, M.K., Barber, P.A., Byblow, W.D., 2007. Lateralization of motor imagery following stroke. Clin. Neurophysiol. 118, 1794–1801. https://doi.org/10.1016/j.clinph.2007.05.008

Stoeckel, L., Garrison, K., Ghosh, S., Wighton, P., 2014. Optimizing real time fMRI neurofeedback for therapeutic discovery and development. NeuroImage Clin.

Stoffers, D., Bosboom, J.L.W., Deijen, J.B., Wolters, E.C., Stam, C.J., Berendse, H.W., 2008. Increased cortico-cortical functional connectivity in early-stage Parkinson's disease: An MEG study. Neuroimage 41, 212–222. https://doi.org/10.1016/j.neuroimage.2008.02.027

Strauss, C., Cavanagh, K., Oliver, A., Pettman, D., 2014. Mindfulness-based interventions for people diagnosed with a current episode of an anxiety or depressive disorder: A meta-

analysis of randomised controlled trials. PLoS One 9. https://doi.org/10.1371/journal.pone.0096110

Subramanian, L., Hindle, J. V, Johnston, S., Roberts, M. V, Husain, M., Goebel, R., Linden, D., 2011. Real-time functional magnetic resonance imaging neurofeedback for treatment of Parkinson's disease. J. Neurosci. 31, 16309–17. https://doi.org/10.1523/JNEUROSCI.3498-11.2011

Subramanian, L., Morris, M.B., Brosnan, M., Turner, D.L., Morris, H.R., Linden, D.E.J., 2016. Functional Magnetic Resonance Imaging Neurofeedback-guided Motor Imagery Training and Motor Training for Parkinson's Disease: Randomized Trial. Front. Behav. Neurosci. 10, 111. https://doi.org/10.3389/fnbeh.2016.00111

Sugarman, M.A., Loree, A.M., Baltes, B.B., Grekin, E.R., Kirsch, I., 2014. The Efficacy of Paroxetine and Placebo in Treating Anxiety and Depression: A Meta-Analysis of Change on the Hamilton Rating Scales. PLoS One 9, e106337. https://doi.org/10.1371/journal.pone.0106337

Sulzer, J., Haller, S., Scharnowski, F., Weiskopf, N., Birbaumer, N., Bruehl, A.B., Cohen, L.G., Gassert, R., Goebel, R., Herwig, U., Linden, D., Luft, A., Seifritz, E., Sitaram, R., Camillo, O.S., 2013. Real-time fMRI neurofeedback: Progress and challenges. Neuroimage 76, 386–399.

Swann, N.C., de Hemptinne, C., Miocinovic, S., Qasim, S., Wang, S.S., Ziman, N., Ostrem, J.L., San Luciano, M., Galifianakis, N.B., Starr, P.A., 2016. Gamma Oscillations in the Hyperkinetic State Detected with Chronic Human Brain Recordings in Parkinson's Disease. J. Neurosci. 36, 6445–6458. https://doi.org/10.1523/JNEUROSCI.1128-16.2016

Szucs, D., Ioannidis, J.P.A., 2017. Empirical assessment of published effect sizes and power in the recent cognitive neuroscience and psychology literature. PLOS Biol. 15, e2000797. https://doi.org/10.1371/journal.pbio.2000797

Taichman, D.B., Sahni, P., Pinborg, A., Peiperl, L., Laine, C., James, A., Hong, S.-T., Haileamlak, A., Gollogly, L., Godlee, F., Frizelle, F.A., Florenzano, F., Drazen, J.M., Bauchner, H., Baethge, C., Backus, J., 2017. Data Sharing Statements for Clinical Trials — A Requirement of the International Committee of Medical Journal Editors. N. Engl. J. Med. 376, 2277–2279. https://doi.org/10.1056/NEJMe1705439

Takenobu, Y., Hayashi, T., Moriwaki, H., Nagatsuka, K., Naritomi, H., Fukuyama, H., 2014. Motor recovery and microstructural change in rubro-spinal tract in subcortical stroke. NeuroImage Clin. 4, 201–208. https://doi.org/10.1016/j.nicl.2013.12.003

Talairach, J., Tournoux, P., 1988. Co-Planar Stereotaxic Atlas of the Human Brain. New York:

Thieme Medical Publishers.

Tamir, R., Dickstein, R., Huberman, M., 2007. Integration of motor imagery and physical practice in group treatment applied to subjects with Parkinson's disease. Neurorehabil. Neural Repair 21, 68–75. https://doi.org/10.1177/1545968306292608

Tan, H., Wade, C., Brown, P., 2016. Post-Movement Beta Activity in Sensorimotor Cortex Indexes Confidence in the Estimations from Internal Models. J. Neurosci. 36.

Team, J., 2017. JASP (Version 0.8.5.1)[Computer software].

The TADS Team, 2007. The Treatment for Adolescents With Depression Study (TADS). Effectiveness, Long-term Outcomes, Safety. Arch Gen Psychiatry 64, 1132–1144.

The Registered Reports Steering Committee, 2018. Registered Reports: Peer review before results are known to align scientific values and practices. [WWW Document]. URL https://cos.io/rr/?_ga=2.221308700.548657228.1527610721-860017422.1494689112 (accessed 5.29.18).

The Registered Reports Steering Committee, n.d. Published Preregistered Studies. URL: https://www.zotero.org/groups/479248/osf/items/collectionKey/KEJP68G9 (accessed 5.15.17).

Thibault, R.T., Lifshitz, M., Raz, A., 2018a. The climate of neurofeedback: Scientific rigour and the perils of ideology. Brain 141, e11. https://doi.org/10.1093/brain/awx330

Thibault, R.T., Lifshitz, M., Raz, A., 2017. Neurofeedback or neuroplacebo? Brain 140, 862–864. https://doi.org/10.1093/brain/awx011

Thibault, R.T., Lifshitz, M., Raz, A., 2016. The self-regulating brain and neurofeedback: Experimental science and clinical promise. Cortex 74, 247–261. https://doi.org/10.1016/j.cortex.2015.10.024

Thibault, R.T., MacPherson, A., Lifshitz, M., Roth, R.R., Raz, A., 2018b. Neurofeedback with fMRI: A critical systematic review. Neuroimage 172, 786–807. https://doi.org/10.1016/j.neuroimage.2017.12.071

Thibault, R.T., MacPherson, A., Lifshitz, M., Roth, R.R., Raz, A., 2017. Neurofeedback or neuroplacebo? Neuroimage. https://doi.org/10.1016/j.neuroimage.2017.12.071

Thibault, R.T., Raz, A., 2016. When can neurofeedback join the clinical armamentarium? The Lancet Psychiatry. https://doi.org/10.1016/S2215-0366(16)30040-2

Thompson, M., Thompson, L., 2002. Biofeedback for Movement Disorders (Dystonia with Parkinson's Disease): Theory and Preliminary Results. J. Neurother. 6, 51–70. https://doi.org/10.1300/J184v06n04

Tohgi, H., Kawashima, M., Tamura, K., Suzuki, H., 1990. Coagulation-fibrinolysis

abnormalities in acute and chronic phases of cerebral thrombosis and embolism. Stroke 21, 1663–1667. https://doi.org/10.1161/01.STR.21.12.1663

Tolin, D.F., 2010. Is cognitive–behavioral therapy more effective than other therapies?: A meta-analytic review. Clin. Psychol. Rev. 30, 710–720. https://doi.org/10.1016/j.cpr.2010.05.003

Tolin, D.F., Gilliam, C., Wootton, B.M., Bowe, W., Bragdon, L.B., Davis, E., Hannan, S.E., Steinman, S.A., Worden, B., Hallion, L.S., 2018. Psychometric Properties of a Structured Diagnostic Interview for DSM-5 Anxiety, Mood, and Obsessive-Compulsive and Related Disorders. Assessment 25, 3–13. https://doi.org/10.1177/1073191116638410

Turner, D.L., Ramos-Murguialday, A., Birbaumer, N., Hoffmann, U., Luft, A., 2013. Neurophysiology of robot-mediated training and therapy: a perspective for future use in clinical populations. Front. Neurol. 4, 184. https://doi.org/10.3389/fneur.2013.00184

Turner, L., Shamseer, L., Altman, D.G., Weeks, L., Peters, J., Kober, T., Dias, S., Schulz, K.F., Plint, A.C., Moher, D., 2012. Consolidated standards of reporting trials (CONSORT) and the completeness of reporting of randomised controlled trials (RCTs) published in medical journals. Cochrane Database Syst. Rev. 11, MR000030. https://doi.org/10.1002/14651858.MR000030.pub2

Turner, R.M., Bird, S.M., Higgins, J.P.T., 2013. The Impact of Study Size on Meta-analyses: Examination of Underpowered Studies in Cochrane Reviews. PLoS One 8, 1–8. https://doi.org/10.1371/journal.pone.0059202

Tzourio-Mazoyer, N., Landeau, B., Papathanassiou, D., Crivello, F., Etard, O., Delcroix, N., Mazoyer, B., Joliot, M., 2002. Automated Anatomical Labeling of Activations in SPM Using a Macroscopic Anatomical Parcellation of the MNI MRI Single-Subject Brain, NeuroImage. https://doi.org/10.1006/nimg.2001.0978

Uher, R., Dernovsek, M.Z., Mors, O., Hauser, J., Souery, D., Zobel, A., Maier, W., Henigsberg, N., Kalember, P., Rietschel, M., Placentino, A., Mendlewicz, J., Aitchison, K.J., McGuffin, P., Farmer, A., 2011. Melancholic, atypical and anxious depression subtypes and outcome of treatment with escitalopram and nortriptyline. J. Affect. Disord. 132, 112–120. https://doi.org/10.1016/j.jad.2011.02.014

Umemura, A., Jaggi, J.L., Hurtig, H.I., Siderowf, A.D., Colcher, A., Stern, M.B., Baltuch, G.H., 2003. Deep brain stimulation for movement disorders: morbidity and mortality in 109 patients. J. Neurosurg. 98, 779–784. https://doi.org/10.3171/jns.2003.98.4.0779

Umscheid, C., Margolis, D., Grossman, C., 2012. NIH Public Access Key Concepts of Clinical Trials : A Narrative Review. Postgr. Med. 123, 194–204.

https://doi.org/10.3810/pgm.2011.09.2475.Key

van de Ven, R.M., Buitenweg, J.I. V., Schmand, B., Veltman, D.J., Aaronson, J.A., Nijboer, T.C.W., Kruiper-Doesborgh, S.J.C., van Bennekom, C.A.M., Rasquin, S.M.C., Ridderinkhof, K.R., Murre, J.M.J., 2017. Brain training improves recovery after stroke but waiting list improves equally: A multicenter randomized controlled trial of a computer-based cognitive flexibility training. PLoS One 12, e0172993. https://doi.org/10.1371/journal.pone.0172993

van Diermen, L., van den Ameele, S., Kamperman, A.M., Sabbe, B.C.G., Vermeulen, T., Schrijvers, D., Birkenhäger, T.K., 2018. Prediction of electroconvulsive therapy response and remission in major depression: meta-analysis. Br. J. Psychiatry 212, 71–80. https://doi.org/10.1192/bjp.2017.28

van Wijk, B.C.M., 2017. Is Broadband Gamma Activity Pathologically Synchronized to the Beta Rhythm in Parkinson's Disease? J. Neurosci. 37, 9347–9349. https://doi.org/10.1523/JNEUROSCI.2023-17.2017

van Wijk, B.C.M., Beudel, M., Jha, A., Oswal, A., Foltynie, T., Hariz, M.I., Limousin, P., Zrinzo, L., Aziz, T.Z., Green, A.L., Brown, P., Litvak, V., 2016. Subthalamic nucleus phase-amplitude coupling correlates with motor impairment in Parkinson's disease. Clin. Neurophysiol. 127, 2010–2019. https://doi.org/10.1016/j.clinph.2016.01.015

Vann, S.D., Aggleton, J.P., Maguire, E.A., 2009. What does the retrosplenial cortex do? Nat. Rev. Neurosci. 10, 792–802. https://doi.org/10.1038/nrn2733

Verstynen, T., 2004. Ipsilateral Motor Cortex Activity During Unimanual Hand Movements Relates to Task Complexity. J. Neurophysiol. 93, 1209–1222. https://doi.org/10.1152/jn.00720.2004

Vrtička, P., Sander, D., Vuilleumier, P., 2011. Effects of emotion regulation strategy on brain responses to the valence and social content of visual scenes. Neuropsychologia 49, 1067–1082. https://doi.org/10.1016/j.neuropsychologia.2011.02.020

Wagenmakers, E.J., Lodewyckx, T., Kuriyal, H., Grasman, R., 2010. Bayesian hypothesis testing for psychologists: A tutorial on the Savage-Dickey method. Cogn. Psychol. 60, 158–189. https://doi.org/10.1016/j.cogpsych.2009.12.001

Walker, E., Nowacki, A.S., 2011. Understanding equivalence and noninferiority testing. J. Gen. Intern. Med. 26, 192–6. https://doi.org/10.1007/s11606-010-1513-8

Wang, D.D., de Hemptinne, C., Miocinovic, S., Ostrem, J.L., Galifianakis, N.B., San Luciano, M., Starr, P.A., 2018. Pallidal deep brain stimulation disrupts pallidal beta oscillations and coherence with primary motor cortex in Parkinson's disease. J. Neurosci. 10, 431–18.

https://doi.org/10.1523/JNEUROSCI.0431-18.2018

Wang, H.C., Lees, a J., Brown, P., 1999. Impairment of EEG desynchronisation before and during movement and its relation to bradykinesia in Parkinson's disease. J Neurol Neurosurg Psychiatry 66, 442–446. https://doi.org/10.1136/jnnp.66.4.442

Wang, T., Mantini, D., Gillebert, C.R., 2017. The potential of real-time fMRI neurofeedback for stroke rehabilitation: A systematic review. Cortex. https://doi.org/10.1016/j.cortex.2017.09.006

Ward, N.S., Brown, M.M., Thompson, A.J., Frackowiak, R.S.J., 2003. Neural correlates of outcome after stroke: A cross-sectional fMRI study. Brain 126, 1430–1448. https://doi.org/10.1093/brain/awg145

Ward, N.S., Cohen, L., 2004. Mechanisms underlying recovery of motor function after stroke. Arch. Neurol. 61, 1844–1848. https://doi.org/10.1136/pgmj.2004.030809

Watanabe, T., Sasaki, Y., Shibata, K., Kawato, M., 2017. Advances in fMRI Real-Time Neurofeedback. Trends Cogn. Sci. 21, 997–1010. https://doi.org/10.1016/J.TICS.2017.09.010

Waters, S., Wiestler, T., Diedrichsen, J., 2017. Cooperation Not Competition: Bihemispheric tDCS and fMRI Show Role for Ipsilateral Hemisphere in Motor Learning. J. Neurosci. 37, 7500–7512. https://doi.org/10.1523/JNEUROSCI.3414-16.2017

Watts, S.E., Turnell, A., Kladnitski, N., Newby, J.M., Andrews, G., 2015. Treatment-as-usual (TAU) is anything but usual: A meta-analysis of CBT versus TAU for anxiety and depression. J. Affect. Disord. 175, 152–167. https://doi.org/10.1016/j.jad.2014.12.025

Weinberger, M., Mahant, N., Hutchison, W.D., Lozano, A.M., Moro, E., Hodaie, M., Lang, A.E., Dostrovsky, J.O., 2006. Beta Oscillatory Activity in the Subthalamic Nucleus and Its Relation to Dopaminergic Response in Parkinson's Disease. J. Neurophysiol. 96, 3248–3256. https://doi.org/10.1152/jn.00697.2006

Weiskopf, N., 2012. Real-time fMRI and its application to neurofeedback. Neuroimage 62, 682–692. https://doi.org/10.1016/j.neuroimage.2011.10.009

Weiskopf, N., Veit, R., Erb, M., Mathiak, K., Grodd, W., Goebel, R., Birbaumer, N., 2003. Physiological self-regulation of regional brain activity using real-time functional magnetic resonance imaging (fMRI): methodology and exemplary data 19, 577–586. https://doi.org/10.1016/S1053-8119(03)00145-9

Wells, A., Davies, M.I., 1994. The Thought Control Questionnaire: a measure of individual differences in the control of unwanted thoughts. Behav. Res. Ther. 32, 871–8.

Wetzels, R., Matzke, D., Lee, M.D., Rouder, J.N., Iverson, G.J., Wagenmakers, E.J., 2011.

Statistical evidence in experimental psychology: An empirical comparison using 855 t tests. Perspect. Psychol. Sci. 6, 291–298. https://doi.org/10.1177/1745691611406923

Wicherts, J.M., Veldkamp, C.L.S., Augusteijn, H.E.M., Bakker, M., van Aert, R.C.M., van Assen, M.A.L.M., 2016. Degrees of freedom in planning, running, analyzing, and reporting psychological studies: A checklist to avoid P-hacking. Front. Psychol. 7, 1–12. https://doi.org/10.3389/fpsyg.2016.01832

Wichmann, T., DeLong, M., 2003. Pathophysiology of Parkinson's disease: The MTPT Primate Model of the Human Disorder. Annu. New York Acad. Sci. 199–213. https://doi.org/10.1196/annals.1355.002

Williams, R.J., Tse, T., DiPiazza, K., Zarin, D.A., 2015. Terminated Trials in the ClinicalTrials.gov Results Database: Evaluation of Availability of Primary Outcome Data and Reasons for Termination. PLoS One 10, e0127242. https://doi.org/10.1371/journal.pone.0127242

Willingham, S., Volkmer, J., Gentles, A., Sahoo, D., Dalerba, P., Mitra, S., Wang, J., Contreras-Trujillo, H., Martin, R., Cohen, J., Lovelace, P., Scheeren, F., Chao, M., Weiskopf, K., Tang, C., Volkmer, A., Naik, T., Storm, T., Mosley, A., Edris, B., Schmid, S., Sun, C., Chua, M., Murillo, O., Rajendran, P., Cha, A., Chin, R., Kim, D., Adorno, M., Raveh, T., Tseng, D., Jaiswal, S., Enger, P., Steinberg, G., Li, G., So, S., Majeti, R., Harsh, G., Rijn, M. van de, Teng, N., Sunwoo, J., Alizadeh, A., Clarke, M., Weissman, I., 2017. The challenges of replication. Elife 6, 6662–6667. https://doi.org/10.7554/eLife.23693

Winkler, A.M., Ridgway, G.R., Webster, M.A., Smith, S.M., Nichols, T.E., 2014. Permutation inference for the general linear model. Neuroimage 92, 381–397. https://doi.org/10.1016/j.neuroimage.2014.01.060

Wolf, M.E., 2015. Functional TCD: Regulation of cerebral hemodynamics - Cerebral autoregulation, vasomotor reactivity, and neurovascular coupling. Front. Neurol. Neurosci. 36, 40–56. https://doi.org/10.1159/000366236

Woo, C., Krishnan, A., Wager, T., 2014. Cluster-extent based thresholding in fMRI analyses: pitfalls and recommendations. Neuroimage.

Woolrich, M.W., Ripley, B.D., Brady, M., Smith, S.M., 2001. Temporal Autocorrelation in Univariate Linear Modeling of FMRI Data. Neuroimage 14, 1370–1386. https://doi.org/10.1006/nimg.2001.0931

World Health Organization, 2017. Depression and other common mental disorders: global health estimates. World Heal. Organ. 1–24. https://doi.org/CC BY-NC-SA 3.0 IGO

Wray, N.R., Ripke, S., Mattheisen, M., Trzaskowski, M., Byrne, E.M., Abdellaoui, A., Adams,

M.J., Agerbo, E., Air, T.M., Andlauer, T.M.F., Bacanu, S.A., Bækvad-Hansen, M., Beekman, A.F.T., Bigdeli, T.B., Binder, E.B., Blackwood, D.R.H., Bryois, J., Buttenschøn, H.N., Bybjerg-Grauholm, J., Cai, N., Castelao, E., Christensen, J.H., Clarke, T.K., Coleman, J.I.R., Colodro-Conde, L., Couvy-Duchesne, B., Craddock, N., Crawford, G.E., Crowley, C.A., Dashti, H.S., Davies, G., Deary, I.J., Degenhardt, F., Derks, E.M., DIrek, N., Dolan, C. V., Dunn, E.C., Eley, T.C., Eriksson, N., Escott-Price, V., Kiadeh, F.H.F., Finucane, H.K., Forstner, A.J., Frank, J., Gaspar, H.A., Gill, M., Giusti-Rodríguez, P., Goes, F.S., Gordon, S.D., Grove, J., Hall, L.S., Hannon, E., Hansen, C.S., Hansen, T.F., Herms, S., Hickie, I.B., Hoffmann, P., Homuth, G., Horn, C., Hottenga, J.J., Hougaard, D.M., Hu, M., Hyde, C.L., Ising, M., Jansen, R., Jin, F., Jorgenson, E., Knowles, J.A., Kohane, I.S., Kraft, J., Kretzschmar, W.W., Krogh, J., Kutalik, Z., Lane, J.M., Li, Y., Li, Y., Lind, P.A., Liu, X., Lu, L., MacIntyre, D.J., MacKinnon, D.F., Maier, R.M., Maier, W., Marchini, J., Mbarek, H., McGrath, P., McGuffin, P., Medland, S.E., Mehta, Di., Middeldorp, C.M., Mihailov, E., Milaneschi, Y., Milani, L., Mill, J., Mondimore, F.M., Montgomery, G.W., Mostafavi, S., Mullins, N., Nauck, M., Ng, B., Nivard, M.G., Nyholt, D.R., O'Reilly, P.F., Oskarsson, H., Owen, M.J., Painter, J.N., Pedersen, C.B., Pedersen, M.G., Peterson, R.E., Pettersson, E., Peyrot, W.J., Pistis, G., Posthuma, D., Purcell, S.M., Quiroz, J.A., Qvist, P., Rice, J.P., Riley, B.P., Rivera, M., Saeed Mirza, S., Saxena, R., Schoevers, R., Schulte, E.C., Shen, L., Shi, J., Shyn, S.I., Sigurdsson, E., Sinnamon, G.B.C., Smit, J.H., Smith, D.J., Stefansson, H., Steinberg, S., Stockmeier, C.A., Streit, F., Strohmaier, J., Tansey, K.E., Teismann, H., Teumer, A., Thompson, W., Thomson, P.A., Thorgeirsson, T.E., Tian, C., Traylor, M., Treutlein, J., Trubetskoy, V., Uitterlinden, A.G., Umbricht, D., Van Der Auwera, S., Van Hemert, A.M., Viktorin, A., Visscher, P.M., Wang, Y., Webb, B.T., Weinsheimer, S.M., Wellmann, J., Willemsen, G., Witt, S.H., Wu, Y., Xi, H.S., Yang, J., Zhang, F., Arolt, V., Baune, B.T., Berger, K., Boomsma, D.I., Cichon, S., Dannlowski, U., De Geus, E.C.J., Depaulo, J.R., Domenici, E., Domschke, K., Esko, T., Grabe, H.J., Hamilton, S.P., Hayward, C., Heath, A.C., Hinds, D.A., Kendler, K.S., Kloiber, S., Lewis, G., Li, Q.S., Lucae, S., Madden, P.F.A., Magnusson, P.K., Martin, N.G., McIntosh, A.M., Metspalu, A., Mors, O., Mortensen, P.B., Müller-Myhsok, B., Nordentoft, M., Nöthen, M.M., O'Donovan, M.C., Paciga, S.A., Pedersen, N.L., Penninx, B.W.J.H., Perlis, R.H., Porteous, D.J., Potash, J.B., Preisig, M., Rietschel, M., Schaefer, C., Schulze, T.G., Smoller, J.W., Stefansson, K., Tiemeier, H., Uher, R., Völzke, H., Weissman, M.M., Werge, T., Winslow, A.R., Lewis, C.M., Levinson, D.F., Breen, G., Børglum, A.D.,

Sullivan, P.F., 2018. Genome-wide association analyses identify 44 risk variants and refine the genetic architecture of major depression. Nat. Genet. 50, 668–681. https://doi.org/10.1038/s41588-018-0090-3

Wylie, G.R., Genova, H., Deluca, J., Chiaravalloti, N., Sumowski, J.F., 2014. Functional magnetic resonance imaging movers and shakers: Does subject-movement cause sampling bias? Hum. Brain Mapp. 35, 1–13. https://doi.org/10.1002/hbm.22150

Young, K., Siegle, G.J., Misaki, M., Zotev, V., Phillips, R., Drevets, W.C., Bodurka, J., 2018a. Altered task-based and resting-state amygdala functional connectivity following real-time fMRI amygdala neurofeedback training in major depressive disorder. NeuroImage Clin. 17, 691–703. https://doi.org/10.1016/j.nicl.2017.12.004

Young, K., Zotev, V., Phillips, R., Misaki, M., Drevets, W.C., Bodurka, J., 2018b. Amygdala Real-Time fMRI Neurofeedback for Major Depressive Disorder: A Review. Psychiatry Clin. Neurosci. 1–16. https://doi.org/10.1111/pcn.12665

Young, K.D., Siegle, G.J., Zotev, V., Phillips, R., Misaki, M., Yuan, H., Drevets, W.C., Bodurka, J., 2017. Randomized Clinical Trial of Real-Time fMRI Amygdala Neurofeedback for Major Depressive Disorder: Effects on Symptoms and Autobiographical Memory Recall. Am. J. Psychiatry 174, 748–755. https://doi.org/10.1176/appi.ajp.2017.16060637

Young, K.D., Zotev, V., Phillips, R., Misaki, M., Yuan, H., Drevets, W.C., Bodurka, J., 2014. Real-Time fMRI Neurofeedback Training of Amygdala Activity in Patients with Major Depressive Disorder. PLoS One 9, e88785. https://doi.org/10.1371/journal.pone.0088785

Yousry, T.A., Schmid, U.D., Alkadhi, H., Schmidt, D., Peraud, A., Buettner, A., Winkler, P., 1997. Localization of the motor hand area to a knob on the precentral gyrus. A new landmark. Brain 141–57.

Yu, L.M., Chan, A.W., Hopewell, S., Deeks, J.J., Altman, D.G., 2010. Reporting on covariate adjustment in randomised controlled trials before and after revision of the 2001 CONSORT statement: A literature review. Trials 11, 1–13. https://doi.org/10.1186/1745-6215-11-59

Yuan, W., Altaye, M., Ret, J., Schmithorst, V., Byars, A.W., Plante, E., Holland, S.K., 2009. Quantification of head motion in children during various fMRI language tasks. Hum. Brain Mapp. 30, 1481–1489. https://doi.org/10.1002/hbm.20616

Zaidel, A., Arkadir, D., Israel, Z., Bergman, H., 2009. Akineto-rigid vs. tremor syndromes in Parkinsonism. Curr. Opin. Neurol. 22, 387–393. https://doi.org/10.1097/WCO.0b013e32832d9d67

Zaidel, A., Spivak, A., Grieb, B., Bergman, H., Israel, Z., 2010. Subthalamic span of β oscillations predicts deep brain stimulation efficacy for patients with Parkinson's disease. Brain 133, 2007–2021. https://doi.org/10.1093/brain/awq144

Zaremba, D., Dohm, K., Redlich, R., Grotegerd, D., Strojny, R., Meinert, S., Burger, C., Enneking, V., Forster, K., Repple, J., Opel, N., Baune, B.T., Zwitserlood, P., Heindel, W., Arolt, V., Kugel, H., Dannlowski, U., 2018. Association of brain cortical changes with relapse in patients with major depressive disorder. JAMA Psychiatry 75, 484–492. https://doi.org/10.1001/jamapsychiatry.2018.0123

Zatorre, R.J., Fields, R.D., Johansen-Berg, H., 2012. Plasticity in gray and white: Neuroimaging changes in brain structure during learning. Nat. Neurosci. 15, 528–536. https://doi.org/10.1038/nn.3045

Zhao, Y.J., Wee, H.L., Chan, Y.H., Seah, S.H., Au, W.L., Lau, P.N., Pica, E.C., Li, S.C., Luo, N., Tan, L.C.S., 2010. Progression of Parkinson's disease as evaluated by Hoehn and Yahr stage transition times. Mov. Disord. 25, 710–716. https://doi.org/10.1002/mds.22875

Zhao, Z., Yao, S., Li, K., Sindermann, C., Zhou, F., Zhao, W., Becker, B., 2018. Real-time functional connectivity-based neurofeedback of amygdala-frontal pathways reduces anxiety. bioRxiv.

Zich, C., Debener, S., Schweinitz, C., Sterr, A., Meekes, J., Kranczioch, C., 2017. High-Intensity Chronic Stroke Motor Imagery Neurofeedback Training at Home: Three Case Reports. Clin. EEG Neurosci. 48, 403–412. https://doi.org/10.1177/1550059417717398

Zigmond, A.S., Snaith, R.P., 1983. The hospital anxiety and depression scale. Acta Psychiatr. Scand. 67, 361–70.

Zilles, K., Schlaug, G., Geyer, S., Luppino, G., Matelli, M., Qü, M., Schleicher, A., Schormann, T., 1996. Anatomy and transmitter receptors of the supplementary motor areas in the human and nonhuman primate brain. Adv. Neurol. 70, 29–43.

Zilverstand, A., Parvaz, M.A., Goldstein, R.Z., 2017. Neuroimaging cognitive reappraisal in clinical populations to define neural targets for enhancing emotion regulation. A systematic review. Neuroimage 151, 105–116. https://doi.org/10.1016/j.neuroimage.2016.06.009

Zilverstand, A., Sorger, B., Sarkheil, P., Goebel, R., 2015. fMRI neurofeedback facilitates anxiety regulation in females with spider phobia 9, 1–12. https://doi.org/10.3389/fnbeh.2015.00148

Zimmermann-Schlatter, A., Schuster, C., Puhan, M.A., Siekierka, E., Steurer, J., 2008. Efficacy of motor imagery in post-stroke rehabilitation: A systematic review. J. Neuroeng.

Rehabil. 5. https://doi.org/10.1186/1743-0003-5-8

Zotev, V., Krueger, F., Phillips, R., Alvarez, R.P., Simmons, W.K., Bellgowan, P., Drevets, W.C., Bodurka, J., 2011. Self-Regulation of Amygdala Activation Using Real-Time fMRI Neurofeedback. PLoS One 6, e24522. https://doi.org/10.1371/journal.pone.0024522